Deciphering the recognition features of glycans by human lectins at the molecular level

Maria Pia Lenza

Doctoral Thesis

2021

Supervisors:

Dr. Jesús Jiménez-Barbero

Dr. June Ereño-Orbea

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Contents

Abbreviation

ACE2 angiotensin-converting enzyme 2

ADCC Antibody-Dependent cell-mediated cytotoxicity

BLI Biolayer Interferometry

BMRB Biological Magnetic Resonance Data Bank

CD Circular Dichroism

CHO Chinese Hamster Ovary

CMAH cytidine monophosphate-N--acetylneuraminic acid hydroxylase

COVID Coronavirus Disease

CRD Carbohydrate Recognition Domain

cryoEM cryo-electron microscopy

CSP Chemical Shift Perturbation

CTL C-type lectin

CTLD C-type lectin domain

CuAAC copper(I)-catalyzed azide-alkyne cycloaddition

DAMP damage-associated molecular pattern

DC Dendritic Cell

DC-SIGN Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin

DMEM Dulbecco's Modified Eagle Medium

DTT Dithiothreitol

E.coli *Escherichia Coli*

EBOV Ebola Virus

ECD Extra-Cellular Domain

Endo H Endoglycosidase H

ENV Envelope

EO Eosinophil

ER Endoplasmatic Reticulum

FAB Fragment antigen binding

FBS Fetal Bovine Serum

FC Fragment Crystallizable

GA Golgi Apparatus

GBP Glycan Binding Protein

GnTI N-acetylglucosaminyltransferase I

HB Hydrogen Bond

HC Heavy Chain

HEK Human Embryonic Kidney cell

HIV human immunodeficiency virus

HSQC Heteronuclear Single-Quantum Coherence

Ig Immunoglobuline

IPTG Isopropyl β- d-1-thiogalactopyranoside

ITIM Immunoreceptor tyrosine-based inhibitory motif

LB Luria Bertani

LC Light Chain

mAb Monoconal Antibody

MAG Myelin Associated Glycoprotein

MC Mast Cell

MGL Macrophage galactose-type lectin

MR Mannose Receptor

MW molecular weight

NEAA Non-Essential Amino Acids

nLC MS/MS nano-scale liquid chromatographic tandem mass spectrometry

NMR Nuclear Magnetic Resonance

NOE Nuclear Overhauser effect

NOESY Nuclear Overhauser Effect Spectroscopy

PAMP Phatogen-Associated Molecular Pattern

PBS Phosphate Buffered Saline

PDB Protein Data Bank

PMSF phenylmethylsulfonyl fluoride

RBD Receptor Binding Domain

SAR structure–activity relationship

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus-2

SASA Solvent Accesible Surface Area

SDS-PAGE Sodium Dodecyl Sulfate-polyacrylamide Gel Electrophoresis

Siglec Sialic acid-binding immunoglobulin-type lectins

SPR Surface plasmon resonance

STD Saturation Transfer Difference

TEV Tobacco Etch Virus

TOCSY Total correlation spectroscopy

TRIS tris(hydroxymethyl)aminomethane

Tr-NOESY Transferred-Nuclear Overhauser Effect Spectroscopy

TROSY Transverse Relaxation-Optimized Spectroscopy

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

WaterLOGSY Water-Ligand Observed through Gradient Spectroscopy

WT wild type

Resumen

Las macromoléculas biológicas, como las proteínas, los lípidos y los ácidos nucleicos, son las moléculas principales y esenciales para la vida. Todas estas macromoléculas sufren modificaciones durante su proceso de formación. Concretamente, una de las modificaciones mas abundantes dentro de la célula es la glicosilación. La glicosilación de proteínas consiste en la adición sucesiva de diferentes monosacáridos sobre un aminoácido específico en la cadena polipeptídica correspondiente para finalmente obtener una cadena de oligosacáridos compleja denominada glicano. Los glicanos pueden ser lineales o ramificados y pueden estar unidos por enlaces glicosídicos α o β, en distintas posiciones. De esta forma, las diferentes combinaciones de los azúcares crean una enorme diversidad química.

Existen diferentes patrones de glicosilación en las glicoproteínas, pero los dos más frecuentes son la N- y la O-glicosilación. Las funciones de los glicanos son muy diversas: están involucrados en el plegamiento, tráfico y estabilidad de proteínas y regulan muchas actividades celulares, especialmente en aquellas de naturaleza extracelular. Además, la glicosilación también se encuentra en las proteínas de superficies de los virus. De hecho, los virus no son capaces de glicosilar sus proteínas, sino que controlan la maquinaria de glicosilación de la célula huésped para que lo haga. La presencia de múltiples glicanos en las glicoproteínas virales sugiere que contribuyen a su supervivencia y a generar infecciones en sus anfitriones.

Entre todos las funciones críticas que juegan los glicanos, su reconocimiento por lectinas juega un papel crucial en la respuesta inmunológica. Entre las distintas familias de lectinas, se pueden identificar galectinas, lectinas de tipo C y Siglecs, según el epítopo de carbohidrato que reconocen. El conocimiento, a nivel molecular, de cómo los carbohidratos interaccionan con las lectinas puede

ayudar al entendimiento de las señales biológicas en situaciones fisiológicas y patológicas, y eventualmente, a cómo se puede modular esta respuesta.

Existen varias herramientas para el estudio de estas interacciones a escala atómica, como la Resonancia Magnética Nuclear (RMN), la difracción de rayos X y la crio-microscopía electrónica (cryo-EM). En este trabajo, hemos aplicado la espectroscopia de resonancia magnética nuclear (RMN) para obtener conocimientos estructurales sobre diferentes procesos de reconocimiento molecular entre glicanos en glicoproteínas intactas frente a diferentes lectinas humanas. De esta manera, hemos analizado la composición de los glicanos de dos glicoproteínas, el FcεRIα y el RBD de la proteína Spike del SARS CoV-2 y su interacción con lectinas humanas, desde la perspectiva de los glicanos y de las lectinas.

Finalmente, los detalles de la interacción molecular entre Siglec-8 y dos glicomiméticos sintéticos de alta afinidad han sido diseccionados mediante RMN. También se han aplicado técnicas de cristalografía de rayos X para dilucidar la estructura de Siglec-8 en su estado apo, asociada al glico-mimético y a un fragmento de un anticuerpo (Fab).

Capítulo I

En la introducción de esta tesis, he proporcionado la visión general del proceso de glicosilación de proteínas y del reconocimiento de glicanos por parte de lectinas. Además, entre todas las lectinas, hemos enfocado la atención sobre Siglec-8, que es un objetivo farmacológico en la modulación de la respuesta alérgica, tanto mediante el uso de glicomiméticos como de anticuerpos.

En la introducción general se describen las técnicas de cristalografía de Rayos X y de RMN más utilizadas en el estudio de las interacciones entre los glicanos y sus receptores. Entre las técnicas de cristalografía de Rayos X, se describe cómo se puede resolver la estructura cristalográfica de un receptor con su ligando mediante dos metodologías alternativas. La primera (*soaking*) se basa

en la inmersión de los cristales de proteína preformados en una disolución del ligando a alta concentración y la segunda, intenta la co-cristalizacion ab-initio del complejo ligando-proteína preformado.

Desde el punto de vista de la RMN, el estudio de fenómenos de reconocimiento molecular puede hacerse desde el punto de vista del ligando y/o del receptor. Uno de los experimentos de RMN que enfocan la atención sobre el ligando es el experimento de diferencia de transferencia de saturación (STD-NMR) que se basa en la transferencia de magnetización entre el receptor, irradiado selectivamente, y aquellos ligandos que se unen a él. Desde el punto de vista del receptor, los experimentos basados en la correlación heteronuclear de cuanto simple ^{1}H-^{15}N (^{1}H,^{15}N-HSQC) permiten de evaluar los cambios en las señales de RMN de los aminoácidos de la proteína cercanos al sitio de unión del ligando.

Capitulo II

Desde el punto de vista formativo, algunos de los objetivos clave de esta Tesis han sido adquirir conocimientos sobre la expresión y purificación de glicoproteínas y lectinas en diferentes sistemas de expresión, aprender a aplicar metodologías de RMN y de Cristalografía de Rayos X para estudiar la interacción entre carbohidratos y lectinas y dominar conceptos y estrategias para abordar problemas científicos sobre el reconocimiento biomolecular.

Desde la perspectiva científica, los objetivos fueron: primero, establecer una nueva metodología basada en RMN para descifrar la composición de glicanos en una glicoproteína intacta (FcεRIα) y analizar cómo estos glicanos interaccionan con una lectina en condiciones experimentales cercanas a las fisiológicas.

En una segunda parte, y en el contexto de la pandemia actual, el objetivo fue analizar el glicoperfil del RBD del SARS-CoV-2 y evaluar sus interacciones con lectinas de nuestro sistema inmunológico, utilizando la metodología de RMN desarrollada para FcεRIα.

En una tercera parte, el objetivo ha sido iniciar una nueva línea de investigación en el laboratorio, centrada en Siglecs. Sobre esta base, el objetivo de esta Tesis ha sido desarrollar y establecer la metodología y los protocolos necesarios para combinar datos de cristalografía de rayos X y RMN para estudiar las interacciones Siglec-ligando, utilizando Siglec-8 como objetivo inicial. Por tanto, el estudio de la estructura de Siglec-8 y sus interacciones con glicomiméticos y anticuerpos se abordó tanto por RMN como por cristalografía de rayos X.

Capitulo IV

En este capítulo se ha analizado el glicoperfil del FcεRIα expresado en células HEK 293T humanas, identificando la presencia y la abundancia relativa de epítopos de glicanos específicos, utilizando RMN. La región anomérica del espectro ^{1}H, ^{13}C-HSQC muestra un alto grado de heterogeneidad en términos de glicosilación. La asignación de los picos anoméricos se validó mediante el uso de enzimas, donde la glicoproteína se incubó secuencialmente con diferentes glicosidasas. Además, se ha generado y estudiado el mutante Asn132Ala para demonstrar la presencia de un glicano especifico en esta posición. Se ha propuesto también una metodología para deducir las interacciones de glicoproteínas intactas con lectinas, utilizando Gal-3 como modelo.

Capítulo V

En este capítulo se ha descrito la producción del RBD de la glicoproteína spike de SARS-CoV-2 en un cultivo de células humanas (HEK293F), con los N-glicanos etiquetados con isótopos estables de ^{13}C en. Esto ha permitido la caracterización detallada, sin precedentes, de las estructuras específicas de sus glicanos. Además, se han usado experimentos ^{1}H-^{13}C HSQC sobre el RBD para diseccionar la interacción de este dominio con una variedad de lectinas humanas, que se expresan en diferentes órganos y tejidos que pueden verse

afectados durante la infección. Se ha determinado la especificidad de los epítopos de glicano que son responsables de la interacción con cada lectina.

Capítulo VI

En este capítulo, he descrito cómo se ha expresado y purificado el dominio Siglec-8 V en condiciones de marcaje con isótopos estables de 15N y sin marcar.

Se han cristalizado diferentes construcciones de Siglec-8, en estado apo, con un ligando glicomimético de alta afinidad y con AK002, el anticuerpo utilizado en ensayos clínicos específicos para Siglec-8. Se ha resuelto la estructura cristalográfica de la forma apo, mientras que el análisis de los cristales de los complejos aún está bajo investigación.

El uso de la proteína marcada en ^{15}N ha permitido el estudio de las perturbaciones del desplazamiento químico de las señales de RMN de los aminoácidos implicados en la unión a sus ligandos glicomiméticos. Además, el uso del experimento STD NMR ha proporcionado información sobre el epítopo de unión desde el punto de vista de los ligandos. De acuerdo con los datos experimentales del CSP y STD, se ha propuesto un modelo 3D del complejo con un glicomimético de alta afinidad.

Finalmente, hemos evaluado como la Siglec-8 interacciona con los glicanos del FcεRIα, su vecino en la membrana de los mastocitos y eosinófilos.

Abstract

N-linked glycosylation is a protein post-translational modification that results in the covalent attachment of an oligosaccharide onto a given asparagine residue at the polypeptide chain. It is a highly complex process that generates can a wide array of chemical structures. Glycans are involved in protein folding, trafficking, and stability and regulate many cellular activities, especially of extracellular nature. Among all the critical roles that glycans play, their recognition by human lectins play a crucial role in the immune response.

Herein, we have applied Nuclear Magnetic Resonance (NMR) spectroscopy to gain structural insights into different molecular recognition processes between glycans on intact glycoprotein and lectins. We have analyzed the glycan composition of two glycoproteins, the FcεRIα and the RBD of the Spike protein of SARS.CoV-2 as well as their interactions with human lectins, both from the glycans' and lectins' perspective.

Finally, in this Thesis we have developed and established the required methodology and protocols to combine NMR and X-Ray crystallography data to study Siglec-ligand interactions, using Siglec-8 as initial target. Thus, the study of the structure of Siglec-8 and their interactions with glycomimetics and antibodies has been tackled by both NMR and X-Ray crystallography.

In particular, X-ray crystallography techniques havd been applied to elucidate the structure of Siglec-8 in the apo form, as well as bound to a high affinity ligand and to a therapeutic antibody. Additionally, the details of the molecular features between Siglec-8 and two high-affinity glycomimetics have dissected using NMR spectroscopy.

Chapter I

General Introduction

1.1 Glycan in nature

Glycans are everywhere. Glycans are everywhere. Besides their very well-known functions as energy sources and structural components, carbohydrates/glycans/saccharides/sugars play a key role in cellular communication events. They are usually presented as glycoconjugates, especially as glycoproteins and glycolipids although they can also be presented as oligo and polysaccharides [1]. It is known, that glycans form a highly charged layer on the surface of the cells, the glycocalyx. This layer functions as a barrier between a cell and its surrounding, but also serves as a mediator for cell-cell interactions and protects a cell membrane from the direct action of physical forces and stresses. Although glycosylation is largely present on lipids and proteins of the cells surface, large body of data documents the presence of glycoconjugates within the nucleus and cytoplasm. Indeed, the presence of glycans in nucleic acids has also been postulated [2]. Thus, glycans are generally found on the three major classes of macromolecules (proteins, lipids and nucleic acids) in all compartments of the cell and play essential roles for life and disease [3].

In the last decades, it has been demonstrated that glycans are involved in a variety of biological processes. Arbitrarily, these events can be divided into two main groups: i) intrinsic functions and ii) extrinsic functions. In the first case, structural and modulatory properties (such as nutrient storage) can be included. In contrast, in the extrinsic functions, the interactions of the glycans with other partners (specially, with Glycan Binding Proteins (GBPs), such as lectins) and their molecular mimicry by pathogens to avoid the immune system can be considered [1]. Indeed, the mutual recognition between glycans and lectins is a highly regulated process in the immune system. Through the binding to lectins, glycans are essential in various cellular mechanisms that can contribute to the adaptive and innate immune system [4]. This research field is continuously experiencing a boost, providing new discoveries from the fundamental level and

promising new applications in prevention, diagnosis and therapies in the near future.

1.1.1 Glycosylation in proteins

In the protein context, in general, the amino acid sequence of a protein determines its function and its properties. However, during the synthesis and the maturation process, proteins can undergo many post-translational modifications. One of the most abundant modifications is glycosylation, common to all eukaryotic cells [5]. Protein glycosylation consists of the successive addition of different monosaccharides on a specific amino acid in the corresponding polypeptide to finally obtain a complex oligosaccharide chain dubbed glycan [6]. The individual monosaccharides commonly found in mammalian glycoproteins and glycolipids are shown in Figure 1.1 [5].

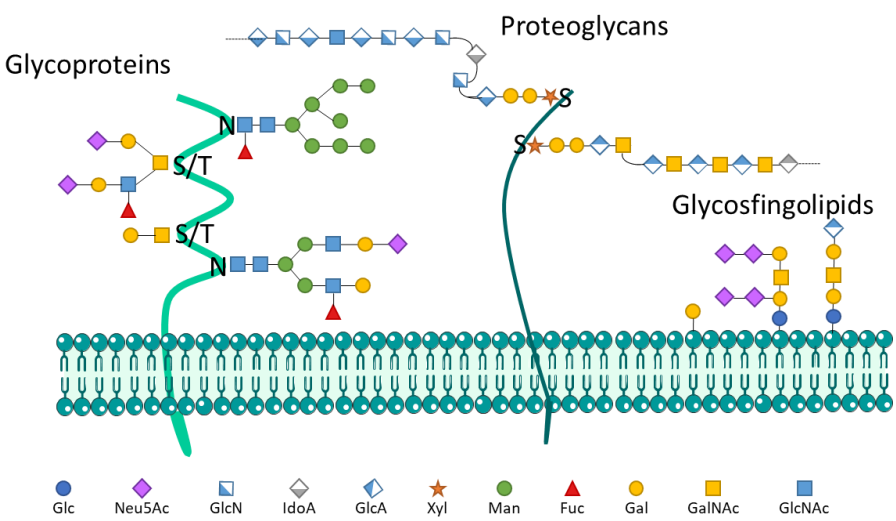


***Figure 1.1**: Schematic representation of the typical glycans on eucaryotic cells. Glycans can be attached to proteins (glycoproteins and proteoglycans) or to lipids (glycosphingolipids). At the bottom of the figure, the most common monosaccharides present in eucariotyc cells are represented as SNFG symbols: D-Glucose (Glc), Sialic Acid (Neu5Ac), D-Glucosamine (GlcN), L-Iduronic Acid (IdoA), D-Glucuronic Acid (GlcA), D-Xylose (Xyl), D-Mannose (Man), L-Fucose (Fuc), D-Galactose (Gal), D-N-Acetyl Galactosamine (GalNAc), D-N-Acetyl Glucosamine (GlcNAc).*

Glycans can display linear or branching presentations and may be bound by α or β- glycosidic linkage at various positions. In this way, the different combination of sugars creates a tremendous chemical diversity [7]. The addition of the first monosaccharides generally occurs in the Endoplasmic Reticulum (ER). The process continues in the cis-, medium- and trans-Golgi apparatus (GA), where a variety of glycosyltransferases and endoglycosidases modify the glycan composition [5].

There are different glycosylation patterns in glycoproteins, but the two more frequent are N-glycosylation and O-glycosylation; these two types are generally found on most proteins trafficking the cellular secretory pathway [6]. N-glycosylation starts with the attachment of N-acetyl glucosamine (GlcNAc) to the nitrogen atom of an asparagine side chain by a linkage β1-N (N-glycosylation), through the recognition of a specific sequence motif, Asn-X-Ser/Thr, where X is any amino acid except proline [8]. N-glycan's structure contains a common $(GlcNAc)_2$ $(Man)_3$ core structure to which other monosaccharides can be attached, including Mannose (Man), GlcNAc, Galactose (Gal), Fucose (Fuc), and Sialic Acid (Neu5Ac), through different glycosidic linkages, to build bi-, tri- or tetra-antennary structures (Figure 1.2A). Based on the sugar composition, N-glycans can also be divided into three different major structures (Figure 1.2A): i) high-Man, ii) hybrid-type and iii) complex-type [6].

In the case of the O-glycosylation, the first step involves the addition of N-acetyl galactosamine (GalNAc) to the hydroxyl group of a serine or threonine, which can then be extended into different structures. GalNAc type O-glycans are also dubbed mucin-type O-glycans, given their ubiquitous presence in these glycoproteins (mucins). They are composed of four major glycan core structures (Core 1-4), while there is one alternative core which is based on the presence of the sialyl-T antigen (as shown in Figure1.2B) [9].

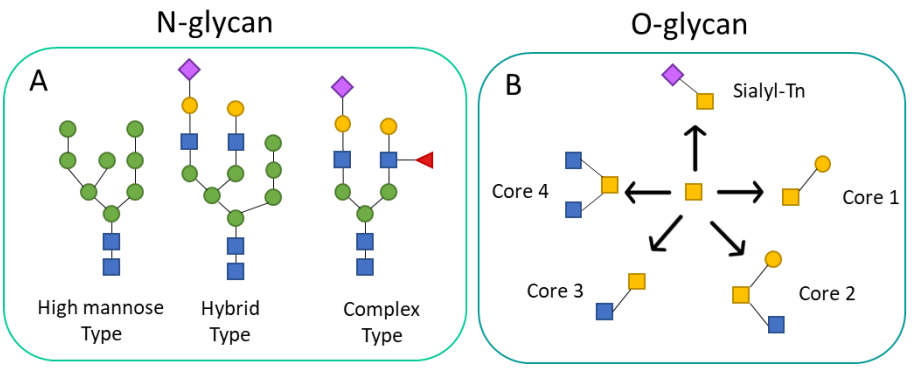


Figure1.2: *The major N-glycan structures (A) and O-glycan structures (B).*

It is essential to highlight that the synthesis of glycans is a rather complex process. Moreover, unlike the genome, exome, or proteome, the glycome is produced in a non-template manner. The final glycan composition in specific cell results from the contribution of different factors: the intrinsic protein structure, the availability of glycan-precursors, the expression level of specific glycosidases and glycosyltransferases in the ER and the GA, and the organization of the different enzymes in the GA [10]. In the mammalian genome, over 200 genes of glycosyltransferases and glycosidases are present. Most of these enzymes suffer transcriptional modulations at the genome level or are localized into different cellular compartments, depending on the cellular phases, energetic level, and oxidation stress. These are the factors that cells use to modify the glycan composition, after internal or external stimuli. Moreover, the physiological state of the cells can also influence the glycome; indeed, a dramatic alteration of the glycosylation pattern is observed in different pathologies such as cancer and autoimmune, infectious and chronic diseases [6]. For example, changes in human Immunoglobulin G (IgG) glycosylation (at the Asn297-linked glycan) have been observed in aging and in various diseases. Differential IgG glycosylation is known to modulate IgG effector functions. In fact, IgG substituted with galactose-deficient N-glycans is pro-inflammatory, while decorated with sialylated N-glycans is anti-inflammatory [11].

1.1.2 Glycans in Viruses

Most enveloped viruses, including influenza, human immunodeficiency virus (HIV), Ebola virus (EBOV), and Coronavirus, display highly glycosylated proteins on their surface [12]. However, a virus particle only contains a viral genome, which is protected by many proteins [13]. Therefore, they are not capable of replicating themselves, being obliged to behave as intracellular parasites. This means that the viruses need to infect a host cell to manipulate the protein synthesis pathway and continue its life cycle [13]. Indeed, enveloped viruses have evolved with envelope proteins (ENV) that display several N-linked glycosylation sites. These linked carbohydrate chains (glycans) are synthesized by the host cell glycosylation machinery. The presence of multiple glycans on viral glycoproteins suggests that they contribute to survival of viruses in their hosts.

The roles of these glycans have been a matter of debate for years. It is now well established that viruses can modulate or alter the host-cell glycosylation machinery [14] and that host cell glycosylation can be involved in several steps of the viral infection [12]. We can assess that there is a consensus on the role of glycans in the following points: i) they provide a protective hydration layer around the virus due to their hydrophilic nature, ii) the sialic acids confer a negative electrostatic charge that may prevent nonspecific adhesion to cell membranes, iii) some glycans specifically bind to a variety of cell surface receptors and iv) glycans may also play a role as a shield (dubbed "*glycan shield*") to efficiently mask key peptide antigens, thus preventing the recognition by neutralizing antibodies. In this way, the efficacy of the protective immune response by the host is highly decreased. However, glycans may also be considered as the "*Achilles Heel*" of the virus, contributing to its vulnerability by promoting its neutralization and destruction. These events are also mediated by natural antibodies that bind to carbohydrate antigens on the virus glycan

shield. Production of these natural antibodies in humans (mammals, in general) takes place along the whole life without the requirement of active immunization.

As shown in Figure 1.3, one of the functional roles of glycans on viral glycoproteins is to influence protein folding and to enhance the stability. Both N- and O-glycans can be involved in the viral particle formation and in the infectivity process.

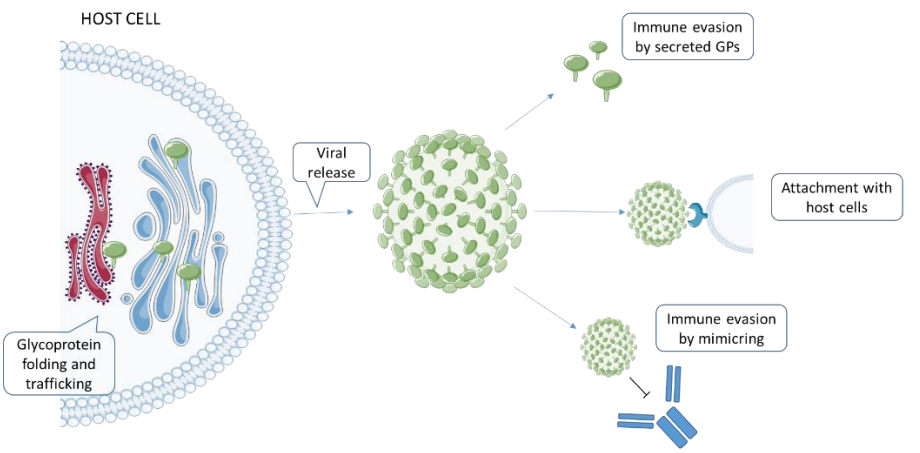


Figure 1.3: *Roles of glycosylation in viral pathogenesis. Glycosylation is important for the protein folding and stability in the ER and GA of host cells. When the virus is released, glycosylation can modulate the virus recognition events, even avoiding the immune response from the host organism.*

For example, in the case of the envelope (Env) protein of HIV, the presence of the N-glycans is essential for stabilizing the assembly of the Env trimer, and its lack causes alteration in the virion formation process [15]. Some viruses can also use glycans to evade the immune system in different manners; for example, viruses can promote the secretion of glycoproteins from host cells, as in EBOV. Indeed, the membrane fusion protein (ssGP) of Ebola is expressed as a transmembrane and secreted form. The role of this secreted sGP is to be an "antigenic subvert" of the immune system, acting as bait for the recruitment of antibodies that recognize the same epitopes of the two sGP isoforms [16]. Another way to escape from the immune system control is to use the host-

derived glycans to decorate the envelope protein, thus masking the virus epitopes and preventing the antibody recognition of the underlying protein surface. This is the case of coronavirus, which is the cause of the health emergency we live in this period. The Spike protein of SARS-CoV-2 is a trimeric protein with 22 N-glycosylation sites and 5 O-glycosylation sites [17]. It has been proposed that the high glycosylation on the coronavirus can facilitate the immune evasion [18][19]. Different studies [17][20][21][22] have demonstrated the presence of a high a variety of N-glycan composition, suggesting that most of the sites can be modified by glycosyl transferases in the GA. Regarding O-glycans, the presence of mucin-like structures has also been proposed [23]. The involvement of the glycans on the key open-closed presentation of the spike receptor binding domain (RBD) to interact with the angiotensin-converting enzyme 2 (ACE2) has been also discussed [17][24]. It is obvious that the elucidation of the glycan composition of the spike protein can provide vital information to understand the viral binding with the target cells, followed by the fusion and replication events. In the long term, it could also provide new information for the design of antigens for future vaccines.

1.2 Human Lectins

The term “lectin” (from latin: *legere*) was firstly coined in 1954 and, since the 1970’s, it has been systematically employed to refer to GBPs [1]. Lectins are ubiquitous in nature, in all kingdoms: microorganisms, plants and animals, and at diverse cellular locations. Lectins work as information mediators in a plethora of molecular recognition processes and interact with specific glycan epitopes on endogenous or exogenous saccharides in glycoproteins and glycolipids, without modifying them (non-enzymatic interaction). Worth noting, their heterogeneity in many aspects, including their function, structure, specificity, cellular location and phylogenetic distribution makes it difficult to establish general classification criteria. In *Animalia*, the categories are defined

following both structural signatures and specific physiological roles and subcellular location (Figure 1.4). For instance, Galectins orchestrate immunological responses, participating in glycan crosslinking at the extracellular matrix, while Siglecs are located at the cell surface and modulate cell-cell adhesion events through the cis- and trans-interaction with endogenous sialic acid residues.

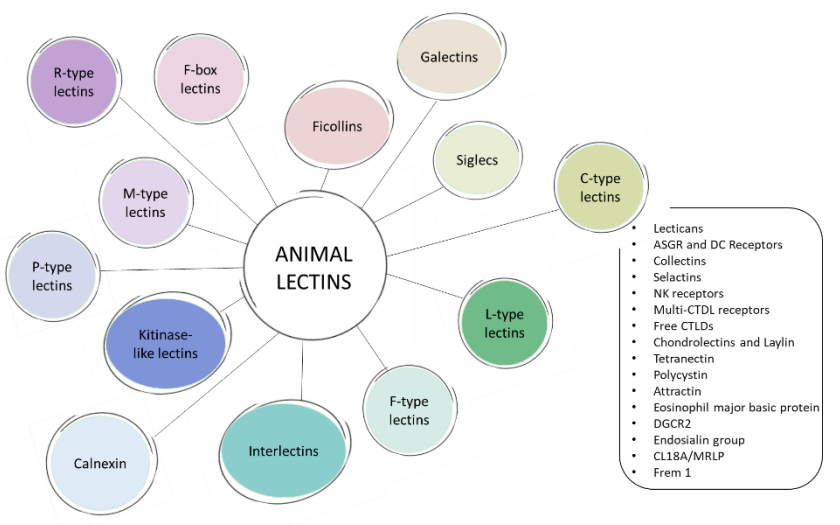


Figure 1.4: *Schematic representation of lectin's classification.*

Glycan antigens are key for maintaining homeostasis. In particular, human lectins are GBPs generally expressed by immune cells, which can recognize glycans as pathogen-associated molecular patterns (PAMPs) but also damage-associated molecular patterns (DAMPs). Therefore, lectins play a crucial role in the immunologic response against infections caused by pathogens and in immune pathologies [25]. From a structural perspective, this lectin-sugar cross-talk is fairly complex. Indeed, a variety of multiple membrane proteins control these molecular recognition processes, including Toll-like receptors. In the last two decades, lectins have become targets for therapeutics. The aim is to use them for modulating or precluding the progression of pathologies arisen from

incorrect immune outcomes. Thus, a detailed knowledge of the recognition features of lectins is mandatory for the development of specific treatments.

Lectins interact with the glycans through their carbohydrate recognition domains (CRDs). Usually, glycans fit into shallow but well-defined binding pockets on the lectin's surfaces (see below in 1.3). There are different classifications (Figure 1.4) for mammalian lectins. One of the most commonly employed is the following one: i) S-type lectins (Galectins), ii) C-type lectins (CTL), and iii) I-type lectins (Siglecs) [1].

1.2.1 S-type lectins (Galectins)

Nowadays known as Galectins, S-type lectins are one of the most expressed lectins in all organisms [1]. As the name suggests, these proteins are characterized by the presence of a β-galactoside binding site, which recognizes Gal-terminated oligosaccharides at glycoproteins or glycolipids [26].

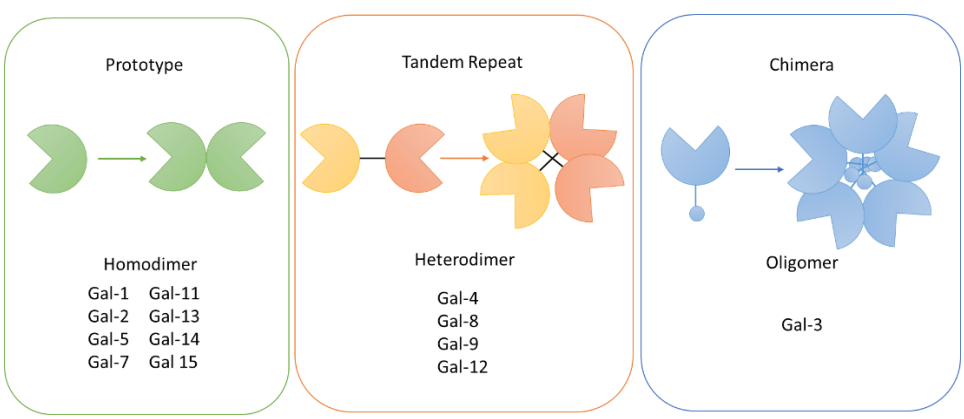


Figure 1.5: *Schematic representation of Galectins. Based on the CRD and its assembly, Galectins can be divided into prototype, tandem repeat and chimera type.*

Galectins can be classified (Figure 1.5) into three major types: prototypical galectins (Gal-1, -2, -7, 10, 11, -13, -14, and -15) that are organized in homodimers, chimera-type galectins (Gal-3), which just display one single CRD able to form oligomers through a N-terminal polypeptide chain, and tandem-

repeat galectins (Gal-4, -8, -9 and -12), which contain two CRDs connected by a linker [27].

The sequence identity between galectins is fairly high. Moreover, the structural pattern of the CRDs is conserved and formed by six (from S1 to S6) and five (from F1 to F5) antiparallel beta-strands (Figure 1.6). The β-Gal-binding site is located at the S-face, on subsite C and may be further extended away through vicinal subsites A, B, D and E. These extensions display important variations between the different members of the family (Figure 1.6).

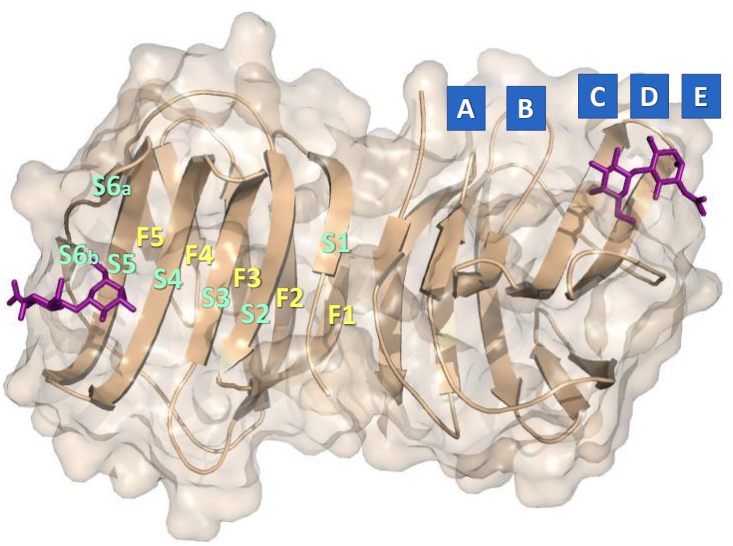


Figure 1.6: *Cartoon and surface representation of the CRD of hGal-1 (dimer) in complex with N-Acetyllactosamine (PDB ID: 1W6P) with subsites (A, B, C, D, E) of the binding site and architecture of the beta strands (S1, S2, S3, S4, S5, S6, F1, F2, F3, F4, F5, F6) highlighted.*

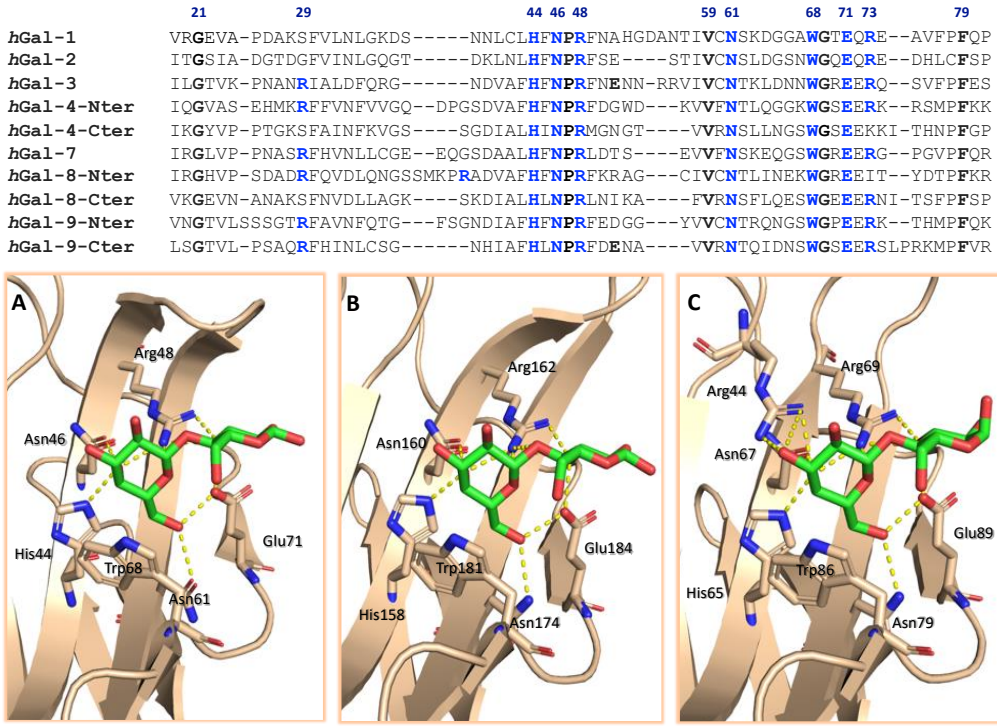


Figure 1.7: *top: Sequence alignment of different galectins. Conserved residues are highlighted in bold; those in blue are involved in the interaction with Gal moieties. Bottom: Reported binding sites for diverse galectin complexes A) hGal-1/lactose (PDB ID: 1GZW), B) hGal-3/lactose (PDB ID: 2NN8) and hGal-8-Nterminal/lactose (PDB ID: 5T7S). Those residues involved in the interaction are displayed as sticks while H bonds displayed as yellow dots. Taken from* [28].

The canonical binding site of galectins for β-galactosides is rather shallow and solvent exposed. The partners interact through hydrogen bonds between specific His, Arg and Asn lectin residues and properly located hydroxyl groups t the β-Gal moiety and through CH-π stacking interactions between a conserved Trp residue and the non-polar face of the β-Gal moiety (Figure 1.7).

Galectins are involved in a wide range of biological activities, both beneficial and harmful, as homeostasis, apoptosis or vascular embryogenesis [29]. They are involved in pathological events, including inflammation [30], host-pathogen interaction [31], antibacterial autophagy [32] and cancer [33][34]. In fact, many members of the family are associated with a plethora of carcinogenesis phenomena: apoptosis, cell transformation, adhesion, migration, invasion,

metastasis, immune escape and angiogenesis [35]. The fine characterization, at the molecular level, of the interactions of galectins with their natural ligands is providing the structural basis for the design of potent antagonists.

Notwithstanding glycans are natural ligands of galectins, they are not the best antagonist candidates because of their lack of good properties to act as pharmacological agents, their low metabolic stability, and their high hydrophilicity. Moreover, the affinities for single carbohydrate-galectin interactions are usually rather weak, in the high µM range, very far from the nM affinity required for entering in the drug discovery process. This fact derives the intrinsic chemical nature of glycans: the absence of hydrophobic patches to match with the protein surface, the hydrogen bond-based interplay with the receptor recognition site that competes with the bulk water and the huge associated enthalpic penalties for the required desolvation of the shallow receptor binding site [36].

Galectins interact by a selective cross-link to the appropriated glycans on the immune system. Then, they play a crucial role in maturation, survival, activation, and infection processes [37][38]. For instance, overexpression of Gal-3 in monocytes can induce differentiation in macrophages [33].

Moreover, galectins are also involved in viral infections. For example, it has been demonstrated that Gal-3 induces cell death in macrophages infected by HIV [39] and that it is also involved in the promoting of in the airway pneumococcal infection. Indeed, Gal-3 can promote the adhesion between the pathogen and broncho-alveolar cells [33].

1.2.2 C-type lectins (CTL)

Among animal lectins, C-type lectins are a set of GBPs that need coordination by a Ca^{2+} ion in the binding site to recognize the glycan epitopes [1]. Conventionally, they are membrane receptors, although they can also be secreted, able to bind PAMPs. They have been classified into 17 subgroups,

based on their phylogeny and domain organization [40]. These groups differ on structural aspects (Figure 1.8) and features linked to the CTLD (C-type lectins domain) organization, as well as in the cell location.

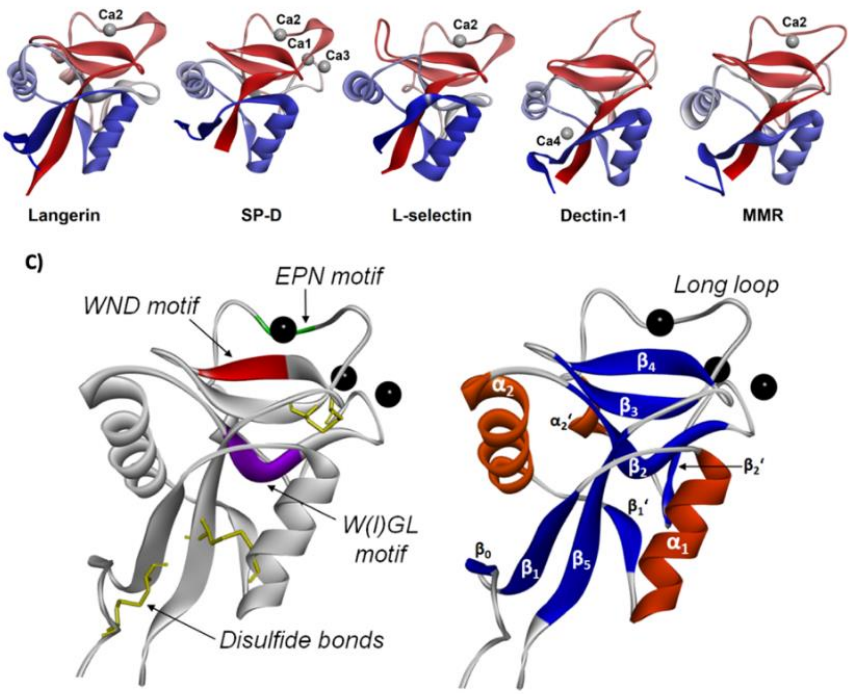


Figure 1.8: *Characteristic structural features of C-type lectins. A) Cartoon representation of five representative CLRs belonging to five different C-type lectin groups (II, III, IV, V and VI). B) Structural comparison between the CRDs of the same five lectins: human langerin (PDB 5G6U), human Surfactant protein D (PDB 4E52), human L-selectin (PDB 3CFW), murine dectin-1 (PDB 2CL8) and human macrophage mannose receptor 1 (CRD2, PDB 5XTS). Calcium ions are depicted in each case. C) Common structural motifs present in the CTLD fold (model: DC-SIGN CRD, PDB 1SL5). On the right, the main secondary structure elements, and on the left, typical conserved residues among different CTLDs and species. Calcium ions are shown as black spheres. Taken from [41].*

Many CTLDs act as mediators in the dissemination and survival of infective pathogens, as well as in the development and progression of autoimmune diseases and certain cancer types, involving the recognition of the self-glycans. Thus, many C-type lectins have become targets to fight diseases such as those caused by HIV, *Mycobacterium tuberculosis*, Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), EBOV, cancer or diabetes. Many of them involve CLRs from antigen presenting cells, including macrophages and

dendritic cells (DC), which are essential players in innate immunity and subsequent guiding of the adaptive response.

Among them, DC-SIGN, the mannose receptor (MR), Langerin, and L-SIGN/DC-SIGNR, display specificity *versus* Man- and Fuc-terminated glycans and recognize exogen epitopes (Figure 1.9). Alternatively, the macrophage galactose lectin (MGL) and L-sectin interact with GalNAc and GlcNAc-terminated glycan structures, thus recognizing self-antigens [41][42].

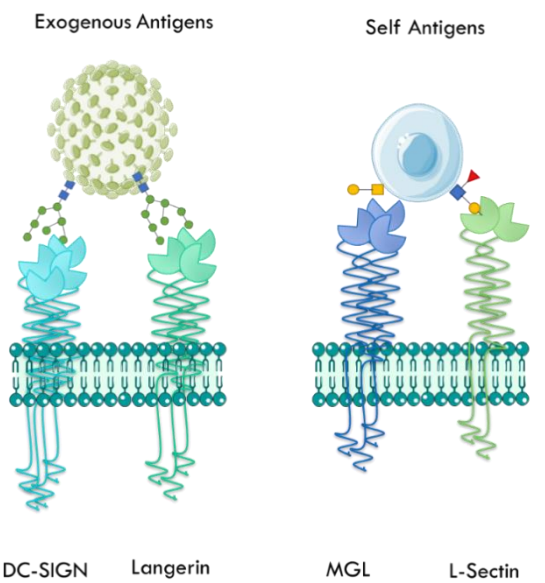


Figure 1.9: *Example of C-type lectins (CTL); CTL are important in the recognition of self-and non self-antigen. For example, some of them interact with the glycan epitopes of pathogens, but also with glycans that are generally expressed on the host cells.*

CTLs are expressed in cells presenting antigens (APC), such as DC and macrophages and they are involved in the self/non-self-recognition process. A key example is DC-SIGN, which can bind many pathogens' glycan-associated epitopes [43]. Moreover, it is also known that CTL can interact with viruses, and that this binding event improves the viral infection. For example, DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-

integrin) is involved in the cis-trans infection of HIV [44], EBOV [44], and Zika [45] viruses. Moreover, a recent study has proposed that DC-SIGN enhances the viral infection of SARS-CoV-2 and that this event can be inhibited by using a glycomimetic molecule, selective for DC-SIGN [46].

1.2.3 I-type lectins (Siglecs)

The sialic acid-binding immunoglobulin (Ig)-like lectins (Siglecs) family in humans is composed of 15 members. Generally speaking, Siglecs are expressed in immune cells [47] [48], although there are exceptions as the myelin-associated glycoprotein (MAG) (Siglec-4), which is expressed on oligodendrocytes and Schwann cells and Siglec-6, on placental trophoblasts. Based on the sequence conservation and evolution, Siglecs are divided in two subgroups: (i) classic Siglecs (including Sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG and Siglec-15); (ii) CD33-related Siglecs (CD33 (Siglec-3), Siglecs-5 to -14 and Siglec -16) [49]. Recognition of their sialylated ligands by the N-terminal variable (V)-Ig-like domain located at the extracellular domain (ECD) triggers cell signalling through their regulatory motifs in their cytoplasmic domains (Figure 1.10) [49].

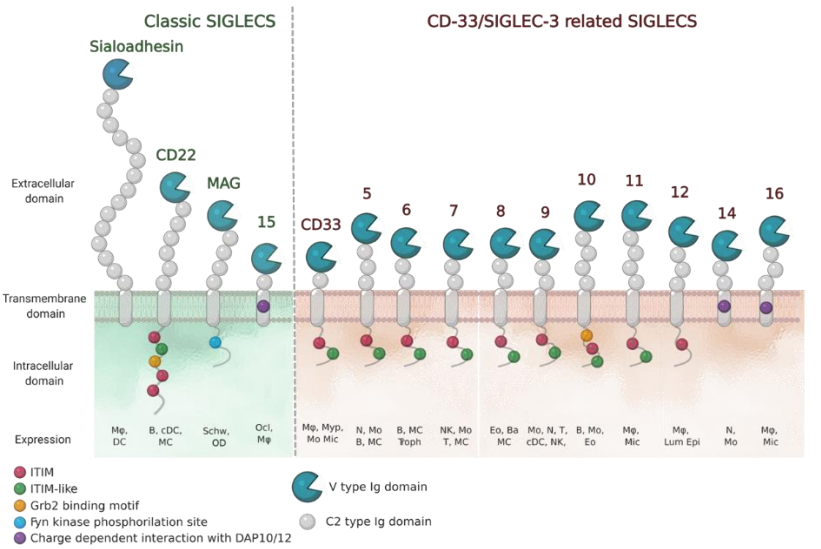


Figure 1.10: *Schematic representation of the human Siglec receptors. Siglecs contain one N-terminal V-type Ig-like domain that mediates sialic-acid recognition and a varying number of constant (C)-type Ig-like extracellular domains. For most Siglecs the intracellular portion contains immunoreceptor tyrosine-bases inhibitory motifs (ITIMs), which serve to recruit phosphatases. In the case of Siglecs-14, -15 and -16, the regulatory domains are immunoreceptor tyrosine-based activatory motifs (ITAM). Siglec-12 in humans has lost the ability to bind sialic acids. The cell-expression patterns are shown (Mø, macrophages; DC, dendritic cell; B, B cells; MC, mast cells; Schw, Schwann cells; OD, oligodendrocytes; Ocl, osteoclasts; Myp, myeloid progenitor; Mo, monocytes; Mic, microglia; N, neutrophils; Troph, trophoblasts; NK, natural-killer cells; T, T cells; Eo, eosinophils; Ba, basophils; Lum epi, lumen epithelia cells).*

Even though all Siglecs share a common N-terminal V domain that binds to sialic acid, each member presents an exclusive specificity and preferences profile towards the terminating sialic acid. Sialic acids refer to a family of nine carbon (C1-C9) sugars derived from neuraminic acid (Neu). There are more than fifty forms of naturally occurring sialic acids, all of which are derived from substituting the amine or the hydroxyl groups. From all of them, just three are mainly expressed in mammals: N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc) and 2-keto-3-deoxynonic acid (Kdn) (Figure 1.11). However, only Neu5Ac is present in humans, since a deletion occurred in the cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMAH) enzyme gene that is responsible for converting Neu5Ac into Neu5Gc [50][51]. Some natural sialic acids bear an O-acetylation in the C9 position, which has a

strong negative effect in most receptors, such as human CD22 and mouse Siglec-1 [52][53]. Regarding the C5 position of Neu5Ac, some Siglecs show different preferences toward the type of N-acyl group at that position. As an example, human and murine Sialoadhesins strongly prefer Neu5Ac over Neu5Gc; nevertheless, murine CD22 accommodates Neu5Gc better than Neu5Ac, while the human orthologue recognizes both of them [54][55].

Sialic acids can be linked to the underlying sugars by different linkages, in most of the cases by α2-3 and α2-6 type linkage to the galactose and by α2-8 to another Neu5Ac (Figure 1.11). In short, by summing up all forms of sialic acids, the type of linkage to the subterminal sugar, the structure of the rest of the oligosaccharide and other possible post-translational modifications (such as sulfation or N-acetylation), there are plenty of potential patterns that can be recognized with variable specificities by the Siglec receptors, which will trigger a biological response accordingly.

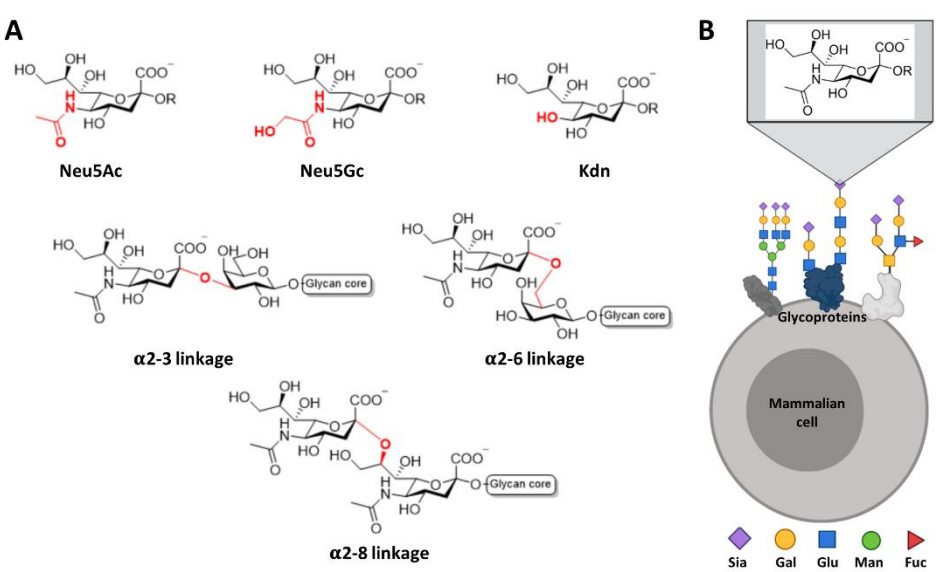
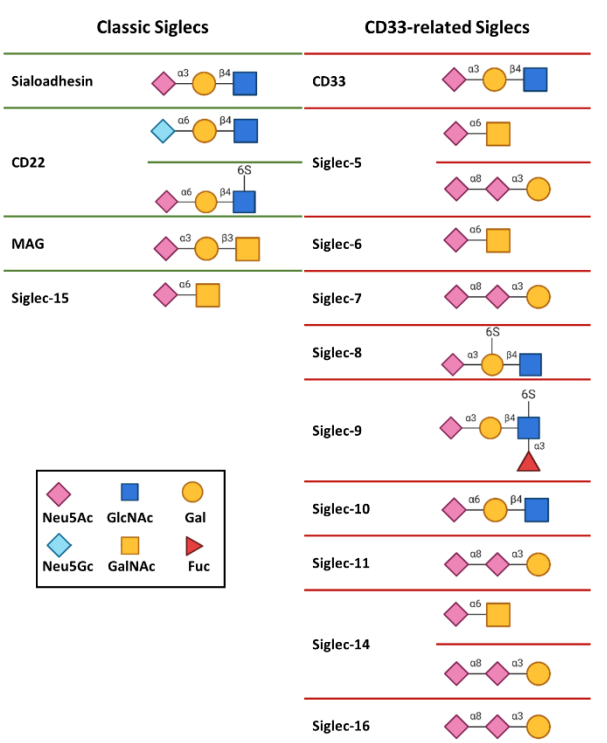


Figure 1.11: *Most common sialic acids in mammals. A) Chemical representation of the most common type of sialic acids in mammals and their linkage to the subterminal glycan. B) Sialic acids are found at the outer most exposed non-reducing end of glycan chains on glycoproteins or glycolipids on the cell surface.*

The binding affinities of Siglecs for isolated Neu5Acα2-6Gal and Neu5Acα2-3Gal moieties are rather low, with dissociation constants ranging from 0.1 mM to 3 mM. Despite the low binding affinity, each Siglec shows a unique specificity profile. For instance, receptor CD22 presents a strong preference for α2-6 linked sialosides, like Neu5Acα2-6Gal and Neu5Gcα2-6Gal [54][56], while Sialoadhesin (Table 1.1) leans towards α2-3 linkages [57]. On the other hand (Table 1.1), Siglec-7 and Siglec-11 have marked selectivity for the Neu5Acα2-8Neu5Ac structure [58][59]. The relative position of the sulfate group regarding the same sialic acid can be also a determining specificity factor. Such is the case of Siglec-8 and Siglec-9, both of which prefer Neu5Acα2-3Galβ1-4GlcNAc as ligand. However, for Siglec-8, the sulfate group at the Gal residue shows improved affinity (Table 1.1), while Siglec-9 is more prone to bind ligand with sulfate at the glucose (Glc) moiety [60]–[62].

Table 1.1: *Glycan binding specificities of human Siglecs.*



The available structural information on Siglecs by either X-Ray crystallography or NMR spectroscopy is currently limited to the extracellular domain (ECD) of the receptor. The ECDs of Siglecs contain one unique V-type Ig like domain, followed by a varying number of constant (C-) type of Ig-like domains. Structurally, Ig-like domains are composed of 70-110 amino acids, which are defined by two opposing β-sheets connected by disulfide bridges (Figure 1.12). All structures showed that the most N-terminal V-type Ig-like domains are composed by two β-sheets, A(A´)BED and C(C´)FG(G´) linked with one intradomain disulfide bridge (Figure 1.12).

The analysis of the 3D structures of V-domain of Siglecs in complex with di/trisaccharides containing sialic acids has provided important clues about the recognition mode and specificity for sialic acids [63]–[69]. For most Siglecs, productive interactions with the sialoglycans are limited to the sialic acid and the adjacent Gal residues, while additional secondary binding sites have not (yet) been identified. The sialic acid binding pocket is formed by strands F and G and loops C-C´ and C´-D with a key conserved Arg residue, essential for forming the salt bridge with the negatively charged carboxyl group C1 of sialic acid. Mutation of this Arg residue causes a drastic decrease in the binding capacity of all studied Siglecs, being the mutation to the positively charged Lys less detrimental for the recognition than that to Ala. A conserved aromatic amino acid (usually a Trp) is present in all Siglec, which interacts with the glycerol side chain of the sialic acid.

Based on the available structural data, we know that the differences in loops C–C′ and strand G at the ligand-binding pocket are determinant for glycan specificity (Figure 1.12). The sequence variability and the conformation adopted by C-C´ loop dictates specificity for the glycan linkage and Gal moiety. Interestingly, the tip of the C-C´ loop in CD22 displays one extra β-hairpin (C1/C2) with the Tyr64, which is optimally preconfigured to extensively interact with branches of N-glycans with α2-6 linkages [69]. In Siglec-8, the

edge of the CC′ loop contains the Arg56 and Gln59 side chains, to form a salt bridge and hydrogen bond with the sulfated Gal6S moiety, respectively [68]. Interestingly the binding pocket in Sialoadhesin, CD22 and Siglec-8 are preformed to accommodate the ligand (Figure 1.12B). On the contrary, CD33, MAG and Siglecs-7 undergo a conformational rearrangement in the C-C´ loop upon ligand binding. Except for CD22, the G strand has a loop of different length inserted (Figure 1.12). Remarkably, the GG′ loop of Siglec-8 consists of eleven residues, which is substantially long compared with the typically five residues of most Siglecs. In Siglec-8, the long and flexible side chains of Lys120 and Gln122 on the GG′ loop interact with the GlcNAc moiety of the ligand 6´S sLex (Neu5Acα2–3[6S]Galβ1–4[Fucα1–3]GlcNAc) [68].

The 3D structures of C-type Ig domains at the ECD of Siglecs (Figure 1.12) can adopt either C1 (formed by strands ABED and CFG) or C2 topology (containing ABE and C´CFG strands). The C1 or C2 Ig-like domain topology, along with differences in the length of the interdomain linkers, remarkably can change the interface between Ig domains and thus might affect the flexibility of the ECD. As shown by the crystal structure of MAG [65] and the 3D reconstruction from negative-stain EM of CD22 [69], the ECD adopts a semi-rigid rod like structure that helps in projecting the ligand binding pocket at V-domain away from the cell surface (around 190 and 300 Å, respectively). Such conformation could be beneficial in exchanging binding with flexible cis (on the same cell surface) and trans (on interacting cells or molecules) ligands on the surface of the cells [65][69].

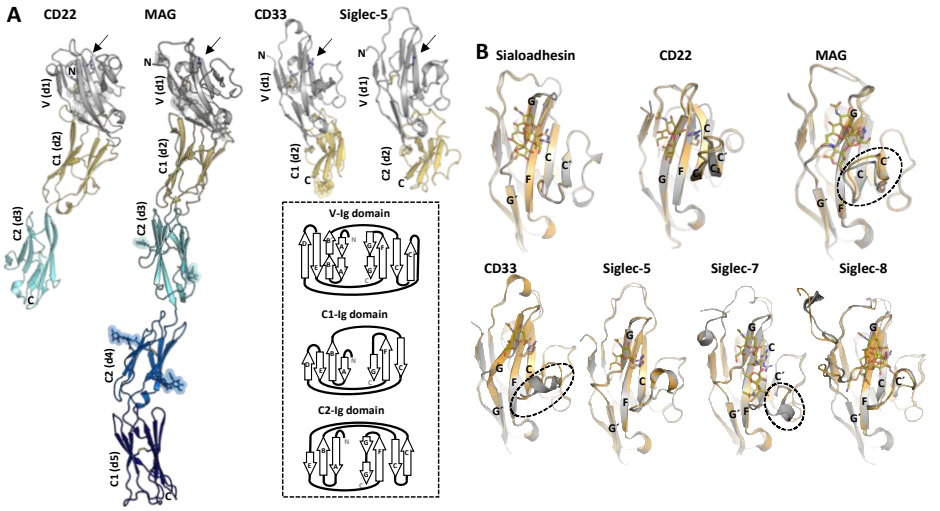


***Figure 1.12**: Three dimensional structures of Siglecs. A) The crystal structures of CD22d1-d3 (PDB ID: 5VKJ), MAGd1-d5 (PDB ID: 5LFU), CD33d1-d2 (PDB ID: 5IHB) and Siglec-5d1-d2 (PDB ID: 2ZG2) in cartoon representation. Domain d1 (in grey) adopts de V-type Ig-like domain and contains the sialic acid binding pocket (indicated with an arrow) with the conserved Arg (in sticks). The N-linked glycans are represented with sticks and spheres. The disulfide bonds are also depicted with sticks. The secondary structure differences between the V- (strands A(A´)B(B´)ED and CC´FG(G´)), C1- (strands ABED and CFG) and C2- (ABE and C´CFG strands) type Ig folding are shown with a diagram (inside the box). B) Superposition of the unliganded (grey) and liganded (orange) structures of d1 from Sialoadhesin (PDB ID: 1QFP and 1QFO), CD22 (PDB ID: 5VKJ and 5VKM), MAG (PDB ID: 5LFR and 5LF5), CD33 (PDB ID: 5IHB and 5J06), Siglec-5 (PDB ID: 2ZG2 and 2ZG3), Siglec-7 (PDB ID: 1O7S and 2HRL) and Siglec-8 (PDB ID: 2N7A and 2N7B).*

Generally, Siglecs regulate the immune response, based on their cellular expression pattern. Moreover, they also play a key role in distinguishing self *versus* non-self recognition. For instance, Sialoadhesin (Siglec-1), on the surface of DC, interact with distinct envelope viruses, in order to allow the infection to proceed [70][71]. Alternatively, the EBOV enters into DC, also through Sialoadhesin, by recognizing the sialylated gangliosides on the viral envelope. The use of a specific block antibody can avoid recognition, thus halting the Ebola infection [71].

Since Siglecs play essential roles in regulating the immune response, these receptors have become important therapeutic targets [72][73][74]. Any therapy targeting Siglecs should exploit their ability to activate or inhibit the target cells and to alter their fate. Additionally, the restricted expression on specific cells can be an advantage for targeted therapies. Siglec-8, for example, has garnered

attention as a target for the treatment of asthma and allergies because of its restricted expression on eosinophils (EO) and mast cells (MC) [75][76] [77][78]. There are numerous strategies to target Siglecs that exploit the characteristics just mentioned. The dominant strategy to target Siglecs is to use monoclonal Abs (mAbs). But there are alternative therapies, standing out the development of chemically modified glycans.

1.2.4 Targeting Siglec-8 as potential modulation of allergic response

Human Siglec-8 consists of an ECD with a unique N-terminal V-type immunoglobulin (Ig)-like domain for binding to sialylated glycans, followed by two C2-type Ig-like domains, a transmembrane domain, and immunoreceptor tyrosine-based inhibitory motifs (ITIMs) localized in the intracellular tail (Figure 1.13). Siglec-8 is a glycoprotein with three N glycosylation sites in the extracellular domain.

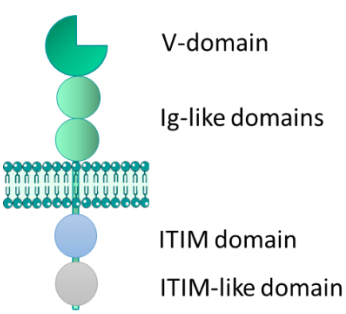


Figure 1.13: *Schematic representation of Siglec-8 structure. The ECD is formed by a V-domain, two Ig-like domains, a single transmembrane domain and an intracellular domain, with one ITIM and one ITIM-like domains.*

Based on the available NMR structure [79], the V-Ig domain comprises two antiparallel β-sheets (β-strands ABED and C′CFG) connected by one intra-sheet disulfide linkage. This binding site (Figure 1.14) contains a unique structural trait in the G-G' β-sheets and the C-C' loop that give the specificity for 6′-sulfo

sialyl Lewis X (6′S sLex) (Neu5Acα2-3[6S]Gal β1-4[Fucα1–3][6S]GlcNAc) [79].

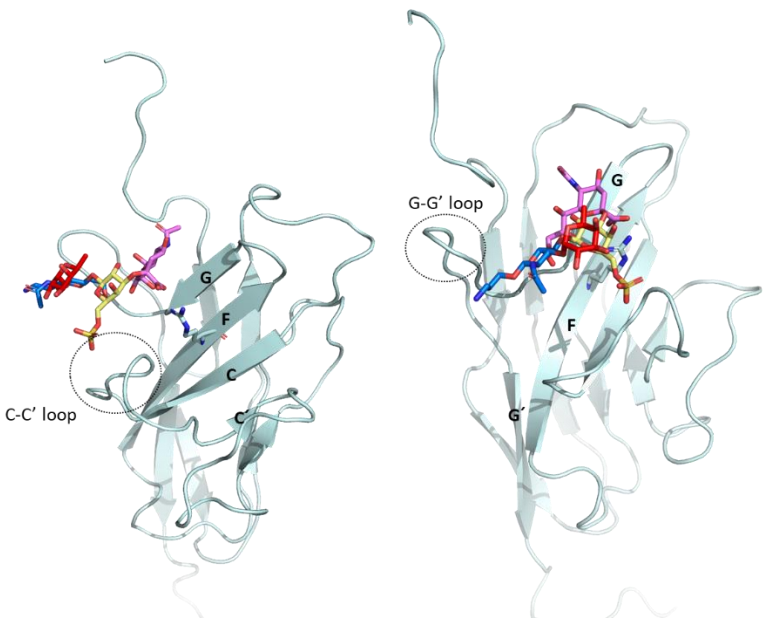


Figure 1.14: *Two perspectives of the interaction of Siglec-8 V domain with its ligand (PDB: 2N7B) according to the NMR studies by Propster et al. [68]. The edge of the CC' loop, in front of the binding site, is directly involved in the binding with the ligand. The GG' loop of Siglec-8 consists of eleven residues. The long and flexible sidechains on the GG' loop interact with the ligand 6'S sLex.*

.

Siglec-8 is expressed on the surface of mast cells, eosinophils, and in very low levels in basophils and displays a critical role in allergy. Generally, the allergic response consists of releasing mediators, such as cytokines and chemokines, by activation of MC and EO, upon the cross-link between the IgE and the FcεRI (Figure 1.15A) [80]. These mediators induce and regulate the acute or chronic inflammation response [81]. FcεRI is the high affinity receptor for the IgE and it is composed by three different subunits (Figure 1.15): the alpha subunit, the beta subunit and two gamma subunits linked by a disulphide bond. The alpha subunit directly interacts with the constan portion of the IgE by the extracellular domain formed by two glycosylated Ig-like domains, while the beta and gamma subunits are responsible of the activation of the signals cascade that leads in the

degranulation of mast-cells and eosinophils [82]. Abnormal activation of this allergic response can lead to allergic and inflammatory diseases [83]. Thus, to avoid an abnormal immune response, there are expressed receptors (as Siglec-8) on the surface of the MC and EO that can inhibit the histamine release. The activation of Siglec-8 induces phosphorylation of its ITIMs, which leads to the inhibition of the IgE-FcεRI-mediated inflammatory mediator release on mast cells, while induces cell apoptosis in eosinophils [84]. These features strongly suggest the key immuno-inhibition role of Siglec-8 under inflammatory conditions (Figure 1.15B).

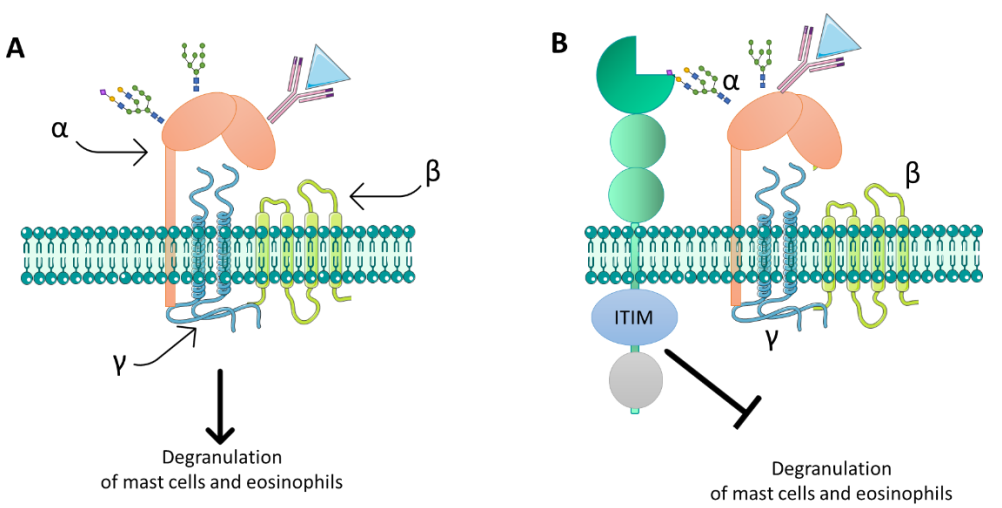


Figure 1.15: *Schematic view of the role of FcεRI in activation of mast cells. The alpha subunit is represented in orange, the beta subunit in green and the gamma in blue.The constant portion of IgE interacts with the alpha subunit of FcεRI. This binding promotes the activation of the mast cell degranulation. On the membrane of mast cells, Siglec- can interact with FcεRI, interrupting the degranulation process through its ITIM domain.*

Therefore, Siglec-8 has been proposed as a therapeutic target for treating allergic and inflammatory diseases. There are several approaches to target a specific protein. The dominant strategy to target Siglecs is to use monoclonal Abs (mAbs). Currently, the humanized non-fucosylated anti-Siglec-8 Ab

AK002 (from Allakos *Inc.*), is undergoing clinical trials to treat allergic and proliferative diseases that affect mast cells and eosinophils (Figure 1.16A).

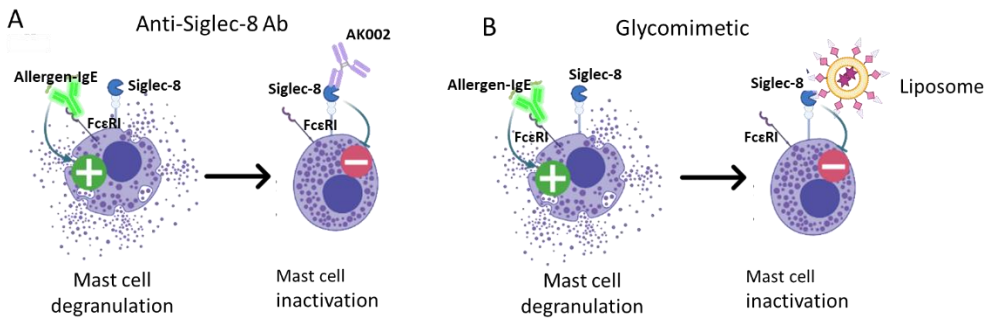


Figure 1.16: *Molecules targeting Siglec-8. A) The AK002 anti Siglec-8-Ab depletes eosinophil activation and mast cell degranulation upon binding. B) liposomes decorated with glycomimetics against Siglec-8 inhibit mast cells and eosinophils activation without depletion.*

The specific binding between AK002 and the ECD of Siglec-8 has been demonstrated, with an affinity of ca. 500 pM [83]. In particular, clinical trials have pointed out that AK002 activates the antibody-dependent cell-mediated cytotoxicity (ADCC) activity against eosinophils in patients with active eosinophilic gastritis or eosinophilic duodenitis [85]. It is also capable of inhibiting the mast cells activity (Figure 1.16) in patients with chronic urticaria, refractory to the conventional therapies [86].

Anti-Siglec monoclonal Abs have emerged to modulate Siglec-sialic acid signalling. In general, the mechanism of action consists in mediating cell depletion on the targeted cell or blocking Siglec-sialic acid interactions [49]. In the case of Siglec-8 and AK002, this antibody can display both actions, based on cell target. The structure of the Siglec-8-AK002 complex remains still unkown and this missing information can help to better understand the mechanism of action of this antibody. Despite the high specificity of the antibody-based therapies, this approach can have some functional limitations such as inadequate tissue accessibility and pharmacokinetics, apart from harmful interactions with the immune system that can cause serious side effects.

Therefore, glycomimetics based on chemically-modified sialic acid moieties have been alternatively proposed to modulate the activity of specific Siglecs.

Chemically modified glycans are drug-like compounds that mimic the structure and function of native glycans, but impart improved affinities, bioavailability and longer serum half-lives [87]. Siglec targeting modified glycans are based on synthetically modified sialic acid scaffolds. These compounds need increased potency for binding at the binding groove masked by the endogenous cis glycans on the target cells. Except for the carboxylic C1 position, which is essential for binding to Siglecs, the rest of the scaffold ranging from C2 to C9 can be potentially modified (Figure 1.17). The first pioneers developing new class of high affinity sialic acid analogues were Kelm et al., with the purpose of addressing the role of the ligand binding domain of CD22 [88]. Most of the variables introduced at positions C5 and C9 on Neu5Ac had negative effect on the binding. However, some substituents such as an -NH2 group in C9 (9-NH2-Neu5Ac/Me) and a fluoroacetate group at C5 (Neu5FAc/Me) enhanced the affinity considerably. The improvements were due to the extra hydrogen bonding and lipophilic interactions between the synthetic ligand and CD22. These observations opened the door to the design and synthesis of new unnatural glycans against CD22 [89]–[91].

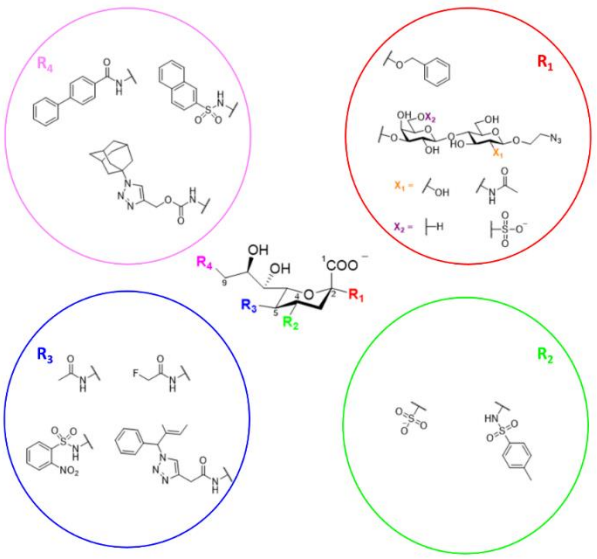


Figure 1.17: *Structure of the sialic acid molecular platform and examples of diverse chemical substituents (R_1-R_4) used to generate specific and high-affinity modified glycans against Siglecs.*

Recently, a potent mimetic (20-fold affinity increase) of the Siglec-8 preferable ligand (6´-sulfo-sialyl LewisX), which is modified with carbocyclic Gal moiety and sulfonamide in the C9 position of Neu5Ac, has been reported [92]. The mentioned modified sialic acids were designed by using classical synthetic structure activity relationship (SAR) studies, which are rather slow and time consuming. In this context, a new high-throughput strategy has been developed by using the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction, which eases the synthesis and screening of sialic acid analogue libraries. The CuAAC reaction has allowed to introduce a 1,2,3-triazole scaffold plus any substituent at the desired position [93]. The analogues are printed as a microarray on glass slides, where can be tested with fluorescently labelled recombinant Siglecs fused to fragment crystallizable (Fc) region of Ab. Using such approach Rillahan et al. identified ligands for CD33 and Siglecs-5, -7, -9 and -10 in the absence of structural information for the majority of the family members [93]. Even though no IC50 values were measured, the approach allowed comparing the relative affinities towards different Siglec members.

Important information about selectivity was derived, indicating that the most potent ligands were not necessarily the most selective ones, such as for Siglec-5, where its most potent ligand also presented high affinity for Siglec-9.

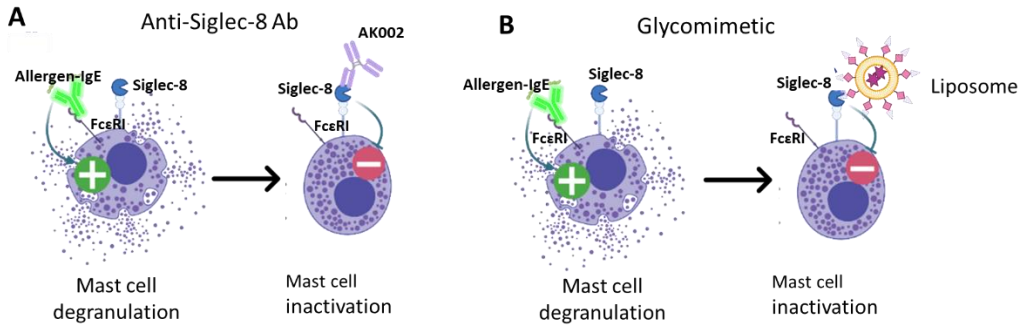


Figure 1.18: *Molecules targeting Siglec-8. A) Anti Siglec-8-Ab, such as AK002, depletes eosinophil activation and mast cell degranulation upon binding. B) liposomes decorated with glycomimetics against Siglec-8 inhibit mast cells and eosinophils activation without depletion.*

For Siglec-8, the research group of Prof. James Paulson at Scripps has identified a sulphonamide-based glycan analogue ligand (Figure 1.18B) that can specifically interact with Siglec-8 [94]. In particular, they synthetized a library of 156 sulfonamidesialoside analogues from two scaffolds, Neu5Acα2−3Galβ1−4GlcNAc, the basic framework, and the 6′-O-sulfo Neu5Acα2−3Galβ1−4GlcNAc analogue, which is selectively recognized by Siglec-8. Moreover, a panel of different substitutions in Neu5Ac C9 was synthesised and, based on the results of the sulfonamide analogue array screened against recombinant Siglec-8, the Neu5Ac modified at C-9 with a 2-naphthyl sulfonyl group (^{NSA}Neu5Ac, NSA = 2-naphthylsulfonamide) was selected as the best ligand. Using in vitro and in vivo mice models, it was demonstrated that liposomes decorated with this glycomimetic (Figure 1.18B) can capture eosinophils expressing Siglec-8 without causing their depletion [94]. Furthermore, the liposomes were also tested on a transgenic mouse line expressing Siglec-8 on mast cells. The results showed that these liposomes

could recruit Siglec-8 to the IgE-FcεRI complex and suppressed activation and desensitized mast cells antigen-induced mast cell degranulation [95].

1.3 Unravelling lectin-glycan interactions: techniques and methodologies.

Glycans directly bind to the carbohydrate recognition domain (CRD) of a lectin, which are usually formed by shallow pockets on the protein surface. Most of the times, the binding process is accompanied by an entropy penalty due to restriction of the conformations of partners as well as by desolvation effects. Nevertheless, the binding event proceeds due to the enthalpically driven interactions generated between the glycan and the lectin [96]. These attractive interactions are those found for any intermolecular ligand-receptor process, with a combination of polar and hydrophobic interactions: solvation effects, hydrogen bonds (HB), coulombic forces, van der Waals interactions, especially CH–π stacking between CH bonds of the sugar and aromatic rings of the protein. For some lectins (C-type lectins), the presence of calcium ions at the protein binding site is also essential [96]. The establishment of hydrogen bonds between sugars and their receptors is important to orient the sugar in the lectin binding site, given the numerous –OH groups in all saccharides. Moreover, additional hydrogen bonds can be mediated by the amine, carbonyl and carboxyl groups present in many sugars (Figure 1.19).

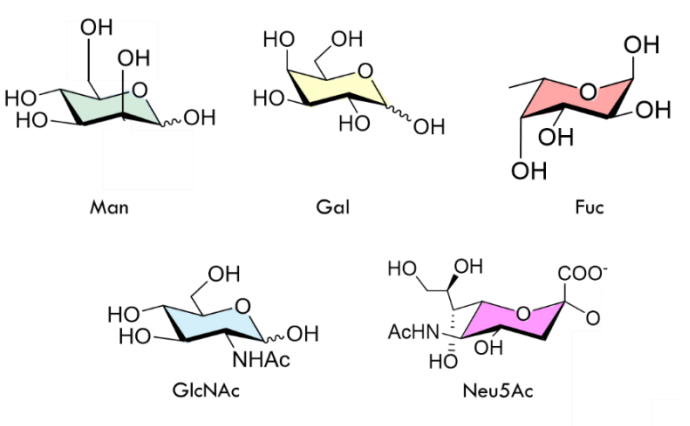


Figure 1.19: *Structure of the common monosaccharides found in glycans, in the pyranose form.*

The hydroxyl groups can act as HB donors or acceptors. Frequently, the same OH can be donor and acceptor at the same time, providing a cooperative HB process [97][98]. Obviously, the sugars establish HB interactions (Figure 1.20A) with the side chain of the polar amino acids, such as aspartic and glutamic acid, asparagine, glutamine, arginine, and serine, using both the side chains and the backbone amide and carbonyl groups [99].

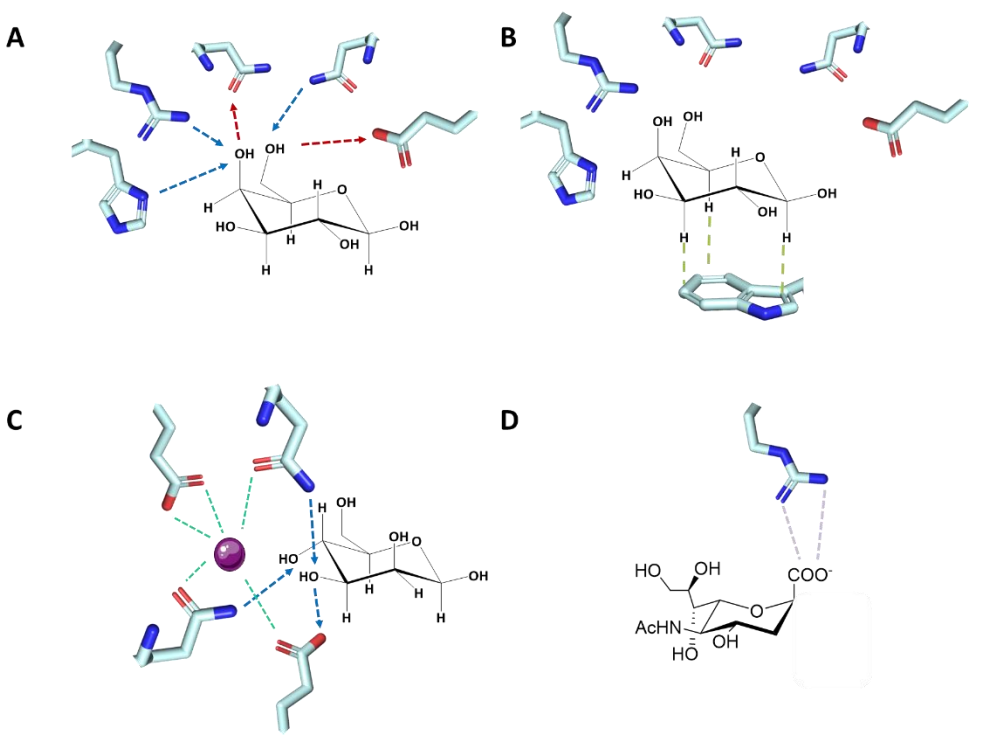


Figure 1.20: *Schematic representation of: A) typical HB between the -OH groups at positions 4 and 6 of a Gal unit and diverse amino acid side chains, B) a typical CH-π interaction between the -CH groups of a Gal unit and the indol moiety of a Trp side chain, C) the involvement of a Ca2+ ion in the molecular recognition of Man by a C-type lectin and D) typical salt bridge between the -COO- of a Neu5Ac and the guanidinium group of Arg side chain*

Additionally, non-polar interactions are also essential and provide the impetus from the enthalpy perspective. Specifically, CH-π stacking interactions between the non-polar faces of certain sugars with adjacent aromatic rings are instrumental to provide the required stabilization. To create the geometry

condition for effective CH-π stacking, at least three CH vectors on the pyranose ring need to be properly oriented towards the same spatial direction (Figure 1.20B) to interact with the aromatic ring [100][101][102].

The role of water molecules should also be emphasised. Besides their role in modulating the entropy, they can also form HBs with the two interacting actors, acting as a "bridge" between the ligand and the receptor. The existence of these cooperative HBs enhance the affinity [103][104].

As mentioned above, CTLs need Ca^{2+} as a key coordinating partner between the lectin and the glycan epitope (Figure 1.20C). The positive charge of the ion manages the interaction of two sugar hydroxyls with certain amino acids in the binding site [105], and the complex is further stabilized by intermolecular HBs.

Coulombic forces are also involved in glycan recognition by protein receptors. Electrostatic interactions between positively charged amino acid residues (Arg, Lys) and saccharide moieties (sialic acid, iduronic acid, glucuronic acid) are instrumental to stabilize the complexes in which these sugars are involved. Related to this Thesis, in the Siglec family, a critically conserved Arg residue interacts (Figure 1.20D) with the sialic acid of the glycans by forming a salt bridge between the guanidinium group and the negatively charged carboxyl group C1 [49].

The understanding of these structural details at the atomic level is essential to design molecules for therapeutics purposes. Different approaches may be employed to decipher these molecular recognition processes with high resolution. Advances in cryo-electron microscopy (cryoEM) are providing exquisite details on mechanisms involving sugars [106], while better and better protocols for structural refinement in the use of X-ray methods for protein-sugar complexes and glycoproteins are also allowing extraordinary advances in the glycoscience arena [1][107]. Alternatively, NMR spectroscopy remains as one of the most rewarding techniques to explore sugar-lectin interactions. In fact,

given the intrinsic dynamic nature of saccharides, NMR affords exquisite structural information at the atomic detail, not accessible by other techniques [101][108][109].

At the macroscopic level, high-throughput microarrays [73][110] can be used for the identification of carbohydrates epitopes to lectins. Moreover, binding affinities can be calculated employing biolayer interferometry (BLI), surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC).

Herein, this Thesis is mainly focused on obtaining high resolution details of the interactions. Therefore, NMR and X-Ray Crystallographic methods have been employed.

1.3.1 X-Ray Crystallography

Since its first application to date, X-ray crystallography has provided more than 156,000 protein structures, which are deposited of proteins in the Protein Data Bank (PDB).

A protein crystal is a three-dimensional array of atoms of identical molecules organized regularly and repeatedly. Each building block of the crystal is named the unit cell. The shape and the size of the unit cell can be described by three axes (a, b, and c) and angles between the axes [111]. When an X-ray impacts an atom, it suffers a scatter in its electrons. Either constructive or destructive wave interferences occur along the different direction as the scattered waves (diffraction pattern) are emitted by atoms at different positions. When atoms have an ordered arrangement, they can emit constructive wave interferences that are directly correlated to atomic structures. The diffraction pattern (Figure 1.21) of the crystal results from all the X-ray reflections by all the atoms in the crystal and represents the electron density maps [111].

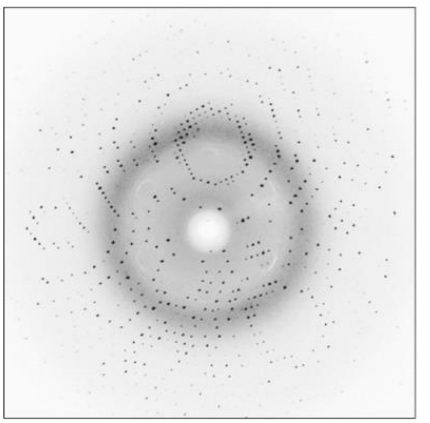


Figure 1.21: *Example of diffraction pattern. Image adapted from Protein Crystallography and Drug Discovery* [112].

Three parameters define each diffraction pattern: i) the amplitude, which measures the spot intensity, ii) the wavelength, which is determined by the X-ray source, and iii) the phase, which is lost during the acquisition. These three parameters must be known for each spot, in order to obtain the crystal structure. The phase is the only parameter that cannot be directly calculated; however, there are indirect methodologies for the phase calculation. Additionally, it is necessary to carry out a Fourier transformation for the conversion of the waves in an atomic map [112].

From the technical perspective, the search of the factors that influence the formation of a given crystal is a major task. Thus, a variety of experimental conditions, by using a high-throughput screening process, are optimized to obtain the best possible crystal. The crystallization of proteins requires generating a supersaturated solution in which the macromolecule is integrated, always employing conditions that do not significantly perturb its natural state [113].

In this context, vapor diffusion methods to generate the required supersaturated conditions: i) the hanging drop and ii) the sitting drop (Figure 1.22A). In both cases, the purified protein is mixed with a buffer that contains crystallizing agents and additives to form a drop. In particular, the mixture is incubated in a

closed system, with a reservoir that contains the crystallizing solution at higher concentration than in the drop. In this close cycle, an equilibrium process is established between the contents of the drop and the reservoir. The primary role of the reservoir solution is to slowly concentrate the crystallization drop under controlled conditions, using the vapor diffusion from the volatile species (water or solvent) [114]. Then, the crystallization of the protein starts with the nucleation step, in which the partially ordered protein starts to assemble in a regular manner [115]. When the nucleus is formed, the crystal starts to grow (Figure 1.22B).

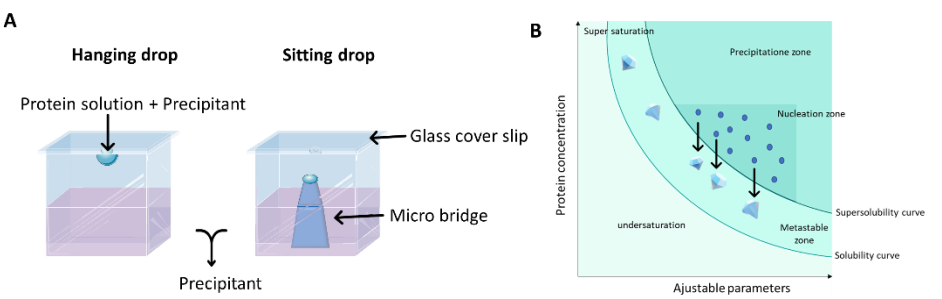


Figure 1.22: *A) the hanging-drop and the sitting-drop vapor-diffusion methods are illustrated in this schematic figure. B) Phase Diagram illustrating the different crystallization steps. The super-solubility curve separates the conditions under which spontaneous nucleation occurs and the metastable zone, ideal for crystals growth. Adapted from* [116]

1.3.1.1 X-Ray crystallography to determine the atomic details of the lectin-glycan interaction

X-ray crystallography has been extensively employed to study protein/carbohydrate complexes [117]. From the technical perspective, the formation of the carbohydrate-lectin complex can be performed using either soaking or co-crystallization methods.

The soaking process requires the addition of an excess of ligand to the already preformed protein crystal. For the soaking process to be sucessful, the binding site needs to be accessible to the ligand in the crystal. Thus, the crystal packing

should allow the conformational changes that are required to accommodate the ligand in the binding site [118]. Different variables should be considered to optimize soaking, including the ligand concentration, which depends on the affinity, as well as the soaking time [119].

If soaking is not possible, the co-crystallization of the ligand and protein to form the complex is an alternative. In this case, a pre-incubation period at a controlled temperature is needed, and new crystallization screening conditions may be required.' Also, in this case, the ligand concentration depends on the binding affinity [118].

1.3.2 Nuclear Magnetic Resonance

NMR has been extensively usied to monitor glycan conformation, dynamics and interactions [109]. There is a vast collection of NMR methodologies that can be employed to this end. From a physical chemistry perspective, the protein-ligand binding event is a chemical exchange event between the free and the bound states. k_{on} represents the constant association of the complex, while k_{off} is dissociation the dissociation one (Equation 1.1). Depending on the corresponding values, different NMR parameters, obtained through specific NMR experiments can be employed to monitor the process and to extract key structural information.

$$[P] + [L] \underset{K_{off}}{\overset{K_{on}}{\rightleftharpoons}} [PL]$$

Equation 1.1: *The equation represents the equilibrium between species: the free protein, the free ligand, and the protein-ligand complex. kon is the association constant and koff is the dissociation one.*

In particular, the process can be monitored either from the glycan or from the receptor's perspective, monitoring the NMR parameters and their changes of one of the entities in the absence and presence of the partner [101].

1.3.2.1 NMR & Molecular recognition form the ligand's perspective

Figure 1.23 schematizes the chemical exchange process and the associated motional characteristics of the partners that are associated to NMR parameters.

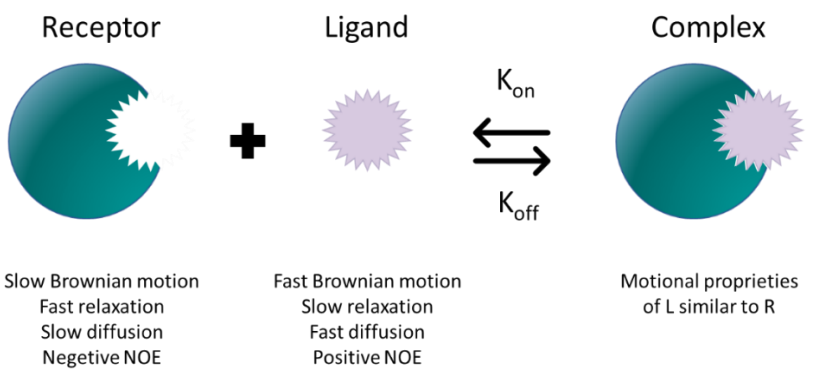


Figure 1.23: *Schematic illustration of the different motional properties that ligands and receptors display in solution and how the process affects them.*

1.3.2.1.1 Relaxation

Small molecules (ligands), display fast rotational motion, in the ps time scale (Figure 1.23). This fast motion is associated to slow NMR relaxation (in the seconds time scale) which, in turn, is translated in the existence of narrow NMR linewidths (below 1 Hz) for these molecules. Large molecules display the opposite features: slow rotational motion (ns time scale), fast relaxation (in the ms timescale), and large linewidths (tens of Hz). A particularly useful NMR approach for studying ligand-protein interactions is to study the relaxation rates of the ligand's resonance signals. When the ligand binds the receptor, its molecular tumbling changes and is more similar to that of the receptor (ns). Then, the bound ligand displays short T2 relaxation times (ms), and the effect

can be visualized in the NMR spectrum, since the signals of the bound ligand resonances are broadened and may even disappear [119].

1.3.2.1.2 Saturation transfer difference (STD-NMR)

The STD NMR experiment is one of the most popular and versatile NMR methods for studying the interaction between ligands and receptors [120] [121]. The STD spectrum represents the difference between two different ^{1}H-NMR spectra performed in a sample that contains the protein-ligand complex, with a large excess of the ligand (usually >50:1 molar ratio). In particular, in the first spectrum, dubbed on-resonance, the protein signals are saturated with a train of selective low power pulses in a spectral region that does not contain any ligand signal (usually around -0.5ppm or the aromatic region). Then, the saturation is kept for a given period (ca. 2 s) and is propagated to all the protein nuclei through a ^{1}H-^{1}H cross-relaxation mechanism, which is very efficient in proteins (spin diffusion). At a given moment, the saturation also reaches the nuclei at the binding site and, provided, that there is a bound ligand, it is also transferred to its nuclei, starting from those protons that are closer to the saturated protein protons. Interestingly, the amount of saturation depends on the proximity of the ligand protons to those at the protein and obviously, there is not any transfer of saturation if there is not any bound ligand. Therefore, the observed intensities in the on-resonance spectrum are lower for those protons that belong to binders and specially for those that are closer to the binding site. To better visualize the effect, a second experiment (off-resonance) is performed using exactly the same conditions, but setting the irradiation frequency at a spectral region without signals of the protein and the ligand (i.e. 100 ppm). Then, this off-resonance spectrum is substracted from the on-resonance one to provide the final STD NMR spectrum. Obviously, if there are no binders in the NMR tube, the intensities of the two experiments are the same and the STD NMR shows no signal at all. For the binders, the signals corresponding to the regions closer to the protein display larger intensities (Figure 1.24).

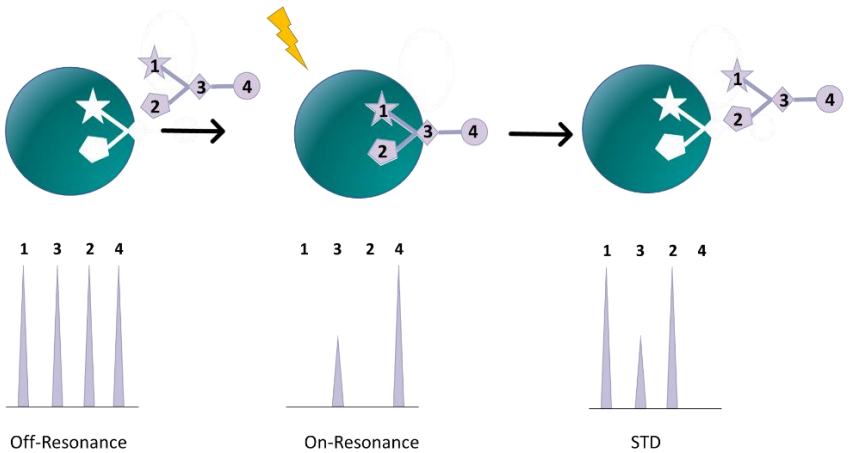


Figure 1.24: *Schematic representation of the STD NMR experiment. The ligand-receptor partners are shown. In the off-resonance experiment, no saturation of the protein signals take place. In the on-resonance spectrum, the saturation of the protein signals is transferred to the bound ligand, especially to those regions (1 and 2) that are closer to the binding site. Therefore, the signals for 1 and 2 are fully saturated. The signals of 3, which is relatively close, are only partially saturated. The STD spectrum, the difference between the off- and the on-resonance ones, only shows the ligand signals that have been affected by the protein saturation. Therefore, the signals of 4, which is far from the binding site, are not present in the STD, since they display the same intensity in the on- and off-resonance spectra.*

Other related experiments that can be used to study protein-ligand molecular recognition events are the WaterLOGSY (Water-Ligand Observed through Gradient Spectroscopy) and the TR-NOESY (transfer- Nuclear Overhauser Effect Spectroscopy). Both are particularly useful for ligands which are not bound tightly and exchange between free and bound state at a reasonably fast rate, faster than the relaxation time. The WaterLOGSY discriminate binders from non-binders, while the transferred nuclear Overhauser enhancement experiment (trNOESY) provides an adequate means to determine the conformation of the bound ligand [122][123].

1.3.2.2 NMR & Molecular recognition form the ligand's perspective

The presence of a ligand in the binding site of a protein modifies the chemical environment of the amino acids directly, or indirectly, involved in the binding event. The essential NMR parameter, the chemical shift, precisely depends on this chemical environment. Therefore, the presence of the ligand should affect

the chemical shifts of the nuclei of the protein, specially to those directly involved in binding, but also to those affected by additional motions or adjustments that can take place when the interaction takes place. This phenomenon is dubbed chemical shift perturbation (CSP). The existence of CSP upon ligand addition is better evaluated using ^{1}H-^{15}N experiments. In particular, ^{1}H,^{15}N-HSQC (Heteronuclear Single-Quantum Coherence) correlation experiments constitute the fingerprint of any protein. [124][125] The cross peaks in this spectrum correspond to all the NH amide groups of the backbone and lateral chains. In certain cases, the NHs of the lateral chains of Lys, Arg, His, and Trp residues may also be observed. The corresponding protein sequence and 3D structure provides a unique chemical environment for each amino acid and therefore, the resulting HSQC spectrum can be considered as its ID, the protein fingerprint. Indeed, the dispersion and specific ^{1}H/^{15}N chemical shifts of the cross peaks are characteristic and unique for each protein. Depending on the size of the protein the ^{1}H-^{15}N correlation spectrum can be obtained using the standard HSQC sequence or the TROSY (Transverse Relaxation-Optimized Spectroscopy) variant, which is specially used for large entities. In this Thesis, both experiments have been employed. In any case, it is required that the protein is ^{15}N-labelled, since the natural abundance of ^{15}N is below 0.5% and the experimental time for natural abundance samples would be prohibitively long.

Essentially, to perform the CSP analysis of a protein/ligand complex, the reference ^{1}H,^{15}N-HSQC spectrum for the apo form is first obtained. Then, sequential ^{1}H,^{15}N-HSQC spectra are recorded upon adding increasing ligand concentrations, ideally until the protein binding site is completely saturated. The presence of the ligand modifies the chemical environment around the nuclei located at the binding pocket and thus, the corresponding peaks display CSP. Depending on the k_{off} of the process, the behaviours of the NMR peaks are different. If the chemical exchange rate between the free and bound states is fast

in the chemical shift timescale, during the titration, the ^{1}H,^{15}N cross-peaks gradually shift from its free position in the free state towards the final position, which corresponds to the fully saturated protein. In contrast, when the exchange rate of the ligand is slow in the chemical shift time scale, two peaks are simultaneously displayed in the HSQC spectrum (the initial one for the free space and a second one for the bound state). Their relative intensities depend on the binding affinity and the added amount of ligand. During the titration, the free state peak gradually decreases, while that for bound from increases (Figure 1.25), until the peak of for the free state peak disappears [126]. The range of protein and ligand concentrations is chosen to cover a wide range of fractional saturations of the lectin by the glycan, ideally to achieve full saturation.

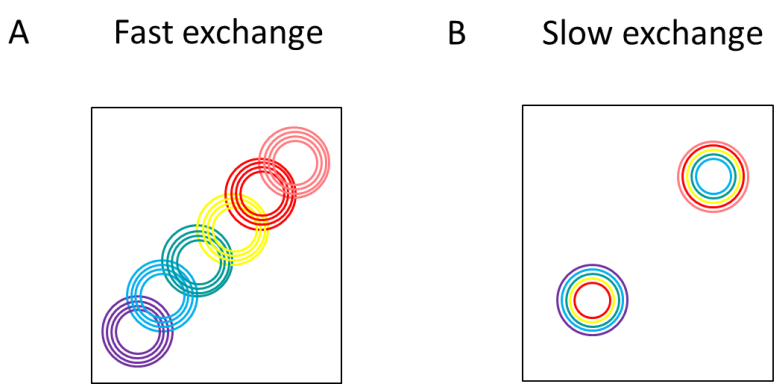


Figure 1.25: *Schematic representation of the effect of the addition of a ligand in the chemical shifts of the ^{15}N-labelled protein cross peaks. A) fast exchange: the addition of increasing amounts of ligand induces a gradual chemical shift perturbation of the ^{15}N and ^{1}H frequencies from the unbound peak to the bound one; B) In the slow exchange regime, the simultaneous presence of the cross peaks for the protein free and bound states are observed.*

Therefore, from the well known equilibrium:

$$[P] + [L] \underset{K_{off}}{\overset{K_{on}}{\rightleftharpoons}} [PL]$$

It is possible to deduce the binding affinities, following these simple algebraic equations, from the total (known) concentrations of ligand and protein:

$$[L]_t = [L] + [PL] \quad \text{and} \quad [P]_t = [P] + [PL]$$

The observed chemical shifts, which depend on those for the free and bound states and the molar fraction of the species

$$\delta_{obs} = \delta_f f_f + \delta_b f_b$$

$$f_b = (\delta_{obs} - \delta_f)/(\delta_b - \delta_f)$$

And then:

$$\Delta\delta_{obs} = \Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{1/2}\right\}\Big/2[P]_t$$

where Δδobs is the change in the observed shift from the free state and Δδmax is the maximum shift change on saturation (deduced as part of the fitting procedure, since it is usually not measurable experimentally). This equation allows to fit K_D from the observed values of the chemical shift at different protein and ligand concentrations. The fitting is easily set up using standard programs.

1.4 References

[1] S. P. H. Varki Ajit , Cummings Richard D , Esko Jeffrey D , Stanley Pamela , Hart Gerald W , Aebi Markus , Darvill Alan G , Kinoshita Taroh , Packer Nicolle H , Prestegard James H , Schnaar Ronald L, *Essential of Glycobiology*, 3th editio. Cold Spring Harbor (NY): Cold Spring Harbor (NY), 2017.

[2] C. R. B. LRyan A. Flynn, Benjamin A. H. Smith, Alex G. Johnson, Kayvon Pedram, Benson M. George, Stacy A. Malaker, Karim Majzoub, Jan E. Carette, "Mammalian Y RNAs are modified at discrete guanosine residues with N-glycans," *bioRxiv Prepr. Serv. Biol.*, 2019, doi: http://dx.doi.org/10.1101/787614.

[3] L. Krasnova and C. H. Wong, *Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis*, vol. 85. 2016.

[4] G. A. Rabinovich, Y. van Kooyk, and B. A. Cobb, "Glycobiology of immune responses," *Ann. N. Y. Acad. Sci.*, vol. 1253, no. 1, pp. 1–15, 2012, doi: 10.1111/j.1749-6632.2012.06492.x.

[5] C. Reily, T. J. Stewart, M. B. Renfrow, and J. Novak, "Glycosylation in health and disease," *Nat. Rev. Nephrol.*, vol. 15, no. 6, pp. 346–366,. 2019, doi: 10.1038/s41581-019-0129-4.

[6] K. T. Schjoldager, Y. Narimatsu, H. J. Joshi, and H. Clausen, "Global view of human protein glycosylation pathways and functions," *Nat. Rev. Mol. Cell Biol.*, vol. 21, no. 12, pp. 729–749. 2020, doi: 10.1038/s41580-020-00294-x.

[7] J. B. Goh, S. K. Ng, J. B. Goh, and S. K. Ng, "Critical Reviews in Biotechnology Impact of host cell line choice on glycan profile Impact of host cell line choice on glycan profile," *Crit. Rev. Biotechnol.*, vol. 38, no. 6, pp. 851–867, 2018, doi: 10.1080/07388551.2017.1416577.

[8] D. Dutta, C. Mandal, and C. Mandal, "Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1861, no. 12, pp. 3096–3108, 2017, doi: 10.1016/j.bbagen.2017.08.025.

[9] P. Nair-Gupta *et al.*, "A novel C2 domain binding CD33xCD3 bispecific antibody with potent T-cell redirection activity against acute myeloid leukemia," *Blood Adv.*, vol. 4, no. 5, pp. 906–919, 2020, doi: 10.1182/bloodadvances.2019001188.

[10] P. Fisher, J. Thomas-Oates, A. J. Wood, and D. Ungar, "The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus," *Front. Cell Dev. Biol.*, vol. 7, pp. 1–11, 2019, doi:

10.3389/fcell.2019.00157.

[11] B. A. Cobb, "The history of IgG glycosylation and where we are now," *Glycobiology*, vol. 30, no. 4, pp. 202–213, 2021, doi: 10.1093/GLYCOB/CWZ065.

[12] Y. Watanabe, T. A. Bowden, I. A. Wilson, and M. Crispin, "Exploitation of glycosylation in enveloped virus pathobiology," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1863, no. 10, pp. 1480–1497. 2019, doi: 10.1016/j.bbagen.2019.05.012.

[13] M. Sevvana, T. Klose, and M. G. Rossmann, "Principles of Virus Structure," in *Encyclopedia of Virology*, vol. 1, Elsevier, 2021, pp. 257–277.

[14] J. F. Cipollo and L. M. Parsons, "Glycomics and glycoproteomics of viruses: Mass spectrometry applications and insights toward structure–function relationships," *Mass Spectrom. Rev.*, vol. 39, no. 4, pp. 371–409, 2020, doi: 10.1002/mas.21629.

[15] L. Kong, I. A. Wilson, and P. D. Kwong, "Crystal structure of a fully glycosylated HIV-1 gp 120 core reveals a stabilizing role for the glycan at Asn262," *Proteins Struct. Funct. Bioinforma.*, vol. 83, no. 3, pp. 590–596, 2015, doi: 10.1002/prot.24747.

[16] G. S. Mohan, W. Li, L. Ye, R. W. Compans, and C. Yang, "Antigenic Subversion: A Novel Mechanism of Host Immune Evasion by Ebola Virus," *PLoS Pathog.*, vol. 8, no. 12, p. e1003065. 2012, doi: 10.1371/journal.ppat.1003065.

[17] A. Shajahan, N. T. Supekar, A. S. Gleinich, and P. Azadi, "Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2," *Glycobiology*, vol. 30, no. 12, pp. 981–988. 2020, doi: 10.1093/glycob/cwaa042.

[18] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281-292.e6, 2020, doi: 10.1016/j.cell.2020.02.058.

[19] T. J. Yang *et al.*, "Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 3, pp. 1438–1446, 2020, doi: 10.1073/pnas.1908898117.

[20] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, and M. Crispin, "Site-specific glycan analysis of the SARS-CoV-2 spike," *Science (80-.).*, vol. 369, no. 6501, pp. 330–333, 2020, doi: 10.1126/science.abb9983.

[21] J. D. Allen, H. Chawla, F. Samsudin, and L. Zuzic, "Site-specific glycosylation steric control of," pp. 1–37, 2021, doi: 10.1101/2021.03.08.433764.

[22] Y. Zhang *et al.*, "Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins using High-Resolution Mass Spectrometry," *bioRxiv*, p. 2020.03.28.013276, 2020, doi: 10.1101/2020.03.28.013276.

[23] I. Bagdonaite *et al.*, "Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells," *Viruses*, vol. 13, no. 4, pp. 1–14, 2021, doi: 10.3390/v13040551.

[24] Y. Watanabe *et al.*, "Vulnerabilities in coronavirus glycan shields despite extensive glycosylation," *Nat. Commun.*, vol. 11, no. 1, p. 2688. 2020, doi: 10.1038/s41467-020-16567-0.

[25] B. Lepenies and R. Lang, "Editorial: Lectins and Their Ligands in Shaping Immune Responses," *Front. Immunol.*, vol. 10, pp. 9–11. 2019, doi: 10.3389/fimmu.2019.02379.

[26] R. L. Schnaar, "Glycans and glycan binding proteins in immune regulation: A concise introduction to glycobiology for the allergist," *J Allergy Clin Immunol.*, vol. 135, no. 3, pp. 609–615, 2015, doi: 10.1016/j.jaci.2014.10.057.

[27] L. Johannes, R. Jacob, and H. Leffler, "Galectins at a glance," *J. Cell Sci.*, vol. 131, no. 9, p. jcs208884, 2018, doi: 10.1242/jcs.208884.

[28] S. Bertuzzi, J. I. Quintana, A. Ardá, A. Gimeno, and J. Jiménez-Barbero, "Targeting Galectins With Glycomimetics," *Front. Chem.*, vol. 8, 2020, doi: 10.3389/fchem.2020.00593.

[29] Y. C. Chan, H. Y. Lin, Z. Tu, Y. H. Kuo, S. T. D. Hsu, and C. H. Lin, "Dissecting the structure–Activity relationship of galectin–Ligand interactions," *Int. J. Mol. Sci.*, vol. 19, no. 2, pp. 1–20, 2018, doi: 10.3390/ijms19020392.

[30] M. F. Brinchmann, D. M. Patel, and M. H. Iversen, "The Role of Galectins as Modulators of Metabolism and Inflammation," *Mediators Inflamm.*, vol. 2018, pp. 1–11, 2018, doi: 10.1155/2018/9186940.

[31] G. R. Vasta, "Roles of galectins in infection," *Nat. Rev. Microbiol.*, vol. 7, no. 6, pp. 424–438, 2009, doi: 10.1038/nrmicro2146.

[32] I. C. Weng *et al.*, "Cytosolic galectin-3 and -8 regulate antibacterial autophagy through differential recognition of host glycans on damaged phagosomes," *Glycobiology*, vol. 28, no. 6, pp. 392–405, 2018, doi: 10.1093/glycob/cwy017.

[33] F. T. Liu, "Regulatory roles of galectins in the immune response," *Int. Arch. Allergy Immunol.*, vol. 136, no. 4, pp. 385–400, 2005, doi: 10.1159/000084545.

[34] M. R. Girotti, M. Salatino, T. Dalotto-Moreno, G. A. Rabinovich, and G. A. Rabinovich, "Sweetening the hallmarks of cancer: Galectins as multifunctional mediators of tumor progression," *J. Exp. Med.*, vol. 217, no. 2, pp. 1–14, 2020, doi: 10.1084/jem_20182041.

[35] R. P. M. Dings *et al.*, "Structure-based optimization of angiostatic agent 6DBF7, an allosteric antagonist of galectin-1," *J. Pharmacol. Exp. Ther.*, vol. 344, no. 3, pp. 589–599, 2013, doi: 10.1124/jpet.112.199646.

[36] R. Hevey, "Bioisosteres of carbohydrate functional groups in glycomimetic design," *Biomimetics*, vol. 4, no. 3, 2019, doi: 10.3390/biomimetics4030053.

[37] O. B. Garner and L. G. Baum, "Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling," *Biochem. Soc. Trans.*, vol. 36, no. 6, pp. 1472–1477, 2008, doi: 10.1042/BST0361472.

[38] G. A. Rabinovich and M. A. Toscano, "Turning 'sweet' on immunity: Galectin-glycan interactions in immune tolerance and inflammation," *Nat. Rev. Immunol.*, vol. 9, no. 5, pp. 338–352, 2009, doi: 10.1038/nri2536.

[39] M. Okamoto, A. Hidaka, M. Toyama, and M. Baba, "Galectin-3 is involved in HIV-1 expression through NF-κB activation and associated with Tat in latently infected cells," *Virus Res.*, vol. 260, no. 2018, pp. 86–93, 2019, doi: 10.1016/j.virusres.2018.11.012.

[40] G. D. Brown, J. A. Willment, and L. Whitehead, "C-type lectins in immunity and homeostasis," *Nat. Rev. Immunol.*, vol. 18, no. 6, pp. 374–389, 2018, doi: 10.1038/s41577-018-0004-8.

[41] P. Valverde, J. D. Martínez, F. J. Cañada, A. Ardá, and J. Jiménez-Barbero, "Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin," *ChemBioChem*, vol. 21, no. 21, pp. 2999–3025. 2020, doi: 10.1002/cbic.202000238.

[42] M. J. Robinson, D. Sancho, E. C. Slack, S. LeibundGut-Landmann, and C. R. Sousa, "Myeloid C-type lectins in innate immunity," *Nat. Immunol.*, vol. 7, no. 12, pp. 1258–1265, 2006, doi: 10.1038/ni1417.

[43] C. del Fresno, S. Iborra, P. Saz-Leal, M. Martínez-López, and D. Sancho, "Flexible signaling of Myeloid C-type lectin receptors in immunity and inflammation," *Front. Immunol.*, vol. 9, pp. 1–13, 2018, doi: 10.3389/fimmu.2018.00804.

[44] C. P. Alvarez, F. Lasala, J. Carrillo, O. Muñiz, A. L. Corbí, and R. Delgado, "C-Type Lectins DC-SIGN and L-SIGN Mediate Cellular Entry by Ebola Virus in cis and in trans," *J. Virol.*, vol. 76, no. 13, pp. 6841–6844, 2002, doi: 10.1128/jvi.76.13.6841-6844.2002.

[45] R. Hamel *et al.*, "Zika virus: epidemiology, clinical features and host-virus interactions," *Microbes Infect.*, vol. 18, no. 7–8, pp. 441–449, 2016, doi: 10.1016/j.micinf.2016.03.009.

[46] M. Thépaut *et al.*, "DC / L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist," pp. 1–34.

[47] M. S. MacAuley, P. R. Crocker, and J. C. Paulson, "Siglec-mediated regulation of immune cell function in disease," *Nature Reviews Immunology*, vol. 14, no. 10. Nature Publishing Group, pp. 653–666,. 11, 2014, doi: 10.1038/nri3737.

[48] J. C. Paulson, M. S. Macauley, and N. Kawasaki, "Siglecs as sensors of self in innate and adaptive immune responses," *Ann. N. Y. Acad. Sci.*, vol. 1253, no. 1, pp. 37–48, 2012, doi: 10.1111/j.1749-6632.2011.06362.x.

[49] M. P. Lenza, U. Atxabal, I. Oyenarte, J. Jiménez-Barbero, and J. Ereño-Orbea, "Current Status on Therapeutic Molecules Targeting Siglec Receptors," *Cells*, vol. 9, no. 12, pp. 1–19, 2020, doi: 10.3390/cells9122691.

[50] A. Irie, S. Koyamat, Y. Kozutsumi, T. Kawasaki, and A. Suzuki, "The molecular basis for the absence of N-glycolylneuraminic acid in humans," *J. Biol. Chem.*, 1998, doi: 10.1074/jbc.273.25.15866.

[51] S. L. Diaz *et al.*, "Sensitive and specific detection of the non-human sialic acid N-Glycolylneuraminic acid in human tissues and biotherapeutic products," *PLoS One*, 2009, doi: 10.1371/journal.pone.0004241.

[52] W. X. Shi, R. Chammas, N. M. Varki, L. Powell, and A. Varki, "Sialic acid 9-O-acetylation on murine erythroleukemia cells affects complement activation, binding to I-type lectins, and tissue homing," *J. Biol. Chem.*, vol. 271, no. 49, pp. 31526–31532. 1996, doi: 10.1074/jbc.271.49.31526.

[53] E. R. Sjoberg, L. D. Powell, A. Klein, and A. Varki, "Natural ligands of the B cell adhesion molecule CD22β can be masked by 9- O-acetylation of sialic acids," *J. Cell Biol.*, vol. 126, no. 2, pp. 549–562. 1994, doi: 10.1083/jcb.126.2.549.

[54] S. Kelm, R. Schauer, J. C. Manuguerra, H. J. Gross, and P. R. Crocker, "Modifications of cell surface sialic acids modulate cell adhesion

mediated by sialoadhesin and CD22," *Glycoconj. J.*, vol. 11, no. 6, pp. 576–585. 1994, doi: 10.1007/BF00731309.

[55] E. C. M. Brinkman-Van Der Linden and A. Varki, "New aspects of siglec binding specificities, including the significance of fucosylation and of the sialyl-Tn epitope," *J. Biol. Chem.*, 2000, doi: 10.1074/jbc.275.12.8625.

[56] L. D. Powell, D. Sgroi, E. R. Sjoberg, I. Stamenkovic, and A. Varki, "Natural ligands of the B cell adhesion molecule CD22β carry N-linked oligosaccharides with α-2,6-linked sialic acids that are required for recognition," *J. Biol. Chem.*, 1993.

[57] B. E. Collins *et al.*, "Binding specificities of the sialoadhesin family of I-type lectins: Sialic acid linkage and substructure requirements for binding of myelin- associated glycoprotein, Schwann cell myelin protein, and sialoadhesin," *J. Biol. Chem.*, 1997, doi: 10.1074/jbc.272.27.16889.

[58] H. Begleiter and B. Porjesz, "Evoked brain potentials as indicators of decision-making," *Science (80-.).*, vol. 187, no. 4178, pp. 754–755, 1975, doi: 10.1126/science.1114321.

[59] T. Yamaji, T. Teranishi, M. S. Alphey, P. R. Crocker, and Y. Hashimoto, "A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to α2,8-disialyl and branched α2,6-sialyl residues. A comparison with Siglec-9," *J. Biol. Chem.*, vol. 277, no. 8, pp. 6324–6332,. 2002, doi: 10.1074/jbc.M110146200.

[60] E. M. Rapoport, G. V. Pazynina, M. A. Sablina, P. R. Crocker, and N. V. Bovin, "Probing sialic acid binding Ig-like lectins (siglecs) with sulfated oligosaccharides," *Biochem.*, vol. 71, no. 5, pp. 496–504, 2006, doi: 10.1134/S0006297906050051.

[61] M. A. Campanero-Rhodes *et al.*, "Carbohydrate microarrays reveal sulphation as a modulator of siglec binding," *Biochem. Biophys. Res. Commun.*, vol. 344, no. 4, pp. 1141–1146, 2006, doi: 10.1016/j.bbrc.2006.03.223.

[62] H. Tateno, P. R. Crocker, and J. C. Paulson, "Mouse Siglec-F and human Siglec-8 are functionally convergent paralogs that are selectively expressed on eosinophils and recognize 6′-sulfo-sialyl Lewis X as a preferred glycan ligand," *Glycobiology*, vol. 15, no. 11, pp. 1125–1135, . 2005, doi: 10.1093/glycob/cwi097.

[63] H. Attrill *et al.*, "The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design.," *Biochem. J.*, vol. 397, no. 2, pp. 271–278. 2006, doi: 10.1042/BJ20060103.

[64] A. P. May, R. C. Robinson, M. Vinson, P. R. Crocker, and E. Y. Jones,

"Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 å resolution," *Mol. Cell*, 1998, doi: 10.1016/S1097-2765(00)80071-4.

[65] M. F. Pronker *et al.*, "Structural basis of myelin-associated glycoprotein adhesion and signalling," *Nat. Commun.*, vol. 7, p. 13584. 2016.

[66] M. A. Zhuravleva, K. Trandem, and P. D. Sun, "Structural implications of Siglec-5-mediated sialoglycan recognition.," *J. Mol. Biol.*, vol. 375, no. 2, pp. 437–447,. 2008, doi: 10.1016/j.jmb.2007.10.009.

[67] M. S. Alphey, H. Attrill, P. R. Crocker, and D. M. F. Van Aalten, "High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family," *J. Biol. Chem.*, vol. 278, no. 5, pp. 3372–3377, 2003, doi: 10.1074/jbc.M210602200.

[68] J. M. Propster, F. Yang, S. Rabbani, B. Ernst, F. H.-T. Allain, and M. Schubert, "Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 29, pp. E4170-9. 2016, doi: 10.1073/pnas.1602214113.

[69] J. Ereño-Orbea *et al.*, "Molecular basis of human CD22 function and therapeutic targeting," *Nat. Commun.*, vol. 8, no. 764, p. 764, 2017, doi: 10.1038/s41467-017-00836-6.

[70] N. Izquierdo-Useros, M. Lorizate, P. J. McLaren, A. Telenti, H.-G. Kräusslich, and J. Martinez-Picado, "HIV-1 Capture and Transmission by Dendritic Cells: The Role of Viral Glycolipids and the Cellular Receptor Siglec-1," *PLoS Pathog.*, vol. 10, no. 7, p. e1004146. 2014, doi: 10.1371/journal.ppat.1004146.

[71] D. Perez-Zsolt *et al.*, "Anti-Siglec-1 antibodies block Ebola viral uptake and decrease cytoplasmic viral entry," *Nat. Microbiol.*, vol. 4, no. 9, pp. 1558–1570, 2019, doi: 10.1038/s41564-019-0453-2.

[72] S. Duan and J. C. Paulson, "Siglecs as Immune Cell Checkpoints in Disease," *Annual Review of Immunology*, vol. 38. pp. 365–395, 2020, doi: 10.1146/annurev-immunol-102419-035900.

[73] M. K. O'Reilly and J. C. Paulson, "Siglecs as targets for therapy in immune-cell-mediated disease," *Trends Pharmacol. Sci.*, vol. 30, no. 5, pp. 240–248, 2009, doi: 10.1016/j.tips.2009.02.005.

[74] A. Bärenwaldt and H. Läubli, "The sialoglycan-Siglec glyco-immune checkpoint–a target for improving innate and adaptive anti-cancer immunity," *Expert Opin. Ther. Targets*, vol. 23, no. 10, pp. 839–853, 2019, doi: 10.1080/14728222.2019.1667977.

[75] J. A. O'Sullivan, D. J. Carroll, Y. Cao, A. N. Salicru, and B. S. Bochner, "Leveraging Siglec-8 endocytic mechanisms to kill human eosinophils and malignant mast cells," *J. Allergy Clin. Immunol.*, vol. 141, no. 5, pp. 1774-1785.e7, 2018, doi: 10.1016/j.jaci.2017.06.028.

[76] B. S. Bochner, "'Siglec'ting the allergic response for therapeutic targeting.," *Glycobiology*, vol. 26, no. 6, pp. 546–552, 2016, doi: 10.1093/glycob/cww024.

[77] T. Kiwamoto, N. Kawasaki, J. C. Paulson, and B. S. Bochner, "Siglec-8 as a drugable target to treat eosinophil and mast cell-associated conditions," *Pharmacol. Ther.*, vol. 135, no. 3, pp. 327–336, 2012, doi: 10.1016/j.pharmthera.2012.06.005.

[78] B. A. Youngblood *et al.*, "Siglec-8 antibody reduces eosinophils and mast cells in a transgenic mouse model of eosinophilic gastroenteritis," *JCI Insight*, vol. 4, no. 19,. 2019, doi: 10.1172/jci.insight.126219.

[79] J. M. Propster *et al.*, "Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 29, pp. E4170-9. 2016, doi: 10.1073/pnas.1602214113.

[80] B. A. Youngblood *et al.*, "Discovery, Function, and Therapeutic Targeting of Siglec-8," *Cells*, vol. 10, no. 1, pp. 1–14, 2020, doi: 10.3390/cells10010019.

[81] M. Krystel-Whittemore, K. N. Dileepan, and J. G. Wood, "Mast cell: A multi-functional master cell," *Front. Immunol.*, vol. 6, pp. 1–12, 2016, doi: 10.3389/fimmu.2015.00620.

[82] T. Kambayashi and G. A. Koretzky, "Proximal signaling events in FcεRI-mediated mast cell activation," *J. Allergy Clin. Immunol.*, vol. 119, no. 3, pp. 544–552, 2007, doi: 10.1016/j.jaci.2007.01.017.

[83] B. A. Youngblood *et al.*, "AK002, a Humanized Sialic Acid-Binding Immunoglobulin-Like Lectin-8 Antibody that Induces Antibody-Dependent Cell-Mediated Cytotoxicity against Human Eosinophils and Inhibits Mast Cell-Mediated Anaphylaxis in Mice," *Int. Arch. Allergy Immunol.*, vol. 180, no. 2, pp. 91–102, 2019, doi: 10.1159/000501637.

[84] H. J. Na, S. A. Hudson, and B. S. Bochner, "IL-33 enhances Siglec-8 mediated apoptosis of human eosinophils," *Cytokine*, vol. 57, no. 1, pp. 169–174, 2012, doi: 10.1016/j.cyto.2011.10.007.

[85] E. S. Dellon *et al.*, "Anti–Siglec-8 Antibody for Eosinophilic Gastritis and Duodenitis," *N. Engl. J. Med.*, vol. 383, no. 17, pp. 1624–1634, 2020, doi: 10.1056/nejmoa2012047.

[86] S. Altrichter *et al.*, "P152 Clinical Activity of Ak002, an Anti-Siglec-8 Antibody, in Multiple Forms of Uncontrolled Chronic Urticaria," *Ann. Allergy, Asthma Immunol.*, vol. 123, no. 5, pp. S27–S28, 2019, doi: 10.1016/j.anai.2019.08.256.

[87] R. Hevey, "Strategies for the development of glycomimetic drug candidates," *Pharmaceuticals*. 2019, doi: 10.3390/ph12020055.

[88] S. Kelm, R. Brossmer, R. Isecke, H. J. Gross, K. Strenge, and R. Schauer, "Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues.," *Eur. J. Biochem.*, vol. 255, no. 3, pp. 663–672. 1998, doi: 10.1046/j.1432-1327.1998.2550663.x.

[89] S. M. W. Van Rossenberg *et al.*, "A Structure-Function Study of Ligand Recognition by CD22β," *J. Biol. Chem.*, vol. 276, no. 16, pp. 12967–12973,. 2001, doi: 10.1074/jbc.M009276200.

[90] S. Kelm, J. Gerlach, R. Brossmer, C. P. Danzer, and L. Nitschke, "The ligand-binding domain of CD22 is needed for inhibition of the B cell receptor signal, as demonstrated by a novel human CD22-specific inhibitor compound," *J. Exp. Med.*, vol. 195, no. 9, pp. 1207–1213, 2002, doi: 10.1084/jem.20011783.

[91] N. R. Zaccai *et al.*, "Structure-guided design of sialic acid-based Siglec inhibitors and crystallographic analysis in complex with sialoadhesin," *Structure*, vol. 11, no. 5, pp. 557–567, 2003, doi: 10.1016/S0969-2126(03)00073-X.

[92] B. S. Kroezen *et al.*, "A Potent Mimetic of the Siglec-8 Ligand 6'-Sulfo-Sialyl Lewisx," *ChemMedChem*, 2020, doi: 10.1002/cmdc.202000417.

[93] C. D. Rillahan, E. Schwartz, R. McBride, V. V Fokin, and J. C. Paulson, "Click and pick: identification of sialoside analogues for siglec-based cell targeting.," *Angew. Chem. Int. Ed. Engl.*, vol. 51, no. 44, pp. 11014–11018. 2012, doi: 10.1002/anie.201205831.

[94] C. M. Nycholat *et al.*, "A Sulfonamide Sialoside Analogue for Targeting Siglec-8 and-F on Immune Cells," *J. Am. Chem. Soc.*, vol. 141, no. 36, pp. 14032–14037. 2019, doi: 10.1021/jacs.9b05769.

[95] S. Duan *et al.*, "Nanoparticles Displaying Allergen and Siglec-8 Ligands Suppress IgE-Fc ε RI − Mediated Anaphylaxis and Desensitize Mast Cells to Subsequent Antigen Challenge," 2021, doi: 10.4049/jimmunol.1901212.

[96] M. del Carmen Fernandez-Alonso, D. Diaz, M. Alvaro Berbis, F. Marcelo, J. Canada, and J. Jimenez-Barbero, "Protein-Carbohydrate

Interactions Studied by NMR: From Molecular Recognition to Drug Design," *Curr. Protein Pept. Sci.*, vol. 13, no. 8, pp. 816–830, 2013, doi: 10.2174/138920312804871175.

[97] P. Çarçabal *et al.*, "Hydrogen bonding and cooperativity in isolated and hydrated sugars: Mannose, galactose, glucose, and lactose," *J. Am. Chem. Soc.*, vol. 127, no. 32, pp. 11414–11425, 2005, doi: 10.1021/ja0518575.

[98] M. López De La Paz, G. Ellis, M. Pérez, J. Perkins, J. Jiménez-Barbero, and C. Vicent, "Carbohydrate hydrogen-bonding cooperativity - Intramolecular hydrogen bonds and their cooperative effect on intermolecular processes - Binding to a hydrogen-bond acceptor molecule," *European J. Org. Chem.*, no. 5, pp. 840–855, 2002, doi: 10.1002/1099-0690(200203)2002:5<840::AID-EJOC840>3.0.CO;2-I.

[99] H. J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, and D. Solís, "From lectin structure to functional glycomics: Principles of the sugar code," *Trends Biochem. Sci.*, vol. 36, no. 6, pp. 298–313, 2011, doi: 10.1016/j.tibs.2011.01.005.

[100] L. Unione *et al.*, "Fluoroacetamide Moieties as NMR Spectroscopy Probes for the Molecular Recognition of GlcNAc-Containing Sugars: Modulation of the CH-pi Stacking Interactions by Different Fluorination Patterns.," *Chemistry*, vol. 23, no. 16, pp. 3957–3965, Mar. 2017, doi: 10.1002/chem.201605573.

[101] A. Ardá and J. Jiménez-Barbero, "The recognition of glycans by protein receptors. Insights from NMR spectroscopy," *Chem. Commun.*, vol. 54, no. 38, pp. 4761–4769, 2018, doi: 10.1039/c8cc01444b.

[102] E. Jiménez-Moreno *et al.*, "A thorough experimental study of CH/π interactions in water: quantitative structure-stability relationships for carbohydrate/aromatic complexes," *Chem. Sci.*, vol. 6, no. 11, pp. 6076–6085, 2015, doi: 10.1039/c5sc02108a.

[103] D. E. Danley *et al.*, "Protein-Carbohydrate Interactions Studied by NMR: From Molecular Recognition to Drug Design," *Curr. Protein Pept. Sci.*, vol. 13, no. 8, pp. 816–830, 2018, doi: 10.2174/138920312804871175.

[104] I. A. Bermejo *et al.*, "Water Sculpts the Distinctive Shapes and Dynamics of the Tumor-Associated Carbohydrate Tn Antigens: Implications for Their Molecular Recognition," *J. Am. Chem. Soc.*, vol. 140, no. 31, pp. 9952–9960, 2018, doi: 10.1021/jacs.8b04801.

[105] F. Marcelo *et al.*, "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.*, vol. 20, no. 49, pp. 16147–16155, . 2014, doi:

10.1002/chem.201404566.

[106] M. Atanasova, H. Bagdonas, and J. Agirre, "Structural glycobiology in the age of electron cryo-microscopy," *Curr. Opin. Struct. Biol.*, vol. 62, pp. 70–78, 2020, doi: 10.1016/j.sbi.2019.12.003.

[107] S. Pérez and D. De Sanctis, "Glycoscience@Synchrotron: Synchrotron radiation applied to structural glycoscience," *Beilstein J. Org. Chem.*, vol. 13, pp. 1145–1167, 2017, doi: 10.3762/bjoc.13.114.

[108] P. M. Nieto, "The use of NMR to study transient carbohydrate-protein interactions," *Front. Mol. Biosci.*, vol. 5, no. APR, pp. 1–7, 2018, doi: 10.3389/fmolb.2018.00033.

[109] A. Gimeno, P. Valverde, A. Ardá, and J. Jiménez-Barbero, "Glycan structures and their interactions with proteins. A NMR view," *Curr. Opin. Struct. Biol.*, vol. 62, pp. 22–30, 2020, doi: 10.1016/j.sbi.2019.11.004.

[110] F. Broecker *et al.*, "Multivalent display of minimal Clostridium difficile glycan epitopes mimics antigenic properties of larger glycans," *Nat. Commun.*, vol. 7, 2016, doi: 10.1038/ncomms11224.

[111] E. S. Ameh, "A review of basic crystallography and x-ray diffraction applications," *Int. J. Adv. Manuf. Technol.*, vol. 105, no. 7–8, pp. 3289–3302, 2019, doi: 10.1007/s00170-019-04508-1.

[112] J. M. Rondeau and H. Schreuder, *Protein Crystallography and Drug Discovery*. Elsevier Ltd, 2015.

[113] A. McPherson and J. A. Gavira, "Introduction to protein crystallization," *Acta Crystallogr. Sect. FStructural Biol. Commun.*, vol. 70, no. 1, pp. 2–20, 2014, doi: 10.1107/S2053230X13033141.

[114] K. V. Dunlop and B. Hazes, "A modified vapor-diffusion crystallization protocol that uses a common dehydrating agent," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 61, no. 8, pp. 1041–1048. 2005, doi: 10.1107/S0907444905013806.

[115] L. Govada and N. Chayen, "Choosing the Method of Crystallization to Obtain Optimal Results," *Crystals*, vol. 9, no. 2, p. 106. 2019, doi: 10.3390/cryst9020106.

[116] N. E. Chayen and E. Saridakis, "Protein crystallization: From purified protein to diffraction-quality crystal," *Nat. Methods*, vol. 5, no. 2, pp. 147–153, 2008, doi: 10.1038/nmeth.f.203.

[117] A. Viegas *et al.*, "Molecular determinants of ligand specificity in family 11 carbohydrate binding modules - An NMR, X-ray crystallography and

computational chemistry approach," *FEBS J.*, vol. 275, no. 10, pp. 2524–2535, 2008, doi: 10.1111/j.1742-4658.2008.06401.x.

[118] I. Müller, "Guidelines for the successful generation of protein–ligand complex crystals," *Acta Crystallogr. Sect. D Struct. Biol.*, vol. 73, no. 2, pp. 79–92, 2017, doi: 10.1107/S2059798316020271.

[119] B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine, and G. Klebe, "Two Methods, One Goal: Structural Differences between Cocrystallization and Crystal Soaking to Discover Ligand Binding Poses," *ChemMedChem*, vol. 16, no. 1, pp. 292–300, 2021, doi: 10.1002/cmdc.202000565.

[120] L. P. Calle, F. J. Cañada, and J. Jiménez-Barbero, "Application of NMR methods to the study of the interaction of natural products with biomolecular receptors," *Nat. Prod. Rep.*, vol. 28, no. 6, pp. 1118–1125, 2011, doi: 10.1039/c0np00071j.

[121] L. Unione, S. Galante, D. Díaz, F. J. Cañada, and J. Jiménez-Barbero, "NMR and molecular recognition. the application of ligand-based NMR methods to monitor molecular interactions," *Medchemcomm*, vol. 5, no. 9, pp. 1280–1289, 2014, doi: 10.1039/c4md00138a.

[122] A. Poveda and J. Jiménez-Barbero, "NMR studies of carbohydrate-protein interactions in solution," *Chem. Soc. Rev.*, vol. 27, no. 2, pp. 133–143, 1998, doi: 10.1039/a827133z.

[123] A. Arda *et al., Recent advances in the application of NMR methodologies to analyze the conformation, dynamics, and interactions of saccharides*, vol. 44. 2021.

[124] W. Braun, C. Bösch, L. R. Brown, N. GO, and K. Wüthrich, "Combined use of proton-proton overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon," *BBA - Protein Struct.*, vol. 667, no. 2, pp. 377–396, 1981, doi: 10.1016/0005-2795(81)90205-1.

[125] L. P. Mcintosh and F. W. Dahlquist, "Biosynthetic Incorporation of 15N and 13C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins," *Q. Rev. Biophys.*, vol. 23, no. 1, pp. 1–38, 1990, doi: 10.1017/S0033583500005400.

[126] M. P. Williamson, "Using chemical shift perturbation to characterise ligand binding," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 73, pp. 1–16, 2013, doi: 10.1016/j.pnmrs.2013.02.001.

Chapter II

Objectives

2 Objectives

From the training point of view, the key objectives of this Thesis have been to acquired knowledge on the expression and purification of glycoproteins and lectins in different expression systems, to master the application of NMR and X-Ray crystallography methodologies to study the interaction between carbohydrates and lectins and to learn concepts and design strategies to address scientific problems on biomolecular recognition.

From the scientific perspective, the aims were: first, to establish a new NMR-based methodology to deciphering the glycan composition on an intact glycoprotein (FcεRIα) and to analyse how these glycans interact with a lectin in experimental conditions close to the physiological ones.

In a second part, and motivated by the current pandemics, the target was to analyse the glycoprofile of the RBD of SARS-CoV-2 and assess its interactions with lectins of our immune system, using the NMR methodology developed for FcεRIα.

In a third part, the target was to initiate a new research line in the laboratory, focused on Siglecs. On this basis, the aim of this Thesis was to develop and establish the required methodology and protocols to combine NMR and X-Ray crystallography data to study Siglec-ligand interactions, using Siglec-8 as initial target. Thus, the study of the structure of Siglec-8 and their interactions with glycomimetics and antibodies was tackled by both NMR and X-Ray crystallography.

Chapter III

Materials and Methods

3.1 Protein expression and purification

In general, all the plasmids used to express the proteins analysed in this study were synthesized by Genscript Biotech. Sequences were subcloned into different vectors based on the expression system employed to produce the proteins. The carbohydrate recognition domains (CRD) of the lectins were expressed in *Escherichia coli (E.coli)* while all the glycoproteins were produced in human embryonic kidney cells (HEK293).

All the proteins were expressed and purified in our laboratory, except for the CRD of MGL [1] that was kindly provided by Dr. Filipa Marcelo, at Universidade NOVA de Lisboa (UNOVA).

3.1.1 Plasmid generation

Depending on the expression system, the different plasmids were chosen. For the proteins expressed in *E.coli,* the pET family of plasmids was employed (Table 3.1).

The sequence of human Siglec-8 CRD (V domain) (Uniprot: Q9NYZ4, amino acid residues 17-155, with the point mutation C42S) was cloned into pET-43.1(a) plasmid, following the construct designed by Propster et al[2]. The C-terminal domain was fused with a thrombin cleavage site and a His_6-tag to facilitate protein purification. The CRD of Galectin-3 (Uniprot: P17931, aminoacid residues 114-250), the N-terminal domain of Galectin-8 (Uniprot: O00214, aminoacid residues 1-155) and Galectin-7 (Uniprot: P47929, aminoacid residues 1-136) were cloned into pET-21a(+), pET-22a(+) and pET-22b(+) respectively. The CRD of DC-SIGN (Uniprot: Q9NNX6, aminoacid residues 254-404) was cloned into pET15b plasmid (Thermo Fisher Scientific).

Table 3.1: *pET family of vectors used in this PhD Thesis.*

Plasmid	Description	Ref.	Size (pb)	Resistence
pET-43.1(a)	T7 lac promoter/ N and C terminal His-Tag/ Thrombine recognition site	[3]	7275	Ampicillin
pET-21a(+)	T7 lac promoter/ C terminal His-Tag	[4]	5443	Ampicillin
pET-22a(+)	T7 lac promoter/ C terminal His-Tag	[5]	5493	Ampicillin
pET-22b(+)	T7 lac promoter/ C terminal His-Tag/ pelB coding sequence (for periplasmic localization)	[5]	5493	Ampicillin
pET15b	T7 lac promoter/ N and C terminal His-Tag/ Nter Thrombine recognition site*	[6]	5708	Ampicillin

*from [6]

For the expression of proteins in HEK293 cells, we cloned the sequences of RBD (Uniprot: P0DTC2, aminoacid residue 328-533), the full extracellular domain of Siglec-10 (uniprot: Q96LC7, aminoacid residue 17-537), Siglec-8 (Uniprot: Q6ZMC9, aminoacid residues), and FcεRIα (Uniprot: P12319, aminoacid residues 26-200) into pHLSec vector [7]. pHLSec contains the pBR322 origin of replication, a cytomegalovirus enhancer and a chick-actin promoter, necessary for the expression in the mammalian system, as well as an ampicillin resistance section for the selection of cells transformed with the plasmid. Inserts were placed between the AgeI and KpnI restriction sites, in order to maintain the secretion signal peptide at the N-terminal and the His-tag in the C-terminal, with a 6 x His tag in the C-terminal domain.

To improve protein solubility and stability, the extracellular domain of Siglec-10 and Siglec-8 were fused with the m-VENUS fluorescent protein [8], separated by a cleavage TEV (Tobacco Etch Virus) site.

3.1.2 Expression and purification of lectins in *E. coli*

The CRDs of Siglec-8, Galectin 3, Galectin 7, Galectin 8 (N-terminal domain), and DC-SIGN were expressed in *E.coli* cells. The advantage of using this expression system is that large amounts of ^{15}N-labeled proteins can be achieved to be used for NMR experiments.

Plasmids were transformed in competent cells (Table 3.2), following the protocol suggested by each provider. First, the plasmid (10 µg) was incubated with cells for 30 min and then permeabilized through the cell membrane using a heat shock at 325 K during 1 min. After the transformation, 300 µL of Luria-Bertani broth (LB) medium was added and cells were grown for 45 min. Finally, 200 µL of cells were seeded in a Agar LB plate with the corresponding antibiotics and incubated at 310 K overnight. Antibiotics allowed the selection of the cells with the plasmid integrated.

Table 3.2: *Different E.coli cells strains used in this PhD thesis.*

Strain	Genotype	Plasmid transfected	Ref
DH5α	*E.coli* F- 80dlacZ M15 (lacZYA-argF) U169 recA1 endA1hsdR17(rk-, mk+) phoAsupE44 -thi-1 gyrA96 relA1	all	[9]
BL21 (D3)	*E.coli* F – ompT hsdSB (rB–, mB–) gal dcm (DE3)	Gal-3, Gal-7, Gal-8 and DC-SIGN	[10]
Rosetta-gami™ 2 (DE3)	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F′[lac+ lacIq pro] gor522::Tn10 trxB pRARE2 (CamR, StrR, TetR)	Siglec-8	[11]

3.1.2.1 Expression of unlabelled proteins

A single bacterial colony selected from the Agar plate was grown overnight in 100 mL of LB media, supplemented with antibiotics, with a gentle shaking. The next day, 20 mL of cell culture were inoculated into 4 Fernbach flasks containing 1,5 L of LB media and antibiotics, at 310 K and 180 rev min^{-1} for 4h. When the OD_{600} (optical density at 600 nm) was 0.6, the culture was induced with 1 mM of Isopropyl β- d-1-thiogalactopyranoside (IPTG), and the grown was prolonged for 24/48h at 310 K (excepted for Siglec-8 that was incubated at 293K) and 180 rev min^{-1}. When the final OD_{600} was 4-5, cells were harvested by centrifugation at 4500 rpm for 30 min. The pellet was collected, frozen and stored at 253 K.

3.1.2.2 Expression of labelled ^{15}N proteins

For the expression of the uniformly ^{15}N-labelled CRD lectins, the transformed cells were inoculated in 200 mL of minimal medium composed by M9 with $^{15}NH_4Cl$, enriched with Biotin and Thiamine, and several other elements, as described in Tables 3.3-3.5, and incubated overnight at 310 K with gentle shaking (170 rpm). 15 mL were inoculated in 1.5 L of complete minimal medium. The growing conditions were as described in the previous section (3.1.1.1).

Table 3.3: *Final composition of M9 media*

M9 media (1L)	
100 mL	Salts Stock (10X)
1 g	$^{15}NH_4Cl$
2 mL	Trace elements solution (250X)
3 g	Glucose
1 mL	$MgSO_4$ (1M)
1 mL	$CaCl_2$ (0.1M)
1 mL	Thiamine (10 mg/mL)
1 mL	Biotin (10 mg/mL)
X	Antibiotics

Table 3.4: *Salts composition*

Stock of salts (1L)	
60 g	Na2HPO4
30 g	KH2PO4
5g	NaCl

Table 3.5: *Composition of trace elements*

Trace elements solution (100X) (for 1L)	
5 g	EDTA
0.83 g	FeCl3 x 6 H2O
84 mg	ZnCl2
13 mg	CuCl2 x 2 H2O
10 mg	CoCl2 x 6 H2O
10 mg	H3BO3
1.6 mg	MnCl2 x 6 H2O

3.1.2.3 Purification of the CRD of Siglec-8

The cell pellet was resuspended in a lysis buffer (20 mM TRIS pH8, 300 mM NaCl, 1 mM of phenylmethylsulfonyl fluoride -PMSF-, and a cocktail of protease inhibitors cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roche). Cells were lysed by sonication for 12 cycles (15 sec on 59 sec off, Amp= 60%) to extract the produced protein from the bacteria. To separate the soluble proteins from the insoluble ones, the lysate was ultra-centrifugated at 30.000 rpm for 1h at 277 K. After ultracentrifugation, the supernatant was filtered through a 0.45 micron sterile filter membrane and was loaded on the 5 mL His-trap FF crude column (GE Life Sciences), previously equilibrated with 5 column volume (CV) of 20 mM Tris pH 8 and 300 mM NaCl. To obtain the pure protein, the column was cleaned with 40 CV of 20 mM Tris pH 8, 300 mM NaCl and 25 mM Imidazole buffer, and then the protein was eluted with 20 mM TRIS pH 8, 300 mM NaCl, and 500 mM Imidazole. The eluted fraction was collected and the protein was concentrated in an Amicon filter with a cut-off of 5 kDa, up to a minimum volume of 10 mL. Subsequently, a gel filtration chromatography step was carried out, using a Superdex 26/600 75 pg (GE Life Sciences), and the protein was eluted in 20 mM of Sodium Phosphate pH 7.4 or 40 mM NaCl buffer or 20 mM TRIS pH 8 and 300 mM NaCl. Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were used to show that Siglec-8 V domain elutes as a monomer from the size exclusion step, with the molecular weight of 16 kDa. Additionally, the protein sequence was verified by nano-scale liquid chromatographic tandem mass spectrometry (nLC MS/MS) in the Proteomics Platform in CIC bioGUNE. Finally, the protein was concentrated using an Amicon filter with a cut-off of 5 kDa at a final concentration of 0.8 mg/mL (50 µM) for NMR analysis or 16 mg/mL (1 mM) for crystallization assays. The remaining protein was, in some cases, flash-frozen using liquid nitrogen and stored at 193 K.

3.1.2.4 Purification of the CRD of Galectins

The collected pellet was resuspended in a lysis buffer with 22 mM tris(hydroxymethyl)aminomethane (Tris-HCl) pH 7.5, 5 mM EDTA and 1 mM dithiothreitol (DTT). Cells were sonicated and ultra-centrifugated for 30 min, and the supernatant was incubated with a α-Lactose-Agarose resin (Sigma-Aldrich). Proteins were eluted with the elution buffer containing PBS 1X and 150 mM of lactose. Before the NMR experiment, proteins were dialyzed to eliminate all the lactose using the elution buffer [12].

3.1.2.5 Purification of the CRD of DC-SIGN

The purification of DC-SIGN followed the same procedure that was previously developed in our laboratory [13]. After collection, the pellet was resuspended in a lysis buffer (10 mM of TRIS pH 8); the bacteria cells were sonicated and ultra-centrifugated at 30,000 rpm for 1 h at 277 K. Since DC-SIGN is unfolded in the inclusion bodies, a chemical refolding procedure was employed to solubilize the protein. First, the insoluble fraction of the sonicated cells was incubated with 6M of urea overnight at 277 K to denature the protein. Then, it was ultra-centrifugated for 2h at 30,000 rpm at 277 K. The soluble part was collected and re-folded by decreasing the urea concentration during three consecutive days (dialysis with 4 M, 2 M, and 0 M of urea, 24 h at 277 K each). For protein purification, mannose-Sepharose beads were used. The protein was eluted using a buffer containing 10 mM of EDTA and purified using Gel filtration chromatography, in a Superdex 75 26 600 column (GE Healthcare). The protein purity and identity was checked using SDS-PAGE gel and mass spectrometry, respectively.

3.1.3 Expression and purification of human glycoproteins from mammalian cells

For the expression of the human glycoproteins two different human embryonic cells were used: i) HEK293T cells in adherent mode; and ii) the suspension

adapted free style HEK293 (HEK293F) and HEK293S (deficient in N-acetylglucosaminyltransferase I -GnTI-enzyme) cells.

Adherent HEK293T cells were used for the expression of the isotope labelled glycoproteins. The specific mammalian cell growth media U-^{13}C,^{15}N Bioexpress 6000 (CIL), was included in the expression system to obtain a uniformly labelled glycoprotein in ^{13}C and ^{15}N [14]. However, this recombinant expression system yields limited quantity of glycoprotein and it is rather expensive.

For this reason, the yields of expression of glycoproteins were highly improved using the suspension HEK293F cells [15]. HEK293F cells are grown in a free fetal bovine serum (FBS) medium that allows to purify the protein without other contaminant proteins. These features allowed obtaining the large amounts of glycoproteins required for the structural studies. The expression to produce ^{13}C labelled glycans on the glycoprotein was further improved by the addition of U-^{13}C-glucose (3 g/L) directly to the media [16]. This methodology allows the access to high amounts of ^{13}C-labelled glycoproteins at relatively low cost.

In contrast with the bacterial expression, the transfection in the mammalian cells needs a huge amount of DNA, which was transfected in *E.coli* DH5α cells, and extracted by a Maxi-prep process.

3.1.3.1 DNA extraction by Maxi-preparation

DNA extraction was performed by the PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen) following the protocol suggests by provider. Briefly, Plasmid DNA was transformed in *E.coli* DH5α cells, which are competent cells generally used for cloning application, and grown in 200 mL of LB media, supplemented with antibiotics over night with shaking. Then, cells pellet was collected by centrifugation and lysed in order to solubilize the DNA. Another centrifugation was performed, to separate the cells debris from soluble fraction, which was charged in anion-exchange resin (provides by the kit) where the

negatively charged phosphates of the DNA backbone interact with the positive charges on the surface of the anion-exchange resin. The eluted DNA was precipitated with isopropanol and washed by a solution with 70% ethanol. Finally, the DNA pellet was resuspended in sterile H_2O at the final concentration to 1 µg/µL.

3.1.3.2 Expression in HEK293T cells

The transient transfection methodology for the expression in adherent cells described by Barbieri et al. was employed [14]. On day 1, 4.4 x 10^6 cells (viability >97%) were seeded in two p150 plates with 20 mL of complete grown media (Dulbecco's Modified Eagle Medium, DMEM, 10% FBS, 1% non essential amino acids (NEAA) and penicillin-streptomycin antibiotics) to prepare the cells for transfection. After 24 h (day 2), a mix transfection was prepared with 25 µg of endotoxin-free DNA and 50 µg of PEI dissolved in 5 mL of DMEM media with 2% of FBS (for the unlabelled protein) or U-^{13}C,^{15}N Bioexpress 6000 media (for the labelled one). The transfection mix was incubated for 20 min. During the incubation step, the cells were washed twice with DPBS to remove all contamination from FBS. Cells were grown in Heracell 150i incubator (Thermo Fisher Scientific) at 310 K in a humidified atmosphere containing 5% CO_2 for 72 h.

3.1.3.3 Expression in HEK293F/S cells

HEK293F/S cells were split every three days at 0.5 x 10^6 cells/mL (with a viability of >97%) in 200 mL of HEK293 Freestyle media in sterile culture flask with vented cap and grown in a Minitron Infors HT orbital shaker incubator at 310 K, 125 rpm, 70% of humidity and 8% CO_2. For transfection, cells were split into flasks with a final density between 0.8-1 x 10^6 cells/mL (with a viability of >97%) in 200 mL of Freestyle media. The transient transfection of HEK293F/S cells was achieved by addition of 50 µg of Endotoxin-free DNA plasmid and 50 µg of Fectopro (Polyplus transfection) transfection reagent.

To produce Fabs, two different plasmids containing the heavy and the light chain, respectively, must be transfected in the same HEK293F/S cells. To optimize the quantity of the recombinant Fab protein obtained, a 2:1 HC/LC ratio of plasmids was kept in the transfection process. Thus, for each flask of 200 mL, a DNA mix was prepared with 40 µg of HC and 20 µg of LC that was mixed with 60 µL of FectoPRO. On the other hand, protein-protein complexes such as, Siglec-8_{d1-d3} -AK002 Fab, were produced in 200 mL cell culture by mixing 20 µg of Siglec-8_{d1-d3} with 20 µg of AK002 HC, and 10 µg of AK002 LC, which was mixed with 50 µg of FectoPRO [15].

The plasmidic DNA was mixed with 5 mL of media and filtered through 0.22 micron filters and then incubated for 10 min with FectoPro at room temperature. Finally, the DNA-Fectopro mix was added directly to the cells and incubated for 7 days at 310 K, 125 rpm, 70 % of humidity and 8 % CO_2 [15]

3.1.3.4 Purification of FcεRIα from adherent HEK293T cells

After 72 h in the incubator, the adherent HEK293T cells were collected by centrifugation, and the supernatant was filtered through a 0.45 micron sterile filter. FcεRIα was purified through a Ni-NTA agarose column, previously equilibrated with PBS 1X pH 7.4. The supernatant was incubated for 30 minutes at 277 K with moderate shaking, and then washed with PBS 1X and 20 mM imidazole five times. Finally, the protein was eluted by adding PBS 1X pH 7 two times, 200 mM imidazole and one time PBS 1X pH 7, 500 mM Imidazole. In order to eliminate the imidazole traces, a buffer exchange step with PBS 1X pH 6.8 was carried out in an Amicon filter with a cut-off of 10 kDa, and the protein was concentrated to a final concentration of 1mg/mL (40 µM). The purity of the protein was checked by Coomassie SDS-PAGE gel and mass spectrometry.

3.1.3.5 Purification of RBD, Siglec-8 d1-d3, Siglec-10 d1-d5 and FcεRIα from suspension HEK293F/S cells.

Cell cultures were harvested at 10,000 rpm for 30 min at room temperature and subsequently filtered using a 0.45 micron filter. The supernatant was loaded on a 5 mL His-trap FF crude column (GE Life Sciences), previously equilibrated with 5 CV of 20 mM Tris pH 8 and 300 mM NaCl. The column was washed with 20 CV of buffer with 25 mM of imidazole and the proteins were eluted with 10 CV of elution buffer with 20 mM Tris pH 8, 300 mM NaCl and 500 mM imidazole. The collected fractions were concentrated using an Amicon filter with a cut-off of 10 kDa up to a final volume of 500 μL. The concentrated protein was then injected in the Superdex 75 10/300 Increase or Superdex 200 10/300 Increase gel filtration columns, previously equilibrated in PBS 1X pH 7.4.

3.1.3.6 Expression and purification of AK002 Fab

After 7 days of incubation, the supernatant of the HEK293F cells was separated by centrifugation and loaded into a 5 mL HiTrap Kappa Select column (GE Healthcare life Sciences) able to bind the constant region of the LC. In this case, the loaded protein was eluted with 20 CV of the elution buffer containing 100 mM Glycine pH 3. To prevent protein denaturation due to the low pH of the elution buffer, 100 mM of Tris pH and injected in a Superdex 75 10 300 Increase gel filtration column. The purity of the protein was checked by gel electrophoresis, under non reducing and reducing conditions, and by mass spectrometry.

3.1.3.7 Purification of the Siglec-8 d1d3-AK002 complex

After 7 days, the supernatant was harvested and loaded into a His column of 5 mL. 20 CV of buffer with 25 mM of Imidazole was used to wash the column, followed by a second step with 500 mM imidazole to elute the protein complex. The eluted protein complex was treated with Endo H for 1h at 37C in order to

eliminate the high mannose glycans. Then, the concentrated protein was loaded into Superdex 200/10 300 increase gel filtration column to separate the aggregates from the Siglec-8-AK002 complex. Finally, the protein complex was concentrated to 10 mg/mL using an Amicon filter with a cut-off of 10 kDa. The purity was checked by SDS-PAGE gel electrophoresis under reducing and non-reducing conditions and by mass spectrometry.

3.2 NMR experiments

All NMR experiments were recorded using an 800 MHz Bruker Avance III spectrometer equipped with a cryoprobe, except for the $^{1}H,^{15}N$-HSQC of MGL that was recorded on a 600 MHz BRUKER AVANCE III spectrometer, equipped with a 5mm inverse detection triple-resonance z-gradient cryogenic probe (CP-TCI), at UNOVA.

3.2.1 NMR assignment of N-glycans of RBD and FcεRIα

FcεRIα and RBD were resuspended in pure D_2O Phosphate Buffered Saline (PBS) 1X at pH 6.8 and 7.4, respectively. 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) at 1 mM concentration was added as an internal reference. 5 mm Shigemi NMR tube with 300 µL of glycoprotein sample concentrated to 60 µM (for FcεRIα) and 625 µM (for RBD) was prepared. Standard TOCSY-HSQC (from 20 to 100 ms mixing times), NOESY-HSQC (200 and 300 ms mixing times) and ^{1}H, ^{13}C-HSQC experiments were used to characterize the glycoprofile of FcεRIα. For the RBD, a combined analysis of 3D H',CH NOESY-HSQC, H',CH TOCSY-HSQC, and H'[C'],CH and [H']C',CH edited HSQC-[$^{13}C,^{13}C$]TOCSY-HSQC allowed to determine the precise structure of the two N-glycans and their glycosidic linkages. TopSpin 3.2.7 (BRUKER) was used for both data acquisition and processing.

3.2.2 Molecular interaction studies from the lectin point of view

60 µM of ^{15}N-lectin samples in the absence or presence of unlabelled RBD (1:1, 1:0.4 or 1:0.5) were analysed by 2D ^{1}H,^{15}N-BEST-TROSY at 310 K, in Shigemi tubes with 5 mm of diameter. The amino acid backbone of each lectin was assigned using the data deposited into the Biological Magnetic Resonance Data Bank (BMRB) (bmrb 4909 for Galectin-3, bmrb 17826 for galectin-7, bmrb 27854 for DC-SIGN, and bmrb 25798 for Siglec-8). Backbone resonance assignment for MGL and galectin-8N were carried out following standard protocols. The software CcpNmr Analysis V2 [17] was used to evaluate experimental data.

3.2.3 Molecular interaction studies from the glycan point of view

To analyse the interaction between N-glycans and lectins, 2D ^{1}H,^{13}C-HSQC spectra of ^{13}C-glycans on the RBD (concentrated at 60 µM) at 310 K was recorded. The cross-peak volumes wwereas integrated without and with 1 (for Gal-3, Gal-7, Siglec-8 and MGL) or 0.2 (Siglec-10 and Gal-8 N-ter) equivalents of lectin. To evaluate the binding, the peak volume was monitored, and the integration of the peaks was carried out with TopSpin 3.5.pI6 software.

3.2.4 ^{1}H,^{15}N-TROSY based Titration experiments for analyzing the binding of lectins to glycan ligands

To analyze the interactions between Siglec-8 V domain and the sialic acid containing ligands, from the point of view of the protein, ^{1}H,^{15}N-TROSY experiments were performed, using systematic increasing concentrations of the ligands. This approach allowed evaluating the most perturbed amino acids on Siglec-8 upon ligand binding. The three selected ligands, Neu5Acα2−3Galβ1−4GlcNAc (**1**), Neu5Acα2−3[6SO$_4$]Galβ1−4GlcNAc (**2**) and ^{NSA}Neu5Acα2−3[6SO$_4$]Galβ1−4GlcNAc (**3**) [18], were gently provided by our collaborators Dr. Nycholat and Prof. Paulson from the Scripps Research Institute (La Jolla, California).

Uniformly ^{15}N-labeled Siglec-8 V was dissolved in 500 µL of phosphate buffer (20 mM sodium phosphate, 40 mM NaCl, pH 7.4) at a concentration of 50 µM using 90% H_2O and 10% D_2O. Due to the different ligand affinities, 50 equivalents, 10 equivalents or 400 equivalents for ligands **1**, **2** and **3**, respectively, were added to completely saturate the protein. For this purpose, standard ^{1}H,^{15}N-TROSY experiments were performed with 2048 (t1) and 256 (t2) complex data points in ^{1}H and ^{15}N dimensions, respectively. All the experiments were recorded at a temperature of 293 K to avoid protein precipitation. NMR data were analysed with CcpNmr Analysis V2.

3.2.5 Saturation transfer difference (STD-NMR) experiments

For the STD experiments of Siglec-8 V with **2** and **3**, the protein was dissolved in deuterated buffer (20 mM Sodium phosphate and 40 mM of NaCl, pD 7.4). A standard 1:50 protein:ligand molar ratio was used, with a protein concentration at 50 µM. All the experiments were performed at 293K.

The 1D STD sequence from Bruker library with spoil and T2 filter (stddiff.3) was employed for the STD-NMR experiments. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at 0.75 ppm, for a total saturation time of 2 seconds. For the reference spectrum (off-resonance), the irradiation frequency was set at 100 ppm.

The STD NMR spectra were obtained by subtracting the on-resonance from the off-resonance spectrum. Additional blank spectra were recorded using the same conditions with only protein or only ligand. The absolute STD value was calculated from the ratio between the intensity of a given signal in the off resonance versus the STD spectra. The 100 % value was assigned to the proton with the strongest STD effect. Based on this, the relative STD effect for all the other protons was calculated.

3.3 Crystallization and crystal structure determination of Siglec-8 alone and in complex with sialylated ligands or Fabs

In this study, the V-domain of Siglec-8 in the apo form, bound to ligand **2**, and the full-length ECD of Siglec-8 in complex with AK002 Fab were crystallized and their structure analyzed by X-Ray crystallography.

3.3.1 Crystallization screening

To find the optimal conditions for crystallizing Siglec-8 V and Siglec-8-AK002, a pre-crystallization test (PTC^{TM}, Hampton) was firstly performed to determine the ideal protein concentration. A preliminary crystallization study was carried out using the sitting drop technique in 96-wells MRC crystallization plates with a protein concentration of 10 mg/mL. Using this methodology, an extensive scale screening, assisted by liquid handling robots and automatic nano-dispensers, was performed, allowing the examination of more than 1800 different solutions from commercial screening available on the CIC bioGUNE X-ray platform (Table 3.6). The TECAN freedom EVO robot (Tecan Group Ltd.) was used for the distribution of 70 µL of precipitating solutions in the corresponding wells on plates, while crystallization drops were prepared with the help of a MOSQUITO nanodispenser (ttplabtech), mixing 200 nL of the solution containing the precipitating agent with 200 nL of the solution containing the purified protein. Plates were stored at 277 K or 291 K, and finally, the presence or the absence of crystals was checked using a MZ12.5 light stereomicroscope (Leica).

Table 3.6: *Commercial crystallization screens used in this PhD Thesis.*

Crystallization Screening	Provider	Catalog Number
PCT™ Pre-Crystallization test	Hampton Research Corp	HR2-140
PEG/Ion	Hampton Research Corp	HR2-126
PEG/Ion 2	Hampton Research Corp	HR2-098
SaltRx 1	Hampton Research Corp	HR2-107
SaltRx 2	Hampton Research Corp	HR2-109
Crystal Screen Cryo™ 1	Hampton Research Corp	HR2-914-01
Crystal Screen Cryo™ 2	Hampton Research Corp	HR2-914-02
Top96 Crystal Screen	Anatrace	TOP96-10ML
Morpheus™	Molecular Dimensions Ltd	MD1-46

Once promising preliminary conditions were identified (Table 3.7), a more specific screening was carried out to optimize the crystallization process. In this second step, the size of the drops was significantly increased, and the hanging drop technique was used in 24-well VDX limbro boxes. The hanging drops contained 1 µL of protein and 1 µL of precipitant. The volume of precipitating solution in the reservoir during the optimization process was 0.5 mL. Plates were incubated at a constant temperature of 277 K or 291 K.

Finally, crystals were collected and flash-frozen into liquid nitrogen using a cryo-protection solution (crystallization buffer with 25 % v/v of glycerol).

Table3.7: *Crystallization conditions for AK002 Fab, Siglec-8 V domain in the unliganded and liganded forms, and for Siglec-8_{a1a3} in complex with AK002.*

Protein	Crystal screen	Crystallization condition	Cryoprotectant
AK002	PEG Ion	20 % (w/v) PEG 3350 0.2 M Ammonium Tartrate dibasic	20% glycerol
Siglec8-AK002	PEG Ion	20 % (w/v) PEG 3350 0.2 M Ammonium Citrate tribasic, pH 7.0	20% glycerol
Siglec-8	Top96	25 % (w/v) PEG 3350 0.1 M BisTris-HCl, pH 6.5	20% glycerol
Siglec-8_NSP	Crystal Screen Cryo	0.075 M HEPES pH 7.5 15% w/v PEG 10,000 25% v/v Glycerol	none
		0.65 M Imidazole pH 7.0 35% v/v Glycerol	none

3.3.2 Soaking and co-crystallization with Siglec-8 V domain and Ligand 3

Two different strategies were followed to obtain Siglec-8 crystals bound to the sialic acid-containing ligand. The first one consisted in the pre-incubation of the protein with the ligand before its crystallization; this methodology is called co-crystallization. The second strategy consisted in soaking the previously obtained crystals of the protein with the ligand. In this case, the ligand needs to penetrate into the crystal and interact with the binding site [19].

Soaking experiments were performed with the Siglec-8 V crystals obtained in a 24-well plate. A solution of ligand **3** (at 3 mM concentration) was added directly

to the drop, at a 1:3 protein:ligand molar ratio. Crystals were incubated with the ligand for different times (1 min, 5 min, 1 hour and overnight), and then flash-frozen into liquid nitrogen using a cryo-protection solution (crystallization buffer containing the ligand with 25 % v/v of glycerol).

For the co-crystallization experiments, the V domain of Siglec-8 (at 1 mM) was incubated with 10 equivalents of **3** (at 10 mM), and different crystallization conditions were screened at 277 K.

3.3.3 X-ray diffraction data collection and analysis

The synchrotron lines used in this study were the MX-XALOC-BL13 line at ALBA synchrotron (Barcelona, Spain) and the MX-SLS Beamline PX III at the Paul Scherren Institute PSI (Villigen, Switzerland). Depending on the crystals, the distance from the detector was set from 2.2 to 3 Å. The number of images and the start angle necessary to achieve 100 % completeness of the spectrum were determined using the EDNA [20] program. Images were taken in intervals of 0.1° with 3600 images for each crystal. The collection parameters used with the diffracting crystals of Siglec-8 V domain are gathered in Table 8. The obtained diffraction data were indexed, integrated and scaled with the autoPROC [21] program.

The structure of Siglec-8 V domain was solved by molecular replacement, using as an initial phase search model the NMR structure of Siglec-8 V (PDB ID: 2N7A). Phases were calculated with the PHASER [22] program integrated into the Phenix package. The structural model was built with the COOT [23] program and underwent several refinement cycles with the Phenix [24] Refine program. The geometric quality of the final structure was validated with Molprobity [25]. The three-dimensional model figures were prepared with PyMOL (http://www.pymol.org/).

Table 3.8: *Parameters of the data collection for X-ray diffraction for Siglec-8 V domain*

Parameters	Siglec-8 V Domain
Number of images	3600
Δ Φ (°/image)	0.1
Beamline	MX-SLS Beamline PX III
λ (Å)	0.979
Exposure (sec)	0.2
Detector	PILATUS 2M-F
Temperature	-173

Table 3.9: *Parameters of the data collection for X-ray diffraction for Siglec-8d1d3-AK002*

Parameters	Siglec-8_{d1d3}-AK002
Number of images	3600
Δ Φ (°/image)	0.1
Beamline	MX-SLS Beamline PX III
λ (Å)	1.00
Exposure (sec)	0.2
Detector	PILATUS 2M-F
Temperature	-173

Table 3.10: *Parameters of the data collection for X-ray diffraction for AK002*

Parameters	AK002
Number of images	3600
Δ Φ (°/image)	0.1
Beamline	MX-SLS Beamline PX III
λ (Å)	1.00
Exposure (sec)	0.2
Detector	PILATUS 2M-F
Temperature	-173

3.4 References

[1] F. Marcelo *et al.*, "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.*, vol. 20, no. 49, pp. 16147–16155,. 2014, doi: 10.1002/chem.201404566.

[2] J. M. Propster, F. Yang, S. Rabbani, B. Ernst, F. H.-T. Allain, and M. Schubert, "Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 29, pp. E4170-9. 2016, doi: 10.1073/pnas.1602214113.

[3] A. C. Theos *et al.*, "Functions of adaptor protein (AP)-3 and AP-1 in tyrosinase sorting from endosomes to melanosomes," *Mol. Biol. Cell*, vol. 16, no., pp. 5356–5372, 2005, doi: 10.1091/mbc.E05.

[4] E. B. Fauman *et al.*, "Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A," *Cell*, vol. 93, no. 4, pp. 617–625, 1998, doi: 10.1016/S0092-8674(00)81190-3.

[5] H. L. Hyung *et al.*, "Crystallization and preliminary X-ray crystallographic analysis of enoyl-acyl carrier protein reductase from Helicobacter pylori," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 58, no. 6 II, pp. 1071–1073, 2002, doi: 10.1107/S0907444902007151.

[6] M. I. Singh and V. Jain, "Tagging the Expressed Protein with 6 Histidines: Rapid Cloning of an Amplicon with Three Options," *PLoS One*, vol. 8, no. 5, 2013, doi: 10.1371/journal.pone.0063922.

[7] A. R. Aricescu, W. Lu, and E. Y. Jones, "A time- and cost-efficient system for high-level protein production in mammalian cells.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 62, no. Pt 10, pp. 1243–1250. 2006, doi: 10.1107/S0907444906029799.

[8] G. J. Kremers, J. Goedhart, E. B. Van Munster, and T. W. J. Gadella, "Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET förster radius," *Biochemistry*, vol. 45, no. 21, pp. 6570–6580, 2006, doi: 10.1021/bi0516273.

[9] B. P. Anton and E. A. Raleigh, "Complete genome sequence of NEB 5-alpha, a derivative of Escherichia coli K-12 DH5α," *Genome Announc.*, vol. 4, no. 6, pp. 6–7, 2016, doi: 10.1128/genomeA.01245-16.

[10] M. J. Rogals, J.-Y. Yang, R. V Williams, K. W. Moremen, I. J. Amster, and J. H. Prestegard, "Sparse isotope labeling for nuclear magnetic resonance (NMR) of glycoproteins using 13C-glucose," *Glycobiology*,

vol. 00, no. 00, pp. 1–11,. 2020, doi: 10.1093/glycob/cwaa071.

[11] M. Fathi-Roudsari, A. Akhavian-Tehrani, and N. Maghsoudi, "Comparison of three Escherichia coli strains in recombinant production of reteplase," *Avicenna J. Med. Biotechnol.*, vol. 8, no. 1, pp. 16–22, 2016.

[12] A. Gimeno *et al.*, "Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3.," *Angew. Chem. Int. Ed. Engl.*, vol. 58, no. 22, pp. 7268–7272, 2019, doi: 10.1002/anie.201900723.

[13] P. Valverde *et al.*, "Molecular Insights into DC-SIGN Binding to Self-Antigens: The Interaction with the Blood Group A/B Antigens," *ACS Chem. Biol.*, vol. 14, no. 7, pp. 1660–1671. 2019, doi: 10.1021/acschembio.9b00458.

[14] L. Barbieri, E. Luchinat, and L. Banci, "Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells," *Nat. Protoc.*, vol. 11, no. 6, pp. 1101–1111. 2016, doi: 10.1038/nprot.2016.061.

[15] J. Ereño-Orbea, T. Sicard, H. Cui, I. Akula, and J.-P. Julien, "Characterization of Glycoproteins with the Immunoglobulin Fold by X-Ray Crystallography and Biophysical Techniques," *J. Vis. Exp.*, no. 137. 2018, doi: 10.3791/57750.

[16] K. Kato, Y. Yamaguchi, and Y. Arata, "Stable-isotope-assisted NMR approaches to glycoproteins using immunoglobulin G as a model system," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 56, no. 4, pp. 346–359, 2010, doi: 10.1016/j.pnmrs.2010.03.001.

[17] L. Mureddu and G. W. Vuister, "Simple high-resolution NMR spectroscopy as a tool in molecular biology," *FEBS J.*, vol. 286, no. 11, pp. 2035–2042, 2019, doi: 10.1111/febs.14771.

[18] C. M. Nycholat *et al.*, "A Sulfonamide Sialoside Analogue for Targeting Siglec-8 and-F on Immune Cells," *J. Am. Chem. Soc.*, vol. 141, no. 36, pp. 14032–14037. 2019, doi: 10.1021/jacs.9b05769.

[19] B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine, and G. Klebe, "Two Methods, One Goal: Structural Differences between Cocrystallization and Crystal Soaking to Discover Ligand Binding Poses," *ChemMedChem*, vol. 16, no. 1, pp. 292–300, 2021, doi: 10.1002/cmdc.202000565.

[20] M. F. Incardona, G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov, and O. Svensson, "EDNA: A framework for plugin-based applications applied to X-ray experiment online data analysis," *J. Synchrotron*

Radiat., vol. 16, no. 6, pp. 872–879, 2009, doi: 10.1107/S0909049509036681.

[21] C. Vonrhein *et al.*, "Data processing and analysis with the autoPROC toolbox," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 293–302, 2011, doi: 10.1107/S0907444911007773.

[22] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007, doi: 10.1107/S0021889807021206.

[23] P. Emsley and K. Cowtan, "Coot: Model-building tools for molecular graphics," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2126–2132, 2004, doi: 10.1107/S0907444904019158.

[24] T. C. Terwilliger *et al.*, "Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 65, no. 6, pp. 582–601, 2009, doi: 10.1107/S0907444909012098.

[25] I. W. Davis *et al.*, "MolProbity: All-atom contacts and structure validation for proteins and nucleic acids," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, pp. 375–383, 2007, doi: 10.1093/nar/gkm216.

Chapter IV

NMR-based glycoprofile analysis of FcεRIα

4.1 Introduction

As mentioned in the Introduction, the most abundant post-translational modification in proteins is the N-glycosylation, and it is common in all eukaryotic cells [1]. Glycans are extremely heterogeneous due to several variables such as the variety of the monosaccharide components, the specific sequence, and the type of glycosidic linkage that link the sugar moieties [2]. Glycans are involved in several biological events, including protein stability, cell signaling, differentiation, tissue differentiation, host-pathogens recognition, immune response, and cancer [2][3]. The pathogenesis of many pathologies like cancer, auto-immune, and inflammatory diseases are intimately related to an alteration in the glycosylation pattern [2].

Therefore, the analysis of the glycan composition on glycoproteins is rather challenging. In general, a combination of mass spectrometry and chemical and/or enzymatic digestion/degradation is employed [4][5]. However, this type of analysis can only identify the sugar composition, while the specific assignment of the type of glycosidic linkage (α) is still undefined. Moreover, the digestion of the glycoprotein may lead to the degradation of certain types of glycosidic linkages, whose information is lost.

In this study, we have developed a novel methodology, based on state-of-the-art NMR spectroscopy, that allows analysing the glycan composition on intact glycoproteins under experimental conditions close to the physiological ones.

As proof-of-concept, we have chosen the extracellular domain of the human high-affinity receptor for IgE (FcεRI) subunit α, a protein of 25 kDa that gathers 2 Ig-like domains and displays 7 N-glycosylation sites (N18, N39, N47, N71, N132, N137, and N163). FcεRI is expressed on the surface of mast cells and basophils. Interestingly, IgE binding leads to the release of inflammatory mediators (histamine and other cytokines), which causes an allergic response [6][7][8].

4.2 Implementing the protocol for FcεRIα glycoprotein production for NMR analysis

As key step for the analysis, FcεRIα was expressed in HEK293T cells, using a rich ^{13}C,^{15}N labeled medium, as previously described [9]. Using this strategy, the analysis of the glycan composition was carried out by using an NMR-based approach. HEK293T cells were chosen, since it is already known that they have the required glycosylation machinery to add all the possible sugars to the glycan chain in contrast to other types of cells (for example, CHO DG44 cells) that can build only certain glycoforms [10].

After the purification process, which is described in the corresponding section, a circular dichroism spectrum of the protein was recorded to evaluate the secondary structure of the FcεRIα (Figure 4.1).

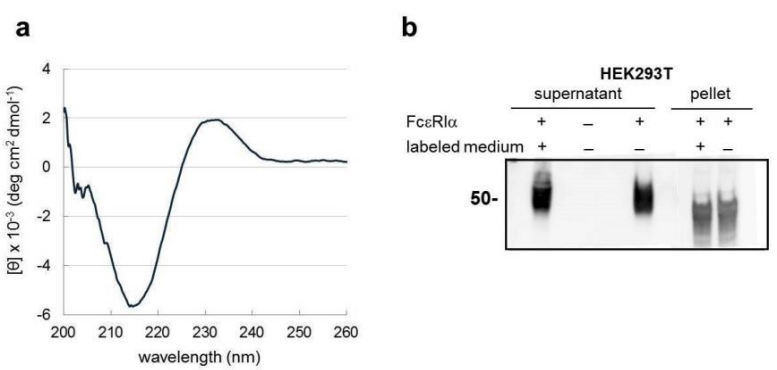


Figure 4.1: *a) Circular dichroism spectrum of soluble FcεRIα; b) western blot analysis of the FcεRIα induced expression in HEK293 cells. The protein was resuspended in 10 mM of phosphate saline buffer at pH 6.8 at the concentration of 5 μM, and the measurement was performed at 298K in a 0.02 cm path length cell.*

The CD spectrum is the same as that previously reported [11], showing the presence of a minimum at 216 nm and a maximum at 230 nm, with a predominant β-sheet structure, typical for Ig domains.

Moreover, the protein was also expressed using an unlabeled medium to evaluate whether the employed conditions could somehow modify the glycosylation output. A Western Blot analysis (Figure 4.1B) was conducted, using the His-tag strategically included at the expressed protein. In detail, the observed MW of the FcεRIα in lane 1 (supernatant from cells transfected with the labeled media) and in lane 3 (supernatant from cells transfected with the unlabeled media) is the same, showing that the glycosylation degree is independent of the growth media. Lane 2, which contains the supernatant from cells that have not been transfected is the control. Furthermore, cell pellets from labeled and unlabeled media (lanes 4 and 5, respectively) were also analyzed. Interestingly, the observed molecular weight is significantly lower, demonstrating that the glycosylation degree is different for the protein remaining inside the cell or for that secreted. This evidence is the consequence of the fact that the glycosylation process of FcεRIα depends on the protein localization. A protein that has suffered all the glycosylation's steps and has been eventually secreted displays a more complex degree of glycosylation than that remaining inside the cell, which shows an immature glycosylation pattern, arising from the machinery present in the endoplasmic reticulum.

Thus, using this protocol with the HEK293T cells, 360 µg of glycoprotein in 20 mL of culture was purified, with uniform labels in the glycan chains. The FcεRIα was then concentrated at 40 µM in 300µL and transferred to a Shigemi 5mm tube for further investigation by NMR spectroscopy.

The first step was the analysis of a standard ^{1}H-^{15}N-HSQC spectrum of the labelled FcεRIα. Basically, no cross peaks corresponding to the protein backbone amides were observed (Fig 4.2A). It has been described that this protocol allows the exclusive labelling of the glycan chains and the Ala residues of glycoproteins [12]. Thus, this simple experiment allowed to easily detect the NH moieties of the amine/amide-containing sugars (Fig. 4.2A). In a further step, the sialic acid residues were removed using neuraminidase A, an enzyme that is

able to cleave the glycosidic linkages involving all types of terminal sialic acid residues, which are usually attached to Gal units at positions O3 or O6. Some differences were observed (Figure 4.2B). Obviously, those cross peaks corresponding to the acetamide signals of the sialic acid moieties disappear, while new signals are now present, which probably belong to the newly formed terminal Gal residues. Finally, a cocktail of glycosidases, including Endo H (to hydrolyze the high-mannose-type glycans), and PNGase F (to cut the hybrid- and complex-type glycans) was employed. In this case, most of the protein precipitated, strongly suggesting that the glycans are essential for protein folding and stability (Figure 4.2C). Nevertheless, a spectrum could still be recorded, which displays the signals for the GlcNAc units directly attached to the protein.

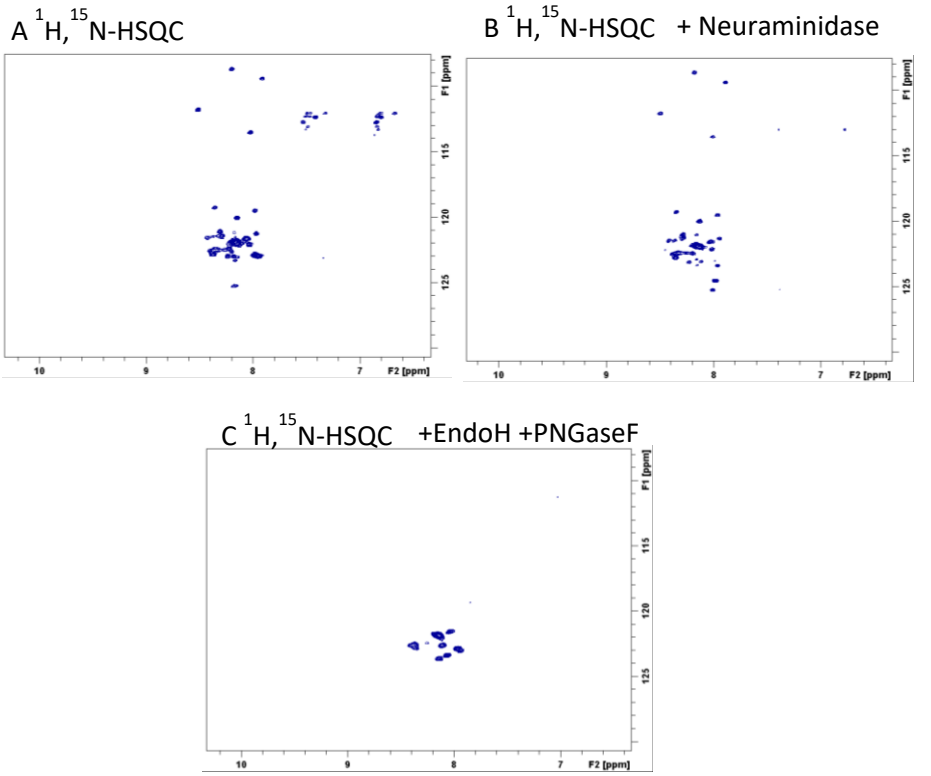


***Figure 4.2**: a) ^{1}H,^{15}N-HSQC of labeled FcεRIα; b) ^{1}H,^{15}N-HSQC of labeled FcεRIα after incubation with Neuroaminidase; c) ^{1}H,^{15}N-HSQC of labeled FcεRIα after incubation with EndoH and PNGase F*

4.3 NMR-based glycoprofile characterization of FcεRIα.

The analysis of the glycan was based on the presence of the ^{13}C-labeled glycans. As mentioned in the introduction, the ^{1}H-^{13}C HSQC spectra of glycans provide a specific cross peak pattern that can be employed as a fingerprint for identification purposes [12]. Although it is well recognized that the quality of the NMR spectra strongly depends on the size of the system scrutinized, due to relaxation issues [13] related to the rotational motion correlation time, the inherent flexibility of glycans allows for relatively fast relaxation in the ps/low ns timescale and allows the use of certain NMR experiments. The quality of the obtained NMR spectra can also be improved when using denaturing conditions. In this case, many intramolecular interactions are destroyed, including those between the glycans and the protein [14]. Therefore, the mobility of the glycan chains further increases, allowing the detection of additional sugar cross peaks.

Herein, the analysis of the glycans was carried out independently for both the native and denaturing conditions. In this second case, the protein was dissolved into 7M deuterated urea solution at pD 6.8.

An NMR-based approach, using heteronuclear 2D and 3D experiments, was implemented to characterize the glycan composition of FcεRIα. First, the ^{1}H,^{13}C-HSQC spectrum (fingerprint) was recorded (Figure 4.3). The analysis of the set of very well dispersed signals present in the anomeric region allowed discriminating certain sugar types and substitutions, which was further defined with the employment of ^{1}H,^{13}C-HSQC-TOCSY and ^{1}H,^{13}C-HSQC-NOESY experiments. Moreover, the comparison of the obtained data with those reported [15][16][17][18] for the diverse glycan epitopes present in N-glycans, including multi-antennae decorations, permitted the identification of the specific glycans in the glycoprotein.

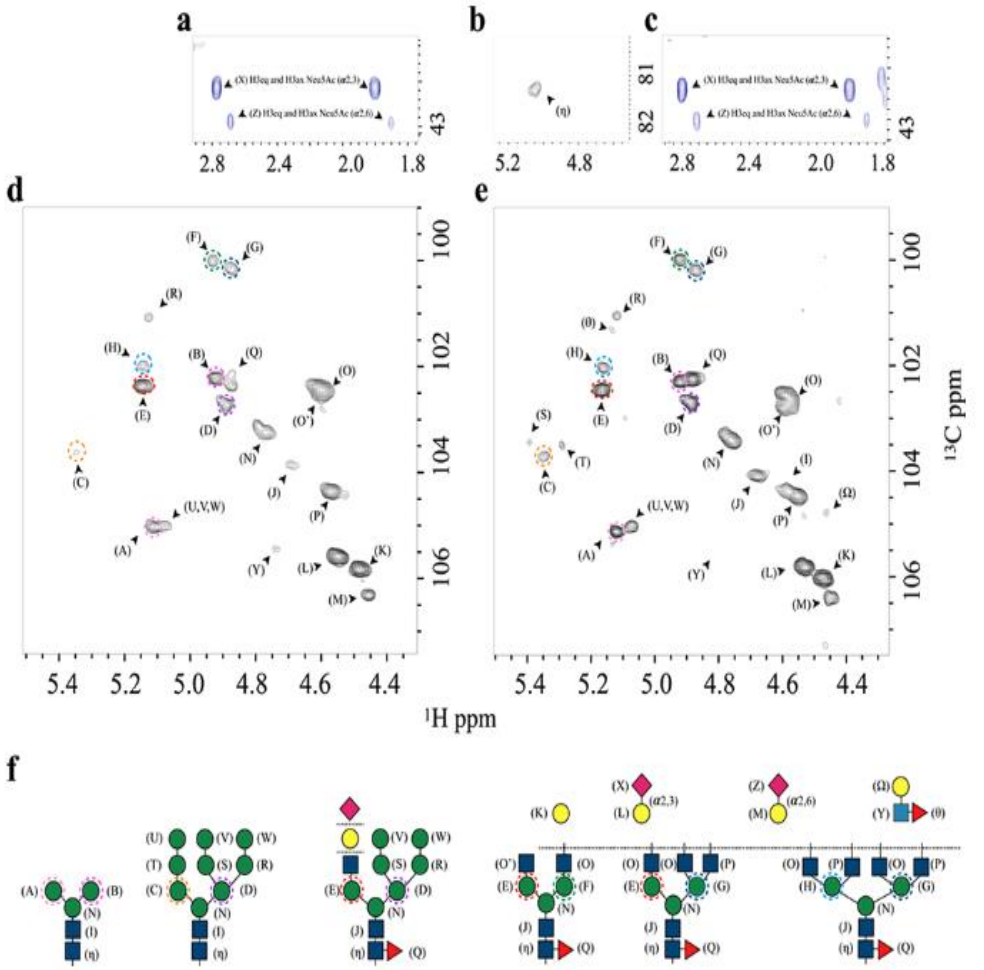


Figure 4.3: *Glycan composition. Expansion of the region containing the axial and equatorial protons of Sialic acid residues in folded (a) and unfolded state (c); (b) Expansion of the spectral region showing the anomeric (C1−H1) cross-peaks of the Asn-linked GlcNAc 1, which appears only under denaturing conditions. (d, e) Expansion of the spectral region showing the anomeric (C1−H1) cross-peaks of the linked saccharides in native and denaturing conditions, respectively. (f) Schematic representation of the different N-glycans that were identified*

Analogous experiments were carried out under denaturing conditions, in the presence of urea, to avoid miss-assignments due to chemical shift perturbations in the canonical glycan signals due to the protein interactions (Figure 4.3). Fittingly, the HSQC pattern for the anomeric region did not showed any difference between the folded and the denatured state. As result of the analysis, the presence of sialic acid (Neu5Ac), galactose (Gal), N-acetylglucosamine

(GalNAc), mannose (Man), and fucose (Fuc), linked in different manners, was assessed. The details of the assignment are shown in Figure 4.3 and Table 4.1.

Table 4.1: *Glycan 13C/1H chemical shifts assignment, referenced to TSP.*

Residue	Residue letter code	C1/H1	C2/H2	C3/H3	C4/H4	C5/H5	C6/H6,H6'	C7/H7	C8/H8	C9/H9,H9'
GlcNAc	η	81.3/5.05	56.7/3.83	74.5/3.71	81.6/3.74 81.8/3.92	78.0/3.74	68.8/3.94,3.79			
Fuc	Q	102.3/4.88	71.2/3.83	72.3/3.91	74.8/3.79	69.8/4.16	18.2/1.27			
GlcNAc	J	104.1/4.68	57.9/3.79	76.1/3.77	82.6/3.72	77.3/3.60	62.8/3.89, 3.77			
GlcNAc	I	104.4/4.59	57.9/3.79	76.1/3.77	82.3/3.71	77.3/3.60	62.9/3.89, 3.77			
Man	N	103.4/4.76	73.0/4.25	83.5/3.79	68.4/3.83 68.6/3.89	76.5/3.77	68.3/3.96,3.77			
Man	E	102.5/5.17	79.7/4.21	72.4/3.93	70.3/3.53	76.6/3.79	65.6/3.87,3.66			
Man	F	100.0/4.92	79.8/4.09	74.6/3.88	70.3/3.53	75.9/3.61	64.5/3.91,3.64			
GlcNAc	O	102.6/4.59	57.8/3.73	74.9/3.74	81.9/3.73	77.4/3.58	63.1/4.00,3.85			
GlcNAc	O'	102.8/4.59	57.8/3.73	-/-	72.8/3.48	78.9/3.46				
GlcNAc	P	104.5/4.55	57.8/3.73	-/-	81.9/3.70	77.4/3.58	63.1/4.00,3.85			
Man	G	100.2/4.87	80.0/4.08	74.5/3.90	70.4/3.43	75.7/3.62	68.8/3.95,3.79			
Man	H	102.0/5.16	79.1/4.24	72.3/3.91	80.9/3.65	74.8/3.75				
Gal	K	106.3/4.47	73.9/3.59	75.4/3.67	71.5/3.95	78.2/3.74	63.8/3.78,3.75			
Gal	L	105.8/4.53	72.2/3.61	78.5/4.11	70.3/3.97	78.2/3.74	63.8/3.78,3.75			
Gal	M	106.4/4.45	73.9/3.59	75.4/3.67	71.5/3.95	76.5/3.81	66.1/3.98,3.59			
Neu5Ac	Z			42.9/1.70,2.68	71.1/3.71	54.8/3.79	75.0/3.72	71.4/3.60	74.6/3.89	63.0/3.67,3.82
Neu5Ac	X			42.5/1.79,2.77	71.1/3.71	54.6/3.83	75.4/3.72	71.1/3.62	74.6/3.89	63.0/3.67,3.82
Man	A	105.1/5.12	72.9/4.07	73.3/3.85	69.7/3.66	68.5/3.72	62.8/3.83,3.63			
Man	B	102.3/4.92	72.8/3.99	73.3/3.85	69.7/3.69	68.5/3.72	62.8/3.83,3.63			
Man	C	103.7/5.35	81.3/4.12	72.8/3.99	71.0/3.79	76.3/3.77	63.9/3.89,3.87			
Man	D	102.7/4.89	72.4/4.14	81.4/3.93	68.6/3.85	73.5/3.885	68.8/3.95,3.79			
Man	R	101.1/5.12	81.4/4.03	72.8/3.99	71.0/3.79	75.8/3.68	63.9/3.89,3.77			
Man	S	103.5/5.39	81.3/4.12	72.8/3.99	71.0/3.79	76.3/3.77	63.9/3.89,3.77			
Man	T	103.5/5.29	81.3/4.12	73.0/4.01	71.0/3.79	76.3/3.77	63.9/3.89,3.77			
Man	U,V,W	105.0/4.75	73.0/4.09	73.5/3.89	71.1/3.72	76.3/3.77	63.9/3.89,3.77			
GlcNAc	Y	105.4/4.75	58.5/3.91	77.4/3.56	78.7/3.91	77.3/3.56	62.4/4.00,3.87			
Fuc	Θ	101.3/5.13	70.6/3.71	72.4/3.80	75.4/3.78	69.5/4.80	18.1/1.18			
Gal	Ω	104.8/4.46	74.6/3.65	73.9/3.50	71.1/4.15	77.7/3.72	63.9/3.74,3.71			

In particular, the presence of the Neu5Ac residues was evidenced by the exclusive cross peaks at high field for the axial and equatorial H3 protons. The existence of both α2,3- (X) and α2,6-linked (Z) Neu5Ac residues was shown given their different chemical shifts. In this manner, the Gal residues at terminal positions (K) were also discriminated from those with α2,3- (L) or α 2,6-sialyl

(M) substitutions. A similar analysis identified the fucose residues at the antennae (Θ) or at the inner core (Q).

The complete analysis showed different glycan structures, such as high mannose and pauci-mannose structure (bottom left, Figure 4.3f), hybrid (bottom central, Figure 4.3f) and complex N-glycans (bottom right, Figure 4.3f) types, with different degrees of fucosylation and sialylation.

The assignment of the anomeric cross-peaks was further validated by using enzymatic trimming, where the glycoprotein was sequentially incubated with different glycosidases, which are able to cleave specific external sugar moieties attached through certain glycosidic linkages. This approach also allowed assigning several of the partially overlapping peaks (Figure 4.4), which were now slightly shifted after the cleavage process, given their new chemical environment. The trimming enzymatic was performed using the folded protein using 10 mM phosphate buffer with 150 mM of NaCl, incubating for 1 hour at 310 K.

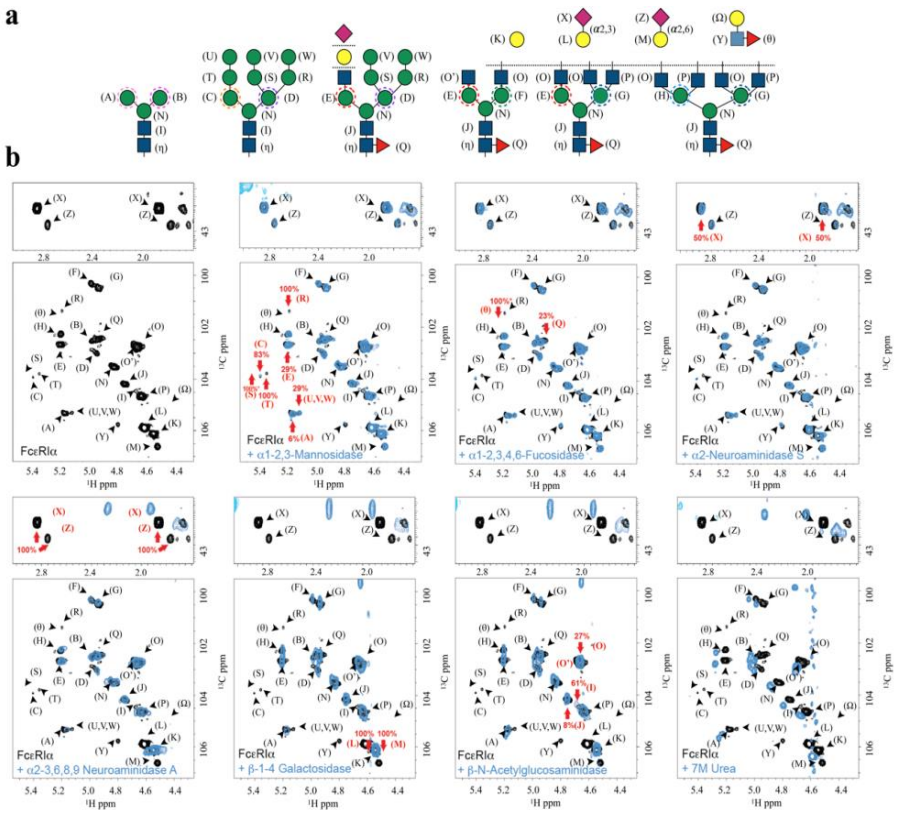


Figure 4.4: *Validation of the NMR assignment through the use of trimming enzymes. The protocol for validating the NMR assignment through the use of glycosidase enzymes. Mannosidases, fucosidases, neuraminidases, galactosidases, GlcNAcminidases were subsequently employed to trim the external mannose, fucose, sialic acids, Gal, and GlcNAc moieties, respectively. All the HSQC spectra, except for those corresponding to the treatment with α-1,4-galactosidase and α-glucosaminidase, were acquired with 1536 x 984 points. These two spectra were recorded using 1536 x 430 points. This lower resolution did not affect the chemical shift values nor the quantification.*

In detail, once the $^{1}H,^{13}C$-HSQC spectrum of the folded protein was recorded, 300µg of FcεRIα were incubated with 1µL of α1-2,1-3-mannosidase, a highly specific exoglycosidase that catalyzes the hydrolysis of external α1-2 and α1-3 linked mannose (Man) residues from the oligosaccharides. The new HSQC showed that more of the Man moieties were removed, suggesting that these residues are exposed on the protein surface. This enzyme has a specific action; indeed, Man D and Man F (α1-6 linkages) were not affected. As second step, a universal α1-2,3,4,6 fucosidase enzyme was employed to cut the fucose (Fuc) residues. The new recorded HSQC spectrum showed that the Fuc moieties on

the branches were eliminated (θ), while the internal Fuc residues (Q) remained, strongly suggesting that they were accessible to the enzymes. The trimming process was followed by the addition of 1µL of α2-3 neuroaminidase S, a highly specific exoglycosidase that catalyses the hydrolysis of α2-3 linked N-acetyl neuraminic acid residues. Half of the sialic acid residues (α2-3 linked) were removed. Upon addition of 1µL of α2-3,6,8,9 neuraminidase A (a broad specificity sialidase, which cleaves linear and branched non-reducing terminal sialic acid residues), the hydrolysis was now total. Finally, a β1-4 galactosidase and a β-N-Acetylglucosaminidase were sequentially used to remove the terminal Gal and GlcNAc moieties, respectively, as shown in Figure 4.4. Interestingly, some residues, as GlcNAc J, remained in the spectra, indicating that they are internal sugars that are not accessible by the enzyme. The quantification of the cross peak intensities was done after protein denaturation to avoid misinterpretation due to restricted motions at certain locations.

4.4 NMR analysis of the FcεRIα glycans under denaturing conditions

Many signals corresponding to similar epitopes in the N-glycans show certain degree of overlapping, underlining the existing structural similarity. However, the cross-peaks from the branched Man moieites (A-H) generate unique signals in the $^{1}H,^{13}C$-HSQC spectrum. Indeed, they show exclusive chemical environments in the different N-glycan species.

The quantification of the integrals of the cross-peaks corresponding to the different anomeric signals of the Man residues (A-H) was employed to describe the glycan heterogeneity on FcεRIα and the presence of diverse epitopes in relative terms. These signals were chosen since they are chemically comparable, they display similar J couplings and show analogous intrinsic relaxation proprieties.

A reference standard (TSP, 1 mM) to obtain the absolute quantification. In particular, the Man A value was used to estimate the amount of pauci mannose

species (Figure 4.5B), Man C for the high-Man N-glycans, Man F for the biantennary structures, and Man H for tetra-antennae. The hybrid- and triantennary complex-types (Figure 4.5B) were evaluated using the difference between Man D and Man C (D-C) and the difference between Man G and Man H, respectively. The percentage of each N-Glycan form was then guessed by dividing each volume by the sum of all volumes (Figure 4.5).

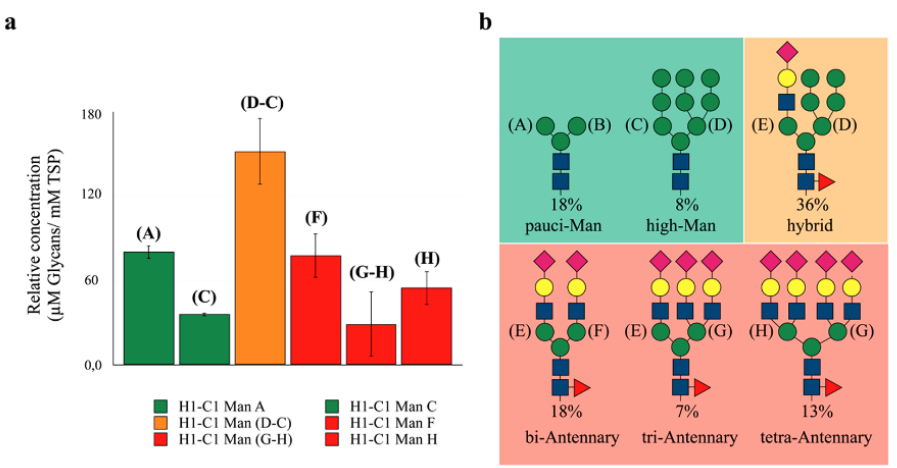


Figure 4.5: *Quantitative Glycoprofile analysis. (a) The relative concentration of the different glycoforms (protein concentration 60 µM was estimated by integrating the HSQC signals of the branched Man residues compared to internal reference TSP (trimethylsilylpropanoic acid 1 mM). Data were analyzed using three different samples. Error bars were calculated by standard deviation analysis. The quantification of the hybrid and triantennary glycans was carried out by subtracting the integrals arising from C to those of D, and H to G, respectively. The color code bar correspond to the N-glycans topology as represented in panel b.*

The analysis showed the presence of 110µM oligo Man glycans (26%, ~1.8 mol glyc/mol prot), 150µM hybrid-type (36%, ~2.5 mol glyc/mol prot), and 160µM complex-type (38%, ~2.7 mol glyc/mol prot). Among the complex-type, biantennary glycans were detected as the major components (80µM, 18% of the total), whereas the amount of tri- and tetra-antennae were also significant 30µM (7%) and 50µM (13%), respectively. Among oligo-Man N-glycans, a 76µM

(18%) concentration was identified as the simple pauci-mannose scaffold, while 34μM (8%) belong to the high-Man type (Figure 4.5).

Using this approach, the composition and proportion of specific glycan epitopes was also estimated. The anomeric cross peaks for the diverse Gal moieties (K, L, and M) were compared, indicating that 46% of the Gal residues were at non reducing terminal positions, 43% were α2-3 linked to a sialic acid, and 11% were α2-6 linked to the sialic acid moieties.

Fucosylation at the inner core or at the antennae was estimated from the relative intensities of the cross peaks of the anomeric protons of the GlcNAc residues. Specifically, 62% of the GlcNAc residues presented an inner core fucosylation (by comparing J and I cross-peaks). In comparison, only 3% showed Lewis-type antigens (by comparing Y intensities with those of O′, O, and P). The type of Lewis antigens (Lea, Leb, LeY, LeX) was deduced by looking for the characteristic downfield chemical shift of Fuc H5 (ca. 4.6 ppm). The comparison of the HSQC pattern in this region for FcεRIα showed a perfect match with that registered for the pure LeX antigen, strongly suggesting that the predominant Lewis antigen in the FcεRIα antennae is LeX (Figure 4.6).

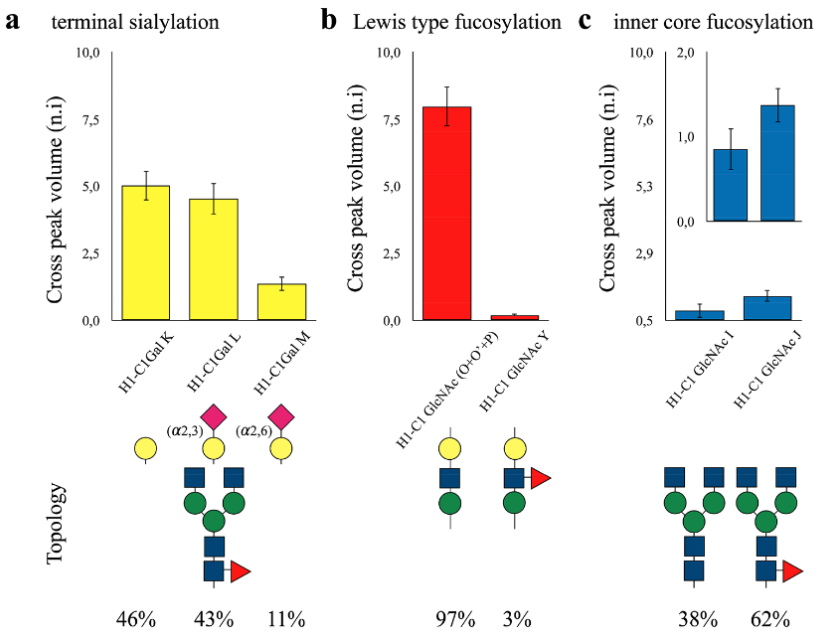


Figure 4.6: *Specific glycan epitope analysis. The different biorelevant N-glycans and epitopes contained in FcεRIα, together with their relevant population. (a) Estimation of the specific terminal sialylation obtained by comparing the cross-peak intensities of the galactose signals measured for the terminal Gal, Neu5Ac(α2−3)Gal, and Neu5Ac(α2−6)Gal moieties.35 (b) Estimation of the degree of fucosylation in the antennae as well as its relative abundance, calculated comparing (O' + O + P) with (Y). (c) Estimation of the degree of inner core fucosylation, calculated comparing (I) and (J).*

4.5 Assessing the presentation of the N-glycans and their dynamics features

As next step, the possibility that some of the glycans interact with the protein surface under the native folded conditions was firstly analysed by evaluating the chemical shift perturbations of the HSQC cross peaks under denaturing conditions compared to those in the folded state. Given the extreme sensitivity of the chemical shift to the chemical environment, if some of the glycans were interacting with the polypeptide chain should show chemical shift variations between both states, while the perturbations should be rather small if they are exposed to the solvent.

As shown in figure 4.7, the observed chemical shift perturbations of the glycan signals between the folded and the unfolded states were minimal, suggesting that they are mostly solvent exposed.

However, the semiquantitative analysis of the cross-peak intensities for the N-glycans attached to denatured FcεRIα using the same methodology described above for the native state, showed significantly different results to those described above for the folded glycoprotein. Very specially, the relative amount of high Man glycans increased more than twice, from 3.3% to 8.3%, while the other types showed less significant variations.

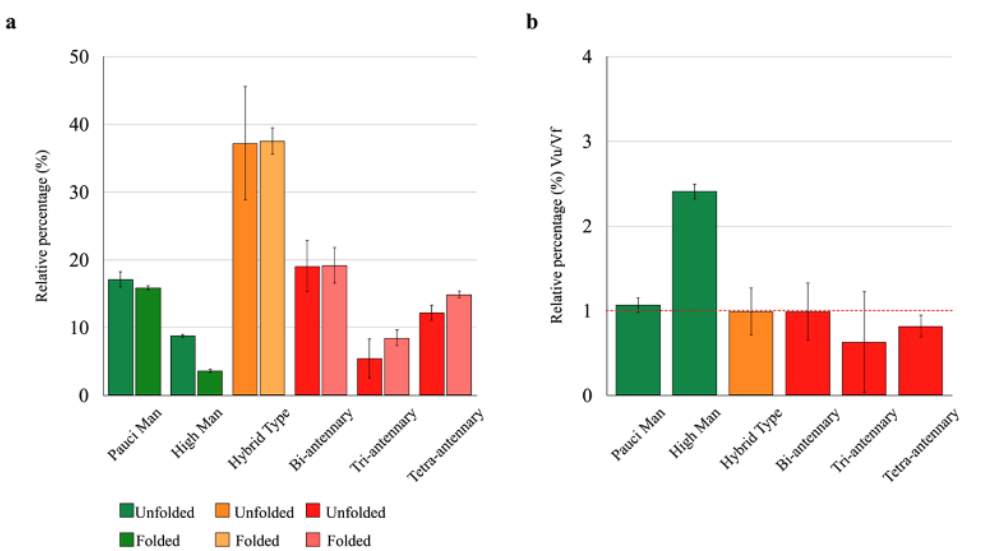


Figure 4.7: *Comparison of the relative population of N-glycan in FcεRIαunder denaturing and native-preserving conditions. (a) Relative percentage of the different types of N-glycans in FcεRIα, as determined in denatured and native conditions. Under denaturing conditions, the percentage always reflects the N-glycan composition while the apparent value obtained in native-preserving solvent is lowered when the N-glycan is interacting with the protein moiety. (b) Ratio of the population in the native and denatured states for the different N-glycan types, where the high mannoses show a statistically significant deviation from unity. Error bars are obtained from independent triplicate measurements.*

Additional experiments were carried out to assess the dynamic features of the glycans when attached to the protein. HSQC experiments were carried out at different temperatures (from 290K to 325K) and the signal intensities of different ^{1}H-^{13}C HSQC cross-peaks were measured. The intensities were

normalized using TSP as internal reference. As shown in Figure 4.8, the terminal residues were fairly more sensitive to the temperature changes compared to those closer to the protein. Indeed, there was a direct correlation between the sugar position in the N-glycan structure and their temperature susceptibility. Moreover, this type of analysis also allowed discriminating the Man moieties at the 1-3 and 1-6 branches. Indeed, Man F was less sensitive to temperature than Man E, probably due to the additional torsional degree of freedom present at the 1,6 arm.

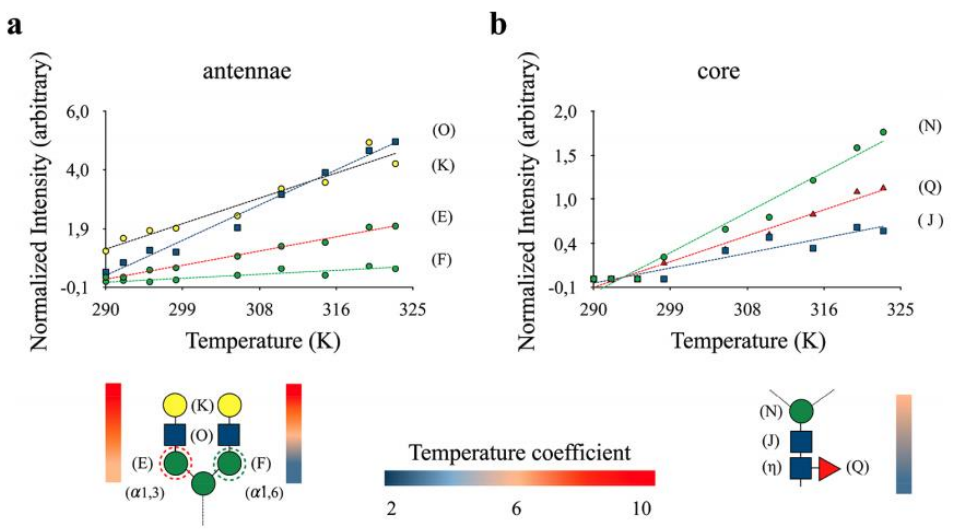


Figure 4.8: *Eximation of N-glycans dynamics. The plots represent the signal intensities measured for the anomeric C1−H1 HSQC cross peaks of each monosaccharide for a biantennary N-glycan, recorded as a function of temperature. (a) The antennae monosaccharides show intense resonances, which increase with temperature in a linear way. The differences among the sugar positions, at either the 1,3 or 1,6 arms, are also highlighted. (b) The inner-core sugars (Fuc, GlcNAc2, and Man3) show extremely weak signals.*

4.6 Towards an integrated 3D structural model of the FcεRIα glycoprotein

The NMR experiments described above have allowed elucidating the glycan complexity present at FcεRIα. However, a limitation of this analysis is that it does not provide a site-specific glycan assignment. Thus, as additional step, a model of the glycan presentation on FcεRIα was generated by integrating the

NMR experimental data with a computational approach. Firstly, the solvent-accessible surface area (SASA) was estimated for the different components, and the accessibility of the glycans on the protein surface was calculated. As key example, the SASA revealed that the residue N132 was the most protected glycosylation site. Indeed, it is located between the two Ig-like domains. Given the biosynthetic pathway established for the glycosylation of proteins in nature [1][19], it is highly likely that the carbohydrate processing enzymes are not accessible to this site and therefore, this N132 mostly retains its primary N-glycan decoration. Therefore, the presence of a high Man-type glycans was anticipated to be predominant at this locus. In contrast, the Asn residues at positions 71, 137 and 163 were very exposed on the protein surface. Therefore, it was hypothesized that these residues show the opposite glycosylation pattern and are decorated with complex type glycans. Finally, N47, N38 and N18 displayed intermediate SASA values. Thus, the presence of hybrid type glycans was proposed (Figure 4.9).

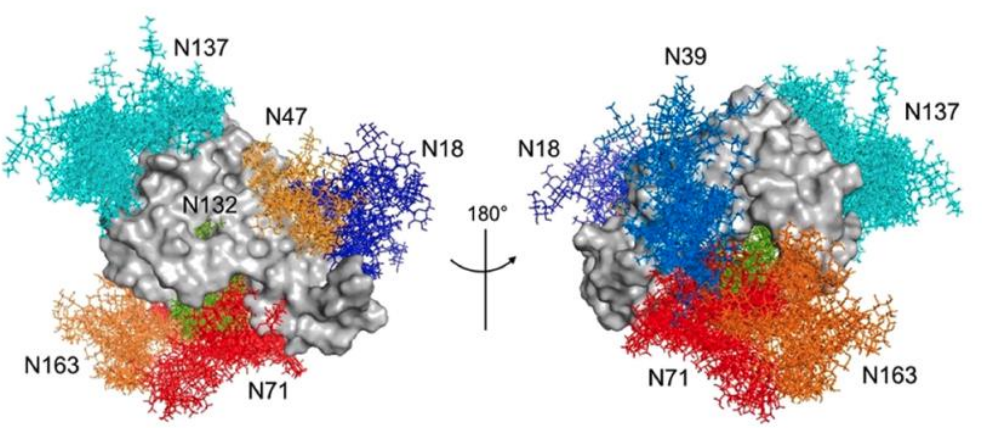


Figure 4.9: *A putative 3D model for FcεRIα glycoprotein. Front and back view for an ensemble of 10 conformers extracted from the 1 μs MD simulation. Structures were superimposed on the peptide backbone atoms. The N-glycans at each N-glycosylation site are coloured as follows: biantennary N-glycan at Asn18 (dark-blue); biantennary N-glycan at Asn39 (light-blue); hybrid N-glycan at Asn47 (light-orange); tetra-antennary N-glycan at Asn71 (red); high mannose N-glycan at Asn132 (green); tetra-antennary N-glycan at Asn137 (cyan); tetra-antennary N-glycan at Asn163 (orange)*

To verify the presence of high Man glycans attached to N132, a (N132A) mutant was designed and expressed. The glycan composition was then compared to that of the WT. In particular, the same amount of protein was obtained for the mutant, using the same experimental conditions than for the WT. Interestingly, a different CD profile was obtained for the N132A, which displayed a higher proportion of random coil structure. This fact strongly suggests that the glycan at N132 is essential for the correct protein folding in the native protein (Figure 4.10).

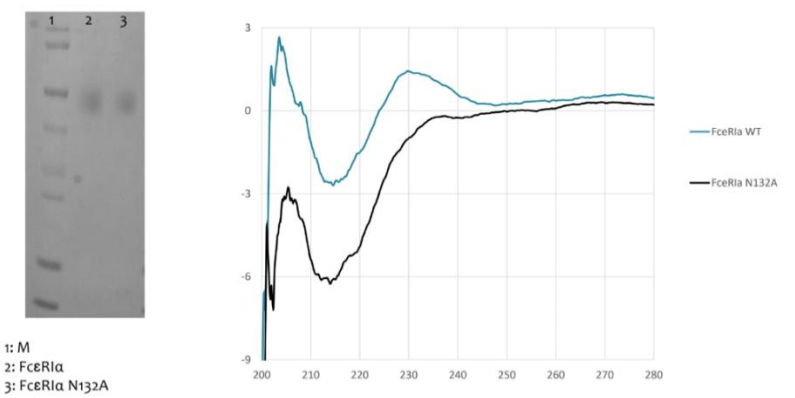


Figure 4.10: *a) SDS-PAGE gel of wt and mutant FcεRIα expressed in HEH293T cells; b) Circular dichroism spectrum of soluble N132A FcεRIα (black) and wt FcεRIα (blue). The profile is significantly different.*

The NMR-based analysis required recording the corresponding ^{1}H-^{13}C-HSQC experiment for the N132A glycoprotein. Fittingly, a major decrease or even disappearance of a variety of specific cross-peaks (C, R, S, T, U, W, V) corresponding to high Man N-glycan species was observed (Figure 4.11). Therefore, these experimental evidence confirm the presence of a high Man glycan attached to N132, supporting the putative N-glycan composition described above.

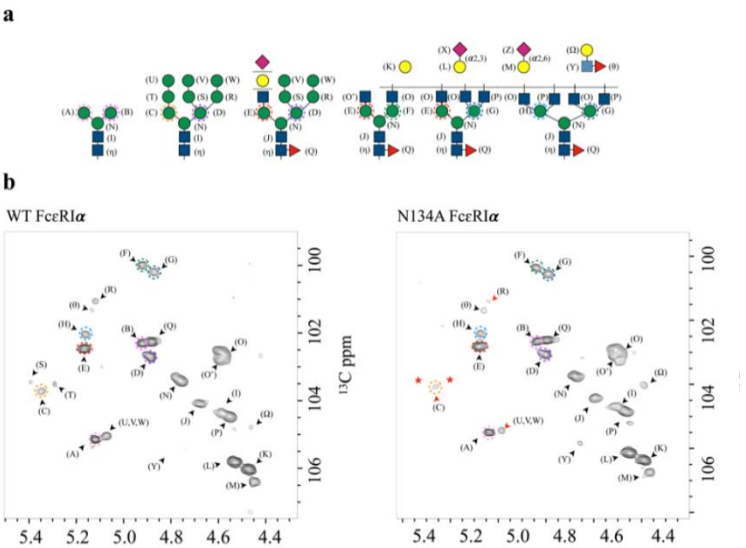


Figure 4.11: *Comparison of the anomeric region in the HSQC spectra recorded for the WT (left) and N132 (right). Many peaks belonging to High man glycans (see panel a) are clearly absent or show decreased intensities in the HSQC of the mutant between WT FcεRIα and N132A mutant*

4.7 Monitoring the direct interactions of intact FcεRIα with lectins by NMR

As mentioned in the introduction, lectins are carbohydrates binding proteins that recognize glycans on the cell surface or within it. The corresponding glycans are usually within glycoconjugates, either glycoproteins or glycolipids. Lectin-glycan binding events are of paramount importance in a variety of biological processes related to health and disease. The analysis of these interactions at the atomic/molecular level is normally achieved using diverse biophysical, chemical, and biochemical methods, employing isolated glycans either in solution or attached at surfaces. However, it is today evident that the presentation of the glycan epitope effectively recognized by the lectin is essential for the interaction to take place [18] There are key issues regarding entropy and dynamics that are affected depending on the actual orientation of the glycan. Therefore, we aimed at analysing the interaction of the FcεRIα glycans using the intact glycoprotein by NMR. As proof of concept, we chose

the carbohydrate recognition domain of Galectin-3 [20] as lectin model. It has been described [21] that the major glycan binding epitopes for hGal-3 are 3′-sialyl N-acetyl-lactosamine (3′SLN) and N-acetyl-lactosamine (LN). In contrast, the lectin does not 6′-sialyl N-acetyl-lactosamine. Since FcεRIα contain all these epitopes, this glycoprotein is an excellent probe to validate the methodology. Thus, ^{1}H-^{15}N-TROSY experiments for isotopically ^{15}N-labeled hGal-3 in the presence and absence of FcεRIα were recorded, and the observed variations in the lectin NMR signals were carefully monitored. Fittingly, the TROSY spectrum of hGal-3 in the presence of the FcεRIα showed significant reductions in many cross peak intensities. The peaks that showed the highest reductions were located in the β-sheets S2, S3, S4, S5, and S6; thus, at the canonical galactose binding-site (Figure 4.12). This evidence suggests that there is a lectin-glycoprotein interaction that exclusively involves the 3'SLN/LN glycan epitopes at FcεRIα.

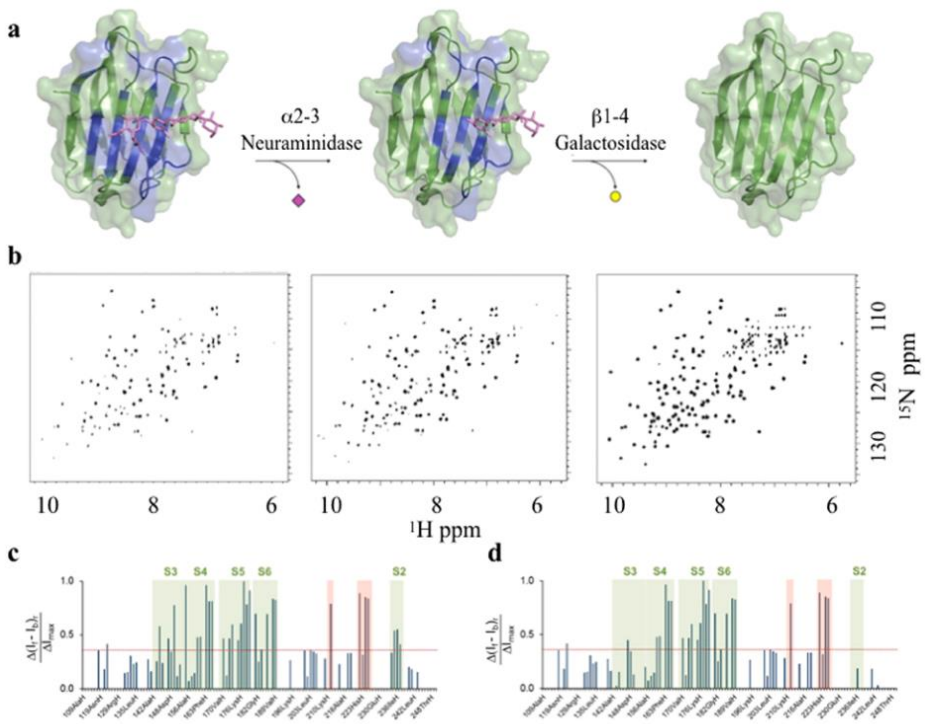


Figure 4.12: *Lectin binding experiments with intact FcεRIα. The recognition of specific exposed epitopes in FcεRIα by hGal3 was monitored by NMR signals intensity analysis of 15N-hGal3 in 1H−15N TROSY NMR experiments. (a) Model of a 3´SLN/hGal-3 interaction generated from the X-Ray crystallographic structure of hGal-3 bound to LN (PDB code: 1A3K) In this cartoon. The hGal3 residues affected by glycoprotein binding are mapped in blue. (b) From left to right, 1H-15N TROSY the binding between Gal-3 and glycans on FcεRIα was monitored by analyzing the variations in the intensities of the cross peaks of 15N. 1 H−15N TROSY NMR spectra of hGal3 in the presence of the complete FcεRIα glycoprotein (left), after enzymatic hydrolysis of the terminal α2,3-linked Neu5NAc moieties, and after subsequent hydrolysis of the β1,4-linked Gal residues. (c) Plot bars of the intensity differences per residue measured between apo hGal3 and FcεRIα-bound hGal3. (d) Plot bars of the intensity differences per residue measured between apo hGal3 and FcεRIα-bound hGal3 after treatment with α2−3 neuraminidase S. Specific residues show meaningful intensity differences and belong to the S2−S6 strands (green boxes). Additional residues located far away from the canonical carbohydrate-binding site are also perturbed (red box).*

To further identify the glycan epitope, FcεRIα was then incubated with neuraminidase S, which specifically cleaves sialyl α2-3Gal linkages. The new ^{1}H-^{15}N-TROSY recorded for hGal-3 in the presence of the treated glycoprotein without 3'SLN glycans showed the partial recovery of the signals of the cross peaks corresponding to the amino acids located at the S2-S3 strand. Interestingly, this locus exactly corresponds to the sialic acid binding pocket [22]. In contrast, the intensities residues in the S4, S5, and S6 remained

invisible. Finally, this glycoprotein sample (without 3'SLN epitopes) was further treated with a β-galactosidase to remove the external galactose residues that are essential for galectin binding. Fittingly, the typical TROSY spectrum observed for hGal-3 in the apo form was now obtained. Thus, only the LN/3'SLN epitopes take place in the interaction of FcεRIα with hGal-3. Finally, a putative 3D model of the interaction between both entities was generated. The PDB coordinates (1A3K) of the X-Ray crystallographic structure of the complex between LN with hGal-3 was used as template. One of the structures generated during the MD-based computational analysis of FcεRIα described above was used, choosing as ligand the lactosamine (LN) epitope of the complex-type glycan attached at N137. Both LN units (at the PDB and at the MD) were superimposed and the corresponding complex was optimized through a short MD protocol. The corresponding 3D structure is shown in Figure 4.13.

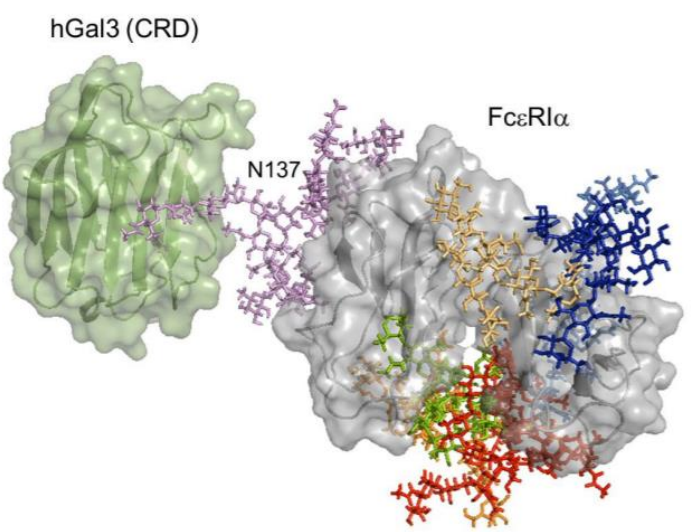


Figure 4.13: *Model for glycoprotein-hGal3 interaction. The model for the lectin-glycoprotein interaction was built by manual docking the N-glycan at Asn137 of the glycoprotein obtained from the MD simulation approach to the crystal structure of the human galectin 3 bound to N-Acetyllactosamine (pdb code 1A3K). Here the proteins are represented in cartoon with green surface (hGal3) and grey surface (FceRIa). The interaction is mediated by the tetra-antennary N-glycan at Asn137. All N-glycans are represented as sticks with unique color code: bi-antennary N-glycan at Asn18 (blue); bi-antennary Nglycan at Asn39 (light-blue); hybrid N-glycan at Asn47 (light-orange); tetra-antennary N-glycan at Asn71 (red); high-mannose N-glycan at Asn132 (green); tetra-antennary N-glycan at Asn137 (pink); tetraantennary N-glycan at Asn163 (orange).*

Therefore, this strategy allows monitoring direct binding of the glycans within intact glycoproteins to lectins and determining the precise glycan epitope involved in the molecular recognition process. demonstrated that glycans exclusively drove the binding between FcεRIα and hGal-3.

4.8 Conclusions

A novel NMR methodology has been presented to determine the glycan composition on an intact glycoprotein. In particular, the detailed analysis of the intrinsic glycan heterogeneity has been solved by combining a suite of NMR 2D and 3D experiments such as $^{1}H,^{13}C$-HSQC, $^{1}H,^{13}C$-HSQC-TOCSY, and $^{1}H,^{13}C$-HSQC-NOESY.

The NMR results, assisted by trimming experiments with glycosidase enzymes, have demonstrated that the FcεRIα glycoprofile entails different N-glycans, including high Man, hybrid and complex type. The specific assignment of the cross peaks of the HSQC to individual glycan epitopes has been achieved. A large variety of NMR signals that identify the presence of high Man, hybrid- and complex-type glycans have been identified in a non-ambiguous manner. These signals provide a fingerprint that can be useful to analyse to the glycan composition and of diverse glycoproteins.

The integration of the experimental NMR data with the results of a computational analysis has allowed generating a putative 3D model of FcεRIα that includes a proposal for the major glycosylation pattern at the different Asn-sites. Indeed, the expression and study of a N132A mutant has demonstrated the presence of a high Man glycan at this Asn site, along with importance of this glycosylation for providing structural stability to the glycoprotein. The relative dynamics features of the different glycans and epitopes has also been evaluated by NMR by using denaturing conditions.

Finally, a methodology for monitoring the interactions of intact glycoproteins to their lectin receptors has been proposed, using hGal-3 as model lectin. The

combined use of $^{1}H,^{15}N$-TROSY NMR experiments recorded for the ^{15}N-labeled lectin with the glycoprotein and other variants generated through sugar hydrolysis with specific glycosidases has allowed assessing the key sugar epitopes involved in the recognition event, along with the lectin recognition site.

Although this methodology has promising applications, there are some limitations. Firstly, the generation of the isotopically labeled (^{15}N and ^{13}C) proteins can be rather costly. Furthermore, although the heterogeneity and the glycan complexity can be analysed, the site-specific occupancy of the N-glycans at the protein remains elusive and cannot be directly assessed. Despite these limitations, a new methodology for the unambiguous detection of specific glycans signatures and their interactions is presented that can be applied to diverse biological problems involving the glycoproteins found for instance in tumour cells or other organisms, such as viruses and parasites.

This methodology has been used in the study of the RBD of the spike glycoprotein of SARS CoV-2, as shown below in the next Chapter V.

4.9 References

[1] K. T. Schjoldager, Y. Narimatsu, H. J. Joshi, and H. Clausen, "Global view of human protein glycosylation pathways and functions," *Nat. Rev. Mol. Cell Biol.*, vol. 21, no. 12, pp. 729–749. 2020, doi: 10.1038/s41580-020-00294-x.

[2] C. Reily, T. J. Stewart, M. B. Renfrow, and J. Novak, "Glycosylation in health and disease," *Nat. Rev. Nephrol.*, vol. 15, no. 6, pp. 346–366. 2019, doi: 10.1038/s41581-019-0129-4.

[3] B. Lin, X. Qing, J. Liao, and K. Zhuo, "Role of Protein Glycosylation in Host-Pathogen Interaction," *Cells*, vol. 9, no. 4, p. 1022,. 2020, doi: 10.3390/cells9041022.

[4] W. Morelle and J. C. Michalski, "Analysis of protein glycosylation by mass spectrometry," *Nat. Protoc.*, vol. 2, no. 7, pp. 1585–1602, 2007, doi: 10.1038/nprot.2007.227.

[5] A. Illiano, G. Pinto, C. Melchiorre, A. Carpentieri, V. Faraco, and A. Amoresano, "Protein Glycosylation Investigated by Mass Spectrometry: An Overview," *Cells*, vol. 9, no. 9, p. 1986. 2020, doi: 10.3390/cells9091986.

[6] G. Gomez, "Current Strategies to Inhibit High Affinity FcεRI-Mediated Signaling for the Treatment of Allergic Disease," *Front. Immunol.*, vol. 10, no., pp. 1–8,. 2019, doi: 10.3389/fimmu.2019.00175.

[7] M. C. Dispenza, B. S. Bochner, and D. W. MacGlashan, "Targeting the FcεRI Pathway as a Potential Strategy to Prevent Food-Induced Anaphylaxis," *Front. Immunol.*, vol. 11, pp. 1–8. 2020, doi: 10.3389/fimmu.2020.614402.

[8] S. C. Garman, S. Sechi, J.-P. Kinet, and T. S. Jardetzky, "The analysis of the human high affinity IgE receptor FcεRIα from multiple crystal forms11Edited by I. A. Wilson," *J. Mol. Biol.*, vol. 311, no. 5, pp. 1049–1062. 2001, doi: 10.1006/jmbi.2001.4929.

[9] L. Barbieri, E. Luchinat, and L. Banci, "Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells," *Nat. Protoc.*, vol. 11, no. 6, pp. 1101–1111. 2016, doi: 10.1038/nprot.2016.061.

[10] M. Shibata-Koyama *et al.*, "The N-linked oligosaccharide at Fc RIIIa Asn-45: an inhibitory element for high Fc RIIIa binding affinity to IgG glycoforms lacking core fucosylation," *Glycobiology*, vol. 19, no. 2, pp. 126–134. 2008, doi: 10.1093/glycob/cwn110.

[11] S. J. Demarest *et al.*, "An Intermediate pH Unfolding Transition Abrogates the Ability of IgE to Interact with Its High Affinity Receptor

FcεRIα," *J. Biol. Chem.*, vol. 281, no. 41, pp. 30755–30767. 2006, doi: 10.1074/jbc.M605190200.

[12] M. J. Rogals, J.-Y. Yang, R. V Williams, K. W. Moremen, I. J. Amster, and J. H. Prestegard, "Sparse isotope labeling for nuclear magnetic resonance (NMR) of glycoproteins using 13C-glucose," *Glycobiology*, vol. 00, no. 00, pp. 1–11. 2020, doi: 10.1093/glycob/cwaa071.

[13] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich, "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 23, pp. 12366–12371, 1997, doi: 10.1073/pnas.94.23.12366.

[14] M. Schubert, M. J. Walczak, M. Aebi, and G. Wider, "Posttranslational modifications of intact proteins detected by NMR spectroscopy: Application to glycosylation," *Angew. Chemie - Int. Ed.*, vol. 54, no. 24, pp. 7096–7100, 2015, doi: 10.1002/anie.201502093.

[15] J. Peng, S. M. Patil, D. A. Keire, and K. Chen, "Chemical Structure and Composition of Major Glycans Covalently Linked to Therapeutic Monoclonal Antibodies by Middle-Down Nuclear Magnetic Resonance," *Anal. Chem.*, vol. 90, no. 18, pp. 11016–11024, 2018, doi: 10.1021/acs.analchem.8b02637.

[16] A. Canales *et al.*, "Breaking the Limits in Analyzing Carbohydrate Recognition by NMR Spectroscopy: Resolving Branch-Selective Interaction of a Tetra-Antennary N-Glycan with Lectins," *Angew. Chemie - Int. Ed.*, vol. 56, no. 47, pp. 14987–14991, 2017, doi: 10.1002/anie.201709130.

[17] B. Fernández de Toro *et al.*, "Avenues to Characterize the Interactions of Extended N-Glycans with Proteins by NMR Spectroscopy: The Influenza Hemagglutinin Case," *Angew. Chemie - Int. Ed.*, vol. 57, no. 46, pp. 15051–15055, 2018, doi: 10.1002/anie.201807162.

[18] A. Gimeno *et al.*, "NMR and Molecular Recognition of N-Glycans: Remote Modifications of the Saccharide Chain Modulate Binding Features," *ACS Chem. Biol.*, vol. 12, no. 4, pp. 1104–1112, 2017, doi: 10.1021/acschembio.6b01116.

[19] P. Fisher, J. Thomas-Oates, A. J. Wood, and D. Ungar, "The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus," *Front. Cell Dev. Biol.*, vol. 7, pp. 1–11, 2019, doi: 10.3389/fcell.2019.00157.

[20] A. Gimeno *et al.*, "Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group

Antigens by Human Galectin-3.,” *Angew. Chem. Int. Ed. Engl.*, vol. 58, no. 22, pp. 7268–7272, 2019, doi: 10.1002/anie.201900723.

[21] S. R. Stowell *et al.*, “Galectin-1, -2, and -3 Exhibit Differential Recognition of Sialylated Glycans and Blood Group Antigens,” *J. Biol. Chem.*, vol. 283, no. 15, pp. 10109–10123,. 2008, doi: 10.1074/jbc.M709545200.

[22] P. M. Collins, K. Bum-Erdene, X. Yu, and H. Blanchard, “Galectin-3 Interactions with Glycosphingolipids,” *J. Mol. Biol.*, vol. 426, no. 7, pp. 1439–1451, 2014, doi: 10.1016/j.jmb.2013.12.004.

Chapter V

Structural characterization of the N-linked glycans in the RBD SARS-CoV-2 spike protein and their interactions with human lectins using NMR spectroscopy

5.1 Introduction

As described in the introduction, glycosylation represents one of the most abundant post-translational modifications in proteins in which a carbohydrate chain is linked to the polypeptide chain via an Asparagine moiety (N-glycosylation) or a Serine/Threonine residue (O-glycosylation) [1]. In the context of viral infections, it is well known that the surface of most viruses is decorated by a glycans that have a tremendous importance for the viral pathobiology. Indeed, it has been demonstrated that glycans are involved in the host recognition process, including the infection and the evasion to the host immune system [1]. Indeed, several viruses, such as HIV [2], Influenza [3], or coronavirus [4], use glycosylation to mask epitopes on their surface as a defence mechanism. In this manner, they occlude their receptor-binding domains (RBD) with the N-glycans, and the immune system cannot recognize the viral invasion [5].

Nowadays, we are devastated by the COVID-19 pandemic caused by the SARS-CoV-2 coronavirus that represents an enormous health and social problem [6]. The infection mechanism has been deciphered: the infection is mediated by the binding of the virus Spike glycoprotein (S) to the human angiotensin-converting enzyme 2 (ACE2) [7]. In particular, the spike interacts with the ACE2 through its RBD. It has been reported that the RBD displays 2 N-glycosylation sites that, in contrast to the evidence for influenza, MERS, or SARS CoV-1 infections [5][8][9], are not directly involved in the binding process with ACE [7][10][11], although other glycans at the spike may be involved in the proper presentation of the RBD to interact with hACE2 [12].

In a parallel manner, we have to consider that the human immune system counterattacks infection processes by different manners. One of them involves the expression of diverse lectins that modulate our immune response in a glycan-dependent manner [13].

In the last months, many efforts have been devoted to decipher the factors (including specific molecules) that may be targeted to avoid or impede the infection. Regarding the role of glycans, different investigations have attempted to clarify the role of sialic acids [14], heparan sulfate [15], or lectins in the infection process. In this scenario, different studies have identified some human lectins (MGL, DC-SIGN, Siglec-9, and Siglec-10) that may bind glycans on the RBD and modulate the viral pathogenesis [16][17].

The glycoprofile analysis of the Spike protein has been extensively studied using mass spectrometry [18][19][20][21][22]. However, as mentioned in the previous chapter, the MS-based methods involve digestion protocols that impede to carry out detailed molecular recognition studies. Ideally, interaction studies with glycoproteins should be carried out using procedures that only minimally alter the partners.

Herein, we have focused our attention on the study of the molecular recognition of the spike RBD with a variety of human lectins expressed in different organs that are affected in the COVID-19 infection. To this aim, RBD has been expressed with uniformly ^{13}C-labeled glycans at the two glycosylation sites (N331 and N343) using HEK293F cells grown with an excess of ^{13}C-glucose (Figure 5.1).

The glycan composition has been deduced using the methodology described in the previous chapter [23], and the fine details of the viral glycan-host lectin interactions have been elucidated at high resolution.

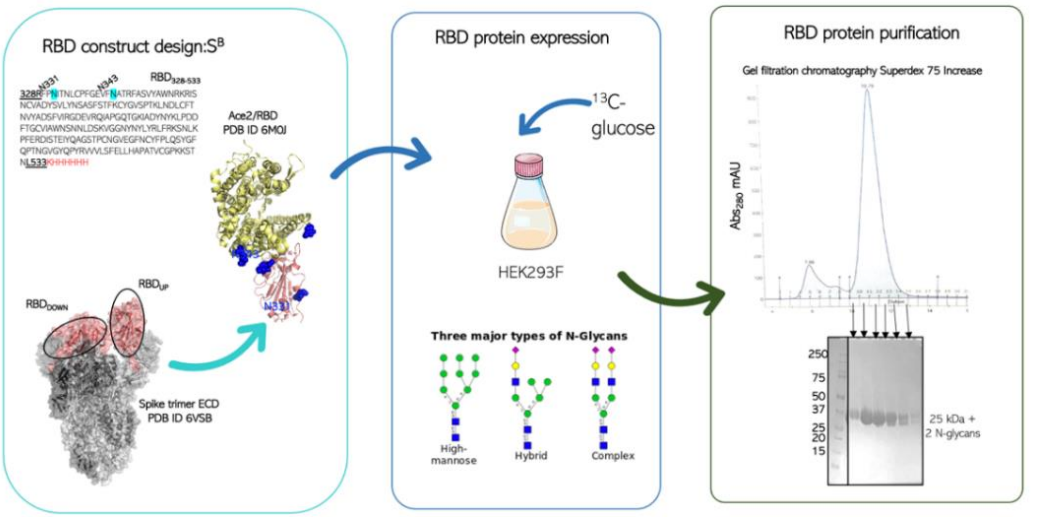
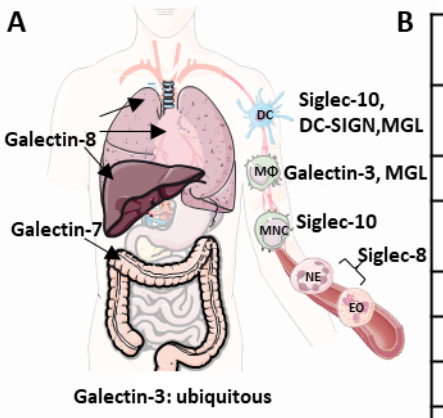


Figure 5.1: *Schematic representation of the RDB construct design, glycoprotein expression and purification. The SB domain (amino acid residues 328-533) of RBD, containing the two N-linked glycans at positions 331 and 343, was selected and expressed in suspension HEK293F cells. ^{13}C glucose was added to the cell media to produce labelled N-glycans on the RBD. Protein purification was achieved by two chromatographic steps: 1) affinity chromatography using His-Trap column that attach to the 6x His Tag at C-term of the RBD; and 2) gel filtration chromatography on Superdex 75 Increase column to separate the RBD monomers (big second peak at the chromatogram). The pure protein has a molecular weight of 35 kDa (SDS-PAGE gel).*

The chosen NMR-based strategy used two complementary approaches. On the one hand, the perturbations on the signals of the ^{13}C-labeled glycans on the RBD after adding a variety of human lectins were monitored using 2D $^{1}H,^{13}C$-HSQC NMR spectra. In particular, lectins expressed in organs affected by the SARS-Cov-2 infection were chosen, including Gal-3 [24], Gal-7 [25], Gal-8 [25], Siglec-8 [26], Siglec-10 [27], DC-SIGN [28] and MGL [29] (Figure 5.2). Additionally, the perturbations in the signals of the ^{15}N-labelled lectins upon addition of the RBD were evaluated using $^{1}H,^{15}N$-TROSY/HSQC spectra. Thus, using this protocol, a complementary view on their specific interactions was provided. Moreover, this approach allowed identifying the specific glycan epitopes on the RBD recognized by the corresponding lectin.



B

Lectin	Glycan preference	k_D (mM)
DC-SIGN	LeX	1
Galectin-7	LacNAc	0.4
Siglec-8	6´S sLeX	0.2
Galectin-3	LacNAc	0.05
MGL	GalNAc	0.01
Galectin-8	3´SLN	0.003
Siglec-10	S´LN	n.d

N-acetylglucosamine ▲ Fucose ● Glucose
○ Galactose ◆ N-acetylneuraminic acid S = SO_3^-

Figure 5.2: *A. Panel of human lectins employed herein along with their locations in human organs and tissues. B. Major glycan specificities and binding affinities for DC-SIGN, galectin-7, Siglec-8, galectin-3, MGL, galectin-8, and siglec-10 are given. Glycans are represented in SNFG symbols [30].*

5.2 Disentangling the glycoprofile of the RBD of the spike protein of SARS CoV-2 produced in HEK293F cells

The glycoprofile analysis of the two N-glycans on the RBD (N331 and N343) was performed using a combined analysis of 3D H',CH NOESY-HSQC, H',CH TOCSY-HSQC, and H'[C'],CH and [H']C',CH edited HSQC-[^{13}C,^{13}C]TOCSY-HSQC [23]. In particular, the [H']C',CH edited HSQC-TOCSY-HSQC was instrumental to assign all carbon resonances for every spin system (Figure 5.3C). Figure 5.3A displays the anomeric region of the ^{1}H-^{13}C-HSQC of the RBD. The corresponding assignment is gathered in Figure 5.3B.

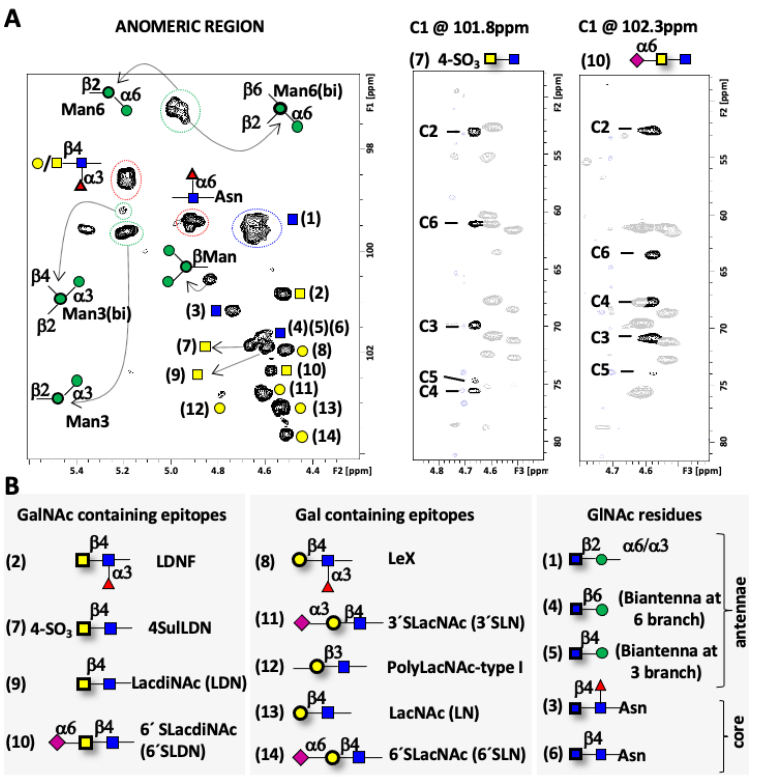


Figure 5.3: *NMR identification of glycan structures on SARS-CoV-2 RBD glycoprotein. A. Left: detail of the $^{1}H,^{13}C$-HSQC of RBD showing the assignment for most anomeric correlations. Anomeric correlations for Gal, GalNAc, and GlcNAc are identified with a number in brackets. Right: 3D [H']C',CH edited HSQC-TOCSY-HSQC of RBD, selected planes for C1 GalNAc on the 4SulLDN fragment and for C1 GalNAc on 6´SLDN, showing the correlations to all 13C atoms within the pyranose spin system. Nearby cross-peaks belonging to other spin systems have been veiled for clarity. B. GalNAc, Gal, and GlcNAc containing epitopes in N-linked glycans on RBD, represented as SNFG symbols.*

Interestingly, the analysis showed a high presence of Galβ1-4GlcNAc (LacNAc) moieties, including its participation as terminal residue (13), as well as substituted with a sialic acid linked via α2,3- (3'SLacNAc, 11) or α2,6- (6'SLacNAc, 13) linkages. Strikingly, a considerable amount of GalNAc containing epitopes was found. Indeed, it was evident the presence of terminal GalNAcβ1-4GlcNAc (LacdiNAc, 9), along with the existence of their α2,6-sialylated and 4-O-sulfated derivatives (6´SLDN, 10, and 4SulLDN, 7). In this context, it is worth mentioning that these epitopes were not identified in the MS analyses of the S protein [18][19][20][21][22].

Additionally, a high degree of fucosylation was observed, both at the core and at terminal positions. The Lewis X antigen was clearly recognized, while the presence of the analogous LDNF epitopes was also identified. Remarkably, the existence of this last epitope was unexpected since it is generally found on the surface of parasites [31]. The analysis also showed the presence of branching to give tri- and tetra antennary N-glycans and of type-I poly-LacNAc structures, although in smaller proportions.

Overall, this NMR-based analysis showed differences regarding the spike protein's glycosylation pattern that has been identified using other methodologies [18][19][20][21][22]. It is known that numerous factors can influence the observed glycosylation patterns [32][33]. One of these can be the adding of the ^{13}C-glucose into the expression medium. Thus, the human FcεRIα receptor was expressed using the same conditions employed for the RBD, and the glycan compositions were evaluated using ^{1}H-^{13}C-HSQC (Figure 5.4).

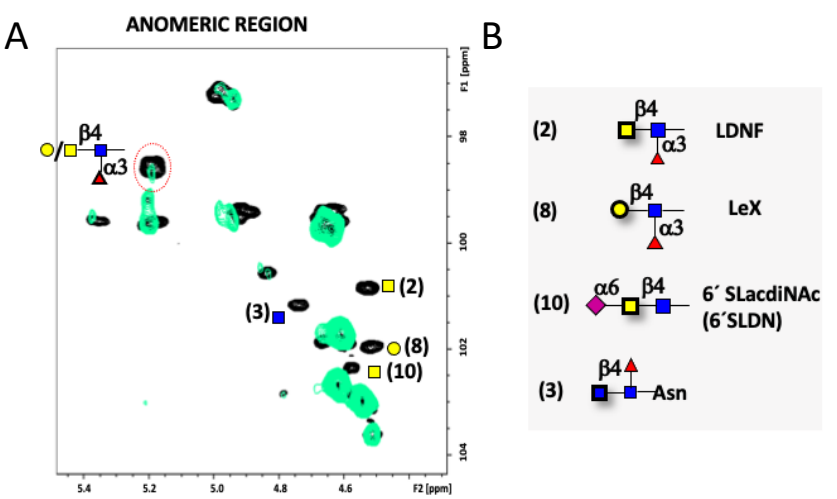
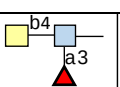
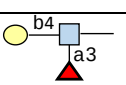
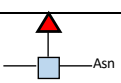
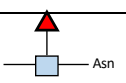
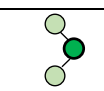
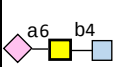
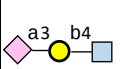
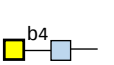
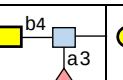
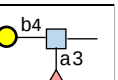


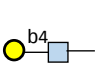
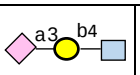
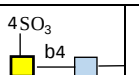
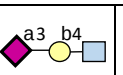
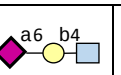
Figure 5.4: *Superimposition of the ^{1}H,^{13}C-HSQC spectra, at the anomeric region, of the two glycoproteins produced in HEK293 as described above. Black: RBD (amino acid residues 328-533) from the spike protein of SARS-COV-2. Green: the human FcεRIα. B: GalNAc, Gal, and GlcNAc containing epitopes in N-linked glycans on RBD, represented as SNFG symbols*

The comparison of the $^{1}H,^{13}C$-HSQC of the glycans on the FcεRIα and the RBD showed the lack of the LDNF, Le^{x} and 6'SLacdiNAc epitopes in the FcεRIα, as described in the previous chapter, when the protein was expressed in different experimental conditions [23]. This fact demonstrates that, at least in this particular case, the grown conditions does not influence the observed glycan profile. It is tempting to speculate that the intrinsic protein structure encodes the required information for the glycosylation outcome [32][33][34][35]. The obtained ^{1}H and ^{13}C chemical shifts for the RBD N-glycans at 310K are shown in Table 5.1.

Table 5.1: *^{1}H and ^{13}C chemical shifts assignation for the RBD N-glycans*

	 Fuc (LDNF)	 Fuc (Le^{x})	 Fuc (core) (a, major)	 Fuc (core) (b,minor)	 Man3	 Manβ
H1/C1	5.19/98.5	5.19/98.5	4.92/99.3	4.92/99.3	5.20/99.6	4.83/100.5
H2/C2	3.76/67.8	3.76/67.8	3.86/68.3	3.82/68.2	4.25/76.6	4.308/70.3
H3/C3	4.00/69.3	4.00/69.3	3.91/69.6	3.91/69.6	3.96/69.5	3.84/80.5
H4/C4	3.90/72.1	3.85/72.1	3.84/71.9	3.84/71.9	3.56/67.4	3.86/65.8
H5/C5	4.91/66.9	4.88/66.7	4.14/66.8	4,08/66.8	3.80/73.6	3.67/74.6
H6/C6	1.31/15.5	1.23/15.4	1.25/15.6	1.19/15.5	3.68, 3.96/61.8	4.02, 3.87/66.0

	 Man6	 GalNAc (2,6Neu5Ac)	 Gal (2,3Neu5Ac)	 GalNAc. (LDN)	 GalNAc (LDNF)	 Gal (Le^{x})
H1/C1	4.99/97.2	4.57/102.3	4.52/103.6	4.59/101.9	4.53/100.8	4.51/101.9
H2/C2	4.17/76.5	4.00/52.6	3.60/70.9	4.01/52.5	4.02/52.6	3.55/71.2
H3/C3	3.95/69.5	3.82/70.8	3.73/72.6	3.83/70.7	3.79/70.8	3.71/72.7
H4/C4	3.55/67.5	4.03/67.5	4.00/68.6	4.01/67.7	3.97/67.5	3.97/68.4
H5/C5	3.68/72.9	3.89/73.8	3.89/73.84	nd	3.64/74.9	3.65/75.0
H6/C6	3.68, 3.96/61.8	4.08, 3.64/63.5	4.05, 3.64/63.5	3.81/61.00	3.80/61.5	3.81/61.6

	 Gal terminal	 Gal (2,3Neu5Ac)	 4SO3-GalNAc	 Neu5Ac α2-3	 Neu5Ac α2-6	 GlcNAc (b1-4GlcNAc)
H1/C1	4.54/103.1	4.61/102.8	4.67/101.8			4.62/101.7
H2/C2	3.61/71.2	3.64/69.5	3.99/53.0			3.80/55.1
H3/C3	3.73/72.7	4.18/75.6	3.98/69.8	2.82, 1.85/39.7	2.73, 1.76/40.2	3.79/72.4
H4/C4	4.00/68.6	4.03/67.6	4.76/75.6	3.77/68.5	3.73/68.3	3.80/78.7
H5/C5	3.80/75.5	3.79/75.3	3.9/74.7	3.91/51.9	3.87/52.0	nd
H6/C6	3.81/61.1	3.81/61.1	3.83/61.0	3.72/73.0	3.77/72.7	4.06, 3.74/60.2
H7/C7				3.66/68.5	3.63/68.5	
H8/C8				3.96/71.9	3.96/71.9	
H9/C9				3.94, 3.72/62.8	3.94, 3.72/62.8	

5.3 Molecular recognition studies between RBD and human lectin

Once the glycoprofile of the RBD was deduced, its molecular recognition features to a panel of human lectins were evaluated. As described above, an NMR-based strategy was developed to decipher the process from two perspective: i) the RBD glycans and ii) the lectin.

From the perspective of the RBD glycans, the observed changes in the ^{1}H-^{13}C HSQC cross-peak intensities upon lectin's addition were monitored. Additionally, the perturbations in the NH signals of the ^{15}N-labelled lectins measured in the ^{1}H,^{15}N-TROSY/HSQC spectra in the absence and in the presence of the RBD were monitored (Figure 5.5).

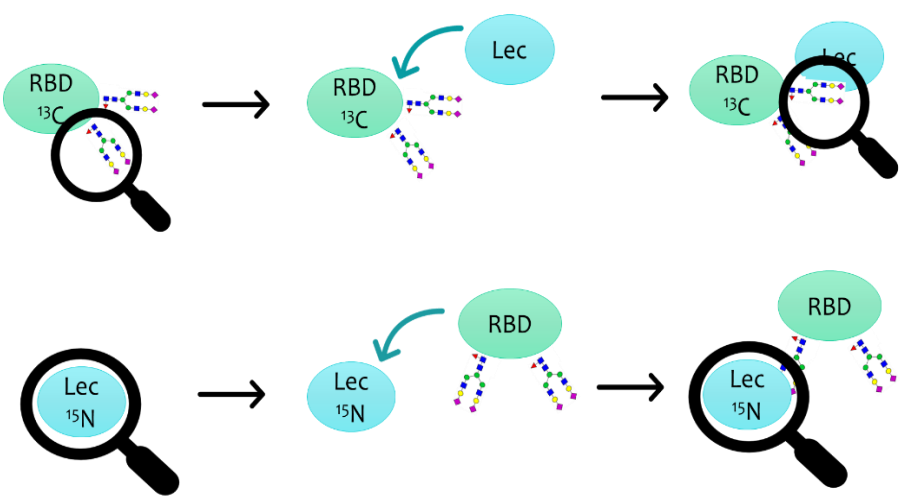


Figure 5.5: *Schematic representation of the protocol applied for the study of the interaction between the RBD and the human lectins.*

5.3.1 The interaction of the glycosylated RBD with Galectins

The carbohydrate recognition domains of Gal-3, full length Gal-7, and the N-terminal domain of Gal-8 were employed.

Regarding Gal-3, a selective reduction in the intensities of diverse cross-peaks was observed in the ^{1}H-^{13}C-HSQC spectrum of the RBD glycans (apo in black) after adding just 1 molar equivalent of Gal-3 (green, Figure 5.6).

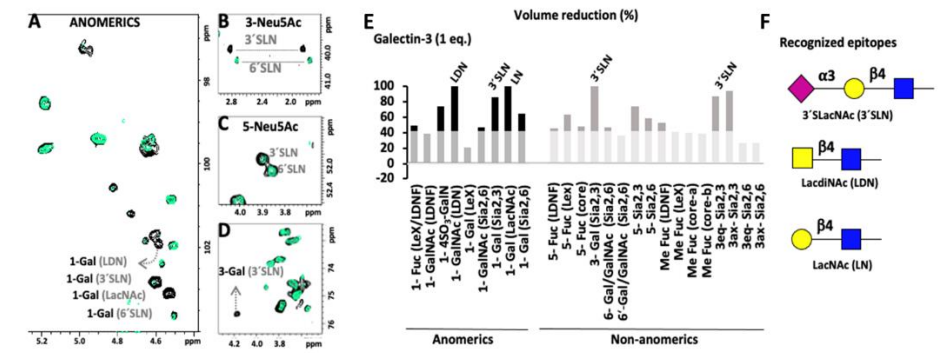


Figure 5.6: *NMR identification of glycan epitopes on the RBD recognized by galectin-3. Different regions of the $^{1}H,^{13}C$-HSQC spectrum of RBD alone (in black) and in the presence of 1 equivalent of galectin-3 (in green). A. Anomeric region: signals for terminal epitopes mostly affected by the presence of galectin-3 are annotated. B and C. Regions of the H3-C3 and H5-C5 (respectively) correlations of terminal Neu5Ac residues: signals for α2-3 linked residues completely disappear, while those for the α 2-6 linked are barely affected. D. Region showing the signal of H3-C3 Gal in 3´SLN epitope. E. The graphical bar representation for the % of volume reduction of selected cross peaks on $^{1}H,^{13}C$-HSQC of RBD upon addition of galectin-3. F: the glycan epitopes on RBD recognized by Gal-3, represented as SNFG symbols*

In the anomeric fingerprint region (Figure 5.6A), a decrease of specific cross-peaks corresponding to LDN, LacNAc, and 3'SLacNAc epitopes was evidenced. Moreover, a clear reduction in the H3-C3 and H5-C5 cross peaks assigned to α2-3-linked sialic acids was observed (Figure **Figure** 5.6B, C & D), while those belonging to α2-6-linked sialic acid were barely affected. The measured % reductions for all epitopes upon addition of Gal-3 is reported in the graphical bars.

From the lectin's perspective, upon addition of the RBD, the $^{1}H,^{15}N$-TROSY of ^{15}N-labeled Gal-3 showed a significant decrease of the intensities of the lectin signals, especially for those residues at the canonical Gal binding site (Figure 5.7**Figure 5.7**).

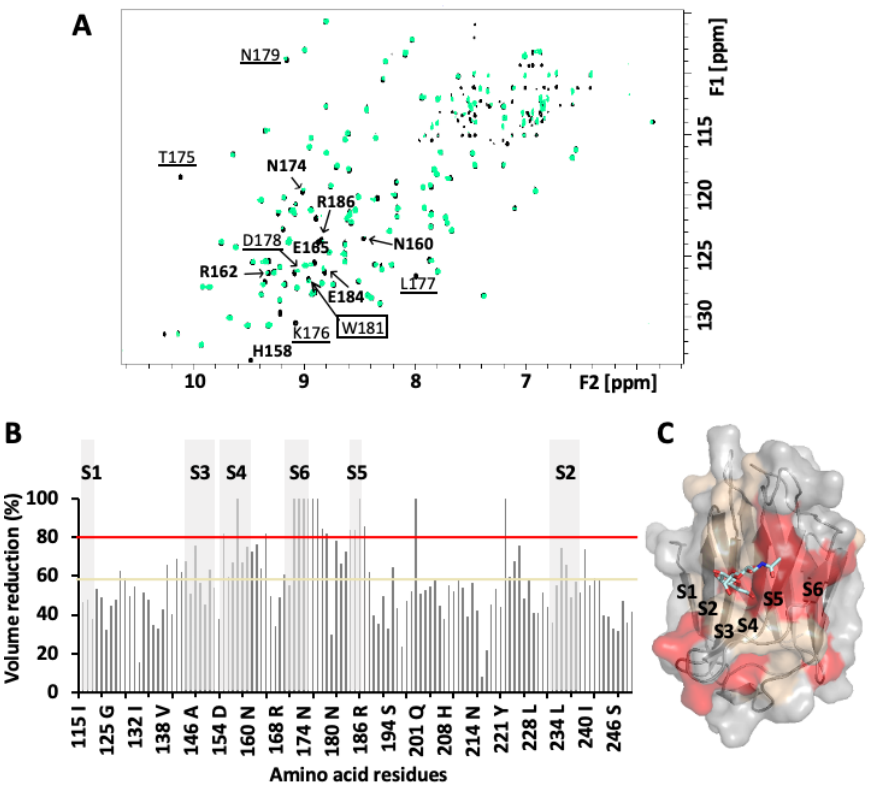


Figure 5.7: *Galectin-3/RBD interaction deduced by NMR from the lectin perspective. A. Superimposition of the $^{1}H,^{15}N$-TROSY for free galectin-3 (black) and galectin-3/RBD (green). Some affected cross-peaks are annotated: amino acids involved in LacNAc interactions are bolded; those interacting with α2-3 sialic acid are underlined. The key signal of W181 is squared. B. % cross peak volume reduction on the 1H,15N-TROSY upon addition of galectin-3. S1-S6 β-strands are depicted with grey boxes. Red and wheat horizontal lines are baselines for cross-peaks suffering 80-100% reduction and 60-80% reduction, respectively. C. Cartoon and surface representation of galectin-3 bound to LacNAc (PDB 1A3K) according to the X-Ray structure. Amino acids are colored based on their perturbation (%volume reduction) due to RBD binding (thresholds in B).*

The TROSY spectrum showed the disappearance of H158 and R186, which are 2 residues essential for the LacNAc binding [24]. Moreover, chemical shift perturbations for those residues in the S5-S6 β-sheet were evident. For instance, W181 is important for stabilizing the Gal with a CH-π staking interaction [24]. Additionally, the T175-N179 region was clearly perturbed, demonstrating the presence of 3'SLN partners at the RBD [24]. Indeed, this region is only affected by the 3'SLN epitope.

Thus, Gal-3 binds the RBD through the canonical Gal binding site and specifically recognizes terminal LN, LDN, and 3'SLN RBD epitopes.

For the N-term domain of hGal-8, the addition of just 1 equivalent of the lectin produced the instantaneous precipitation of the sample. Thus, the NMR analysis was carried out using just 0.2 molar equivalent of the lectin versus the RBD. The analysis of the data showed that only the signals corresponding to the 3'SLN RBD glycan epitope [36] were markedly affected in the presence of the lectin (Figure 5.8).

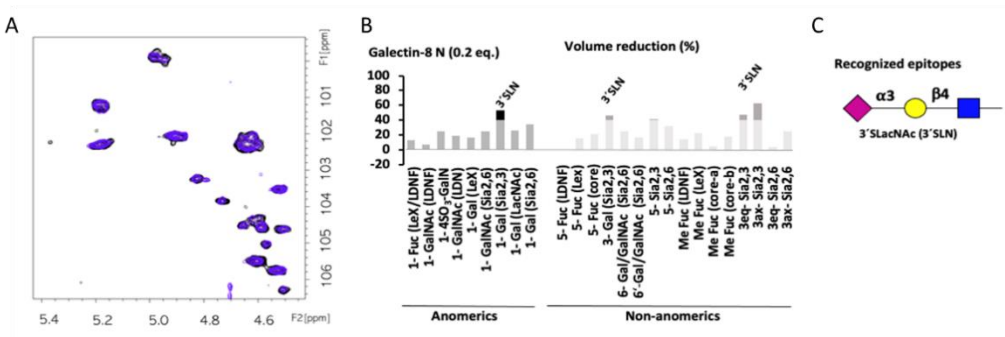


Figure 5.8: *NMR identification of glycan epitopes on the RBD recognized by galectin-8. A. Anomeric region of the $^{1}H,^{13}C$-HSQC spectrum of RBD alone (in black) and in the presence of 1 equivalent of galectin-8 (in purple). B. The graphical bar representation for the % of volume reduction of selected cross peaks on $^{1}H,^{13}C$-HSQC of RBD upon addition of galectin-8.C: the glycan epitopes on RBD recognized by Gal-8, represented as SNFG symbols*

From the lectin's point of view, 0.4 equivalents of unlabelled RBD were added to the solution of ^{15}N-labeled hGal-8 N-term (60µM). The $^{1}H,^{15}N$-TROSY showed a general decrease of the intensities of basically all lectin's cross peaks, strongly suggesting the formation of large supramolecular entities between the RBD and hGal8 N-term (Figure 5.9).

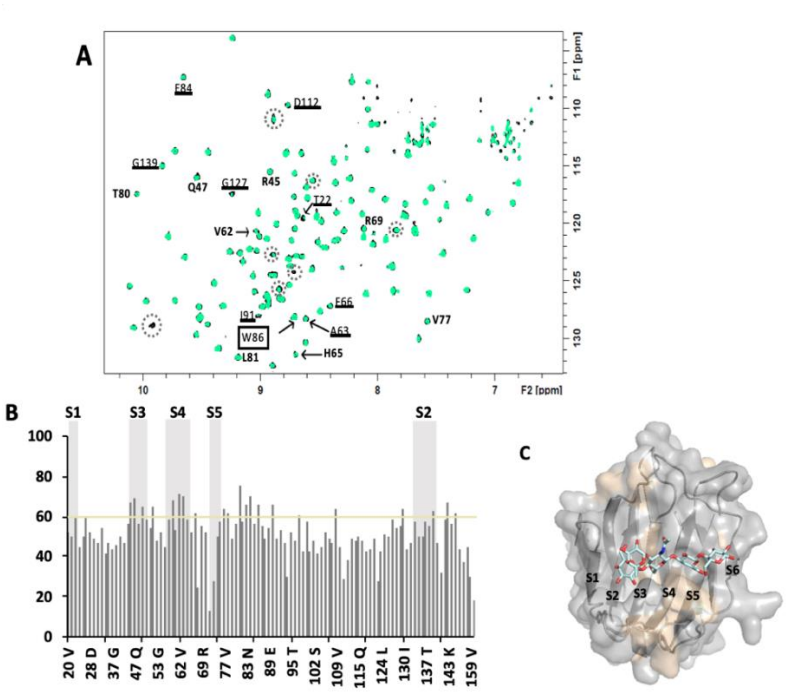


Figure 5.9: *The interaction between galectin-8N and the RBD as deduced by NMR from the lectin perspective. A. Overlay of the 1H,15N-TROSY spectra of galectin-8N (black) and in the presence of 0.4 equivalents of RBD (green). Broken lines indicate not assigned cross peaks. In bold, residues at the canonical glycan binding site and underlined, residues far from the binding site. The conserved Trp at the glycan binding site (W86) is squared. B. Graphical bar representation of the % cross peak volume reduction on the $^{1}H,^{15}N$-TROSY spectrum of galectin-8N upon the addition of 0.4 equivalents of RBD. S1-S6 β-strands are depicted with grey boxes. The wheat horizontal line indicates the NMR cross peaks suffering more than a 60 % of signal reduction. C. Cartoon and surface representation of galectin-8N complexed with 3'SLN (PDB ID 5G7F) according to the crystal structure. Amino acids are coloured based on their perturbation (% cross peak volume reduction) due to the presence of the RBD according to the threshold established in B.*

These results permit underline the different glycan binding preferences between both galectins towards the RBD in terms of epitopes and the different recognition phenomena that occur when the binding epitopes are differently exposed or hidden, especially in multivalent presentations [37].

Finally, the binding with Gal-7 was also analyzed. This galectin is a prototype galectin that forms a homodimer with two identical domains, which are connected through non-covalent interactions. This galectin displays

significantly low glycan affinities than other family members [36] [37] [38] [39] [40].

From the RBD's perspective, the LDN epitope was the most affected one (Figure 5.10), together with the LN and 3'SLN, as expected from the reported affinity data for this lectin [41].

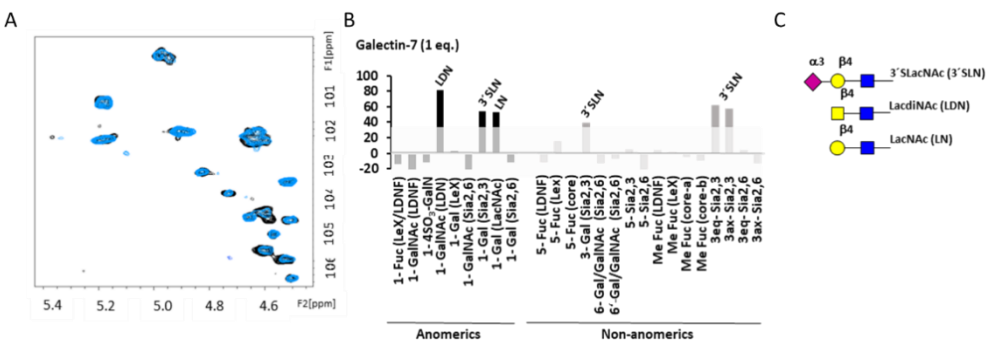


Figure 5.10: *NMR identification of glycan epitopes on the RBD recognized by galectin-7. A. Anomeric region of the $^{1}H,^{13}C$-HSQC spectrum of RBD alone (in black) and in the presence of 1 equivalent of galectin-8 (in blue). B. The graphical bar representation for the % of volume reduction of selected cross peaks on $^{1}H,^{13}C$-HSQC of RBD upon addition of galectin-7.C: the glycan epitopes on RBD recognized by Gal-7, represented as SNFG symbols*

From the lectin's point of view, a decrease in the intensities of the cross-peaks corresponding to the canonical binding site was observed. Nevertheless, other signals, corresponding to amino acids far from the binding site were also perturbed (Figure 5.11). These residues are located at the F face and at the dimer interface. These results suggest that all the lectin is somehow affected by the interaction, in agreement with previous reports that have proposed the existence of inter-domain communication in the presence of the ligands [41], as also reported for hGal-1 [42].

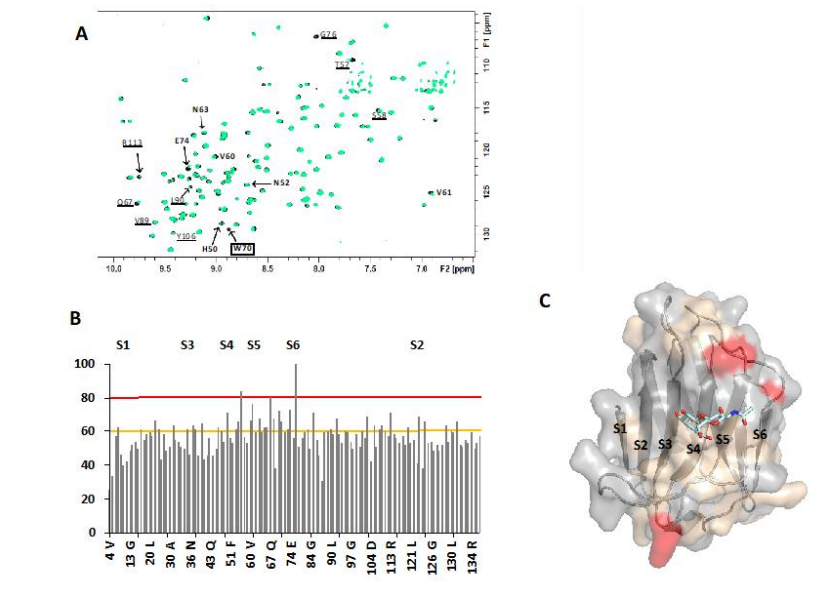


Figure 5.11: *Interaction between galectin-7 and the RBD as deduced by NMR from the lectin perspective. A. Superposition of the $^{1}H,^{15}N$-TROSY spectra of galectin-7 (black) and in the presence of 1 equivalent of RBD (green). In bold, residues at the canonical glycan binding site and underlined, residues far from the binding site. The conserved Trp at the glycan binding site (W70) is squared. B. Graphical bar representation of the % cross peak volume reduction on the $^{1}H,^{15}N$-TROSY spectrum of galectin-7 upon the addition of 1 equivalent of RBD. S1-S6 β-strands are depicted with grey boxes. The red and wheat horizontal lines, are baselines for NMR cross peaks suffering 80-100% signal reduction and 60-80% signal reduction, respectively. C. Cartoon and surface representation of galectin-7 complexed with LacNAc (PDB ID 5GAL) according to the crystal structure. Amino acids are coloured based on their perturbation (% cross peak volume reduction) due to the presence of the RBD according to the thresholds established in B.*

5.3.2 The interaction of the glycosylated RBD with Siglecs

In this proof-of-concept work, two different Siglecs were chosen, the V domain (CRD) of Siglec-8 and the whole extracellular domain of Siglec-10 (d1-d5), attached to the m-VENUS protein. Unfortunately, the CRD (V domain) of Siglec-10 is highly instable in the absence of glycosylation (when expressed in *E. Coli*). Therefore, Siglec-10 was expressed in a mammalian system, using a construct with m-VENUS, a chaperone protein that stabilizes proteins inside mammalian cells.

An analogous protocol to that described above for the galectins was carried out. When the two partners (Siglec-8 CRD and the RBD) were mixed, no changes neither on the glycan signals in the $^{1}H,^{13}C$-HSQC of RBD (Figure 5.12A) nor

in the $^{1}H,^{15}N$-TROSY of ^{15}N-Siglec-8 (Figure 5.12D) were observed. Thus, Siglec-8 does not recognize any glycan on the RBD. This fact is not unexpected, since it has been described that this lectin displays a tight glycan-binding selectivity for terminal 3´SLN and SLeX, but only when they are sulfated at Gal O6 [26]. Indeed, as deciphered by the glycoanalysis described above, this chemical modification is not present in the glycosylated RBD. Its presence would have beed readily identified due to the characteristic chemical shifts of the sulfated C6-Gal moiety.

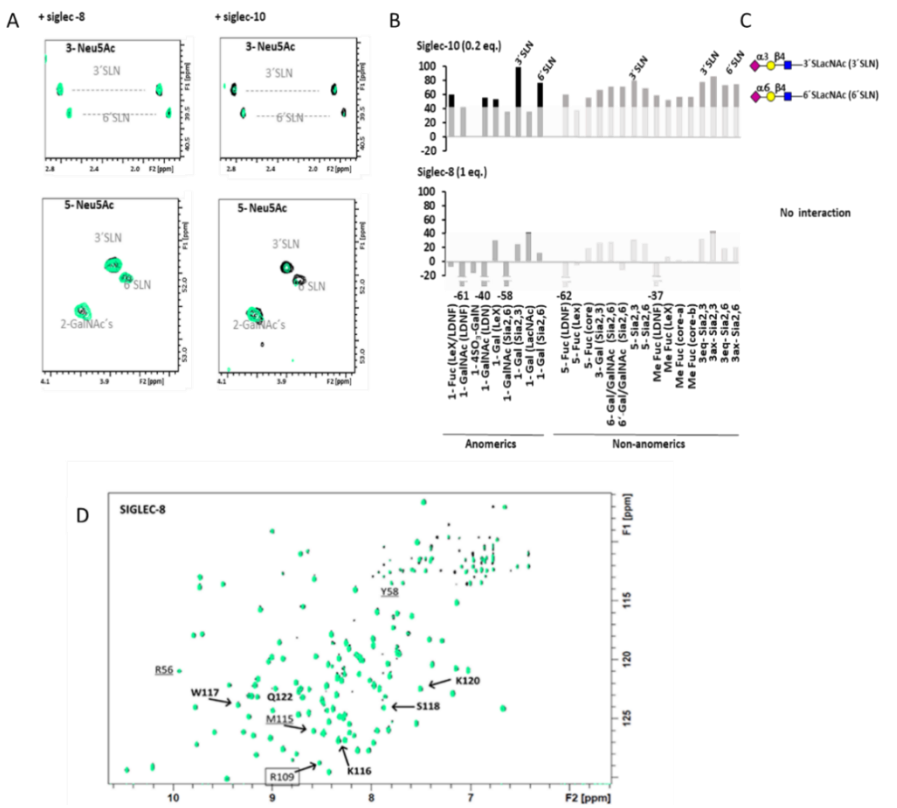


Figure 5.12: *The interaction of RBD with Siglecs -8 and -10 from the glycan perspective. Different regions of the $^{1}H,^{13}C$-HSQC of RBD alone (in black) and with 1 eq. of Siglec-8 (left, superimposed in green), and with 0.2 equivalents of Siglec-10 (right, superimposed in green). Top and middle: regions for the C3-H3 and C5-H5 correlations of Neu5Ac. In the presence of Siglec-8, no signal is affected, indicating that there is no interaction, while in the presence of Siglec-10, the signals of Neu5Ac, both α2-3 and α2-6 linked are affected, indicating that Siglec-10 interacts with the RBD through these epitopes. B. The graphical bar representation for % of volume reduction of cross peaks on the 1H,13C-HSQC of RBD upon adding Siglec-8 and -10. C. the glycan epitopes on RBD recognized by Siglec-10, represented as SNFG symbols D. Superposition of the $^{1}H,^{15}N$-TROSY spectra of siglec-8 (black) alone and in the presence of 1 equivalent of RBD (green). Some cross peaks are annotated. D. the glycan epitopes on RBD recognized by Siglec-10, represented as SNFG symbols*

A completely different situation was found for Siglec-10, for which the addition of 0.2 equivalents of the lectin to the solution containing the RBD caused a general reduction of the cross-peak intensities of the ^{13}C-labelled glycans in the HSQC. The effect was more pronounced for the signals of the terminal 3´SLN and 6´SLN epitopes (Figure 5.12C), in agreement with the reported selectivity for this lectin. Although a preference for 6'SLN over 3'SLN has been described, this is not appreciable from the NMR data gathered herein. In this case, the alternative information from the lectin perspective was not possible due to the lack of access to the ^{15}N-labelled lectin for NMR studies.

5.3.3 The interaction of the glycosylated RBD with C-type lectins: DC-SIGN & MGL.

The addition of 1 equivalent of DC-SIGN caused a selective intensity decrease on specific glycan ^{1}H,^{13}C-HSQC cross-peaks of the RBD (Figure 5.13). In terms of terminal epitopes, the LeX and LDNF signals were the most affected (Figure 5.13). This agrees with the reported preference of DC-SIGN for these moieties [43]. The effects produced in the cross peaks of the RBD glycans by adding 1 equivalent of MGL were markedly different (Figure 5.13), permitting to clearly identify the diverse binding preferences of both lectins. Indeed, the presence of MGL produced an exquisite selective reduction of the signals corresponding to GalNAc-containing epitopes, with the exception of 4SulLDN (Figure 5.13). Thus, terminal LDN, its α2-6 sialylated version (6´SLDN), and the fucosylated LDN (LDNF) are the glycans specifically recognized by MGL [44].

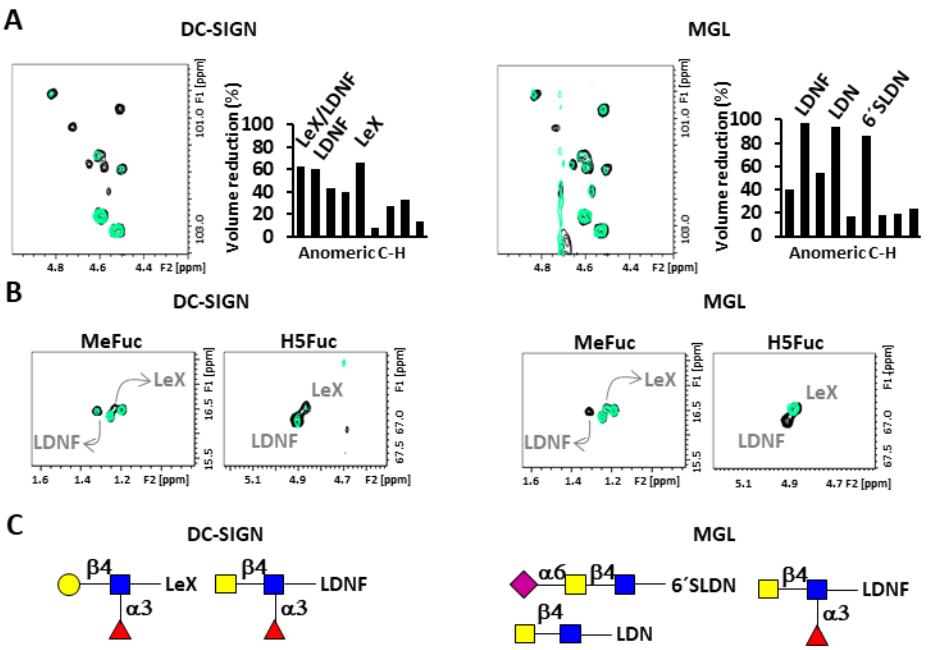


Figure 5.13: *Different interactions of the RBD with C-type lectins DC-SIGN and MGL from the glycan perspective. Selected regions of the $^{1}H,^{13}C$-HSQC of RBD alone (in black), with 1 eq. of DC-SIGN (left, superimposed in green), and with 0,2 equivalents of MGL (right, superimposed in green). A. Anomeric region and graphical bar representation for the % of volume reduction B. Specific regions for C6-H6 (Me) and C5-H5 correlations of Fuc in LDNF and LeX. C. Epitopes on RBD recognised by both lectins, as SNFG symbols.*

Alternatively, from the lectin's perspective, many cross peaks present in the $^{1}H,^{15}N$-TROSY of DC-SIGN (Figure 5.14) exhibited differential intensity loss upon addition of the RBD. The most affected residues are close to the calcium binding site, and are directly involved in interactions with the bound Fuc (N365, D366, N367, K368) [45]. Additionally, the NMR cross peaks for F313 and F374 were completely absent in the presence of the RBD, confirming the placement of Gal/GalNAc close to F313. Interestingly, a number of residues at the secondary calcium site (D320, L321, Q323, G325, T326 and W327) were also affected.

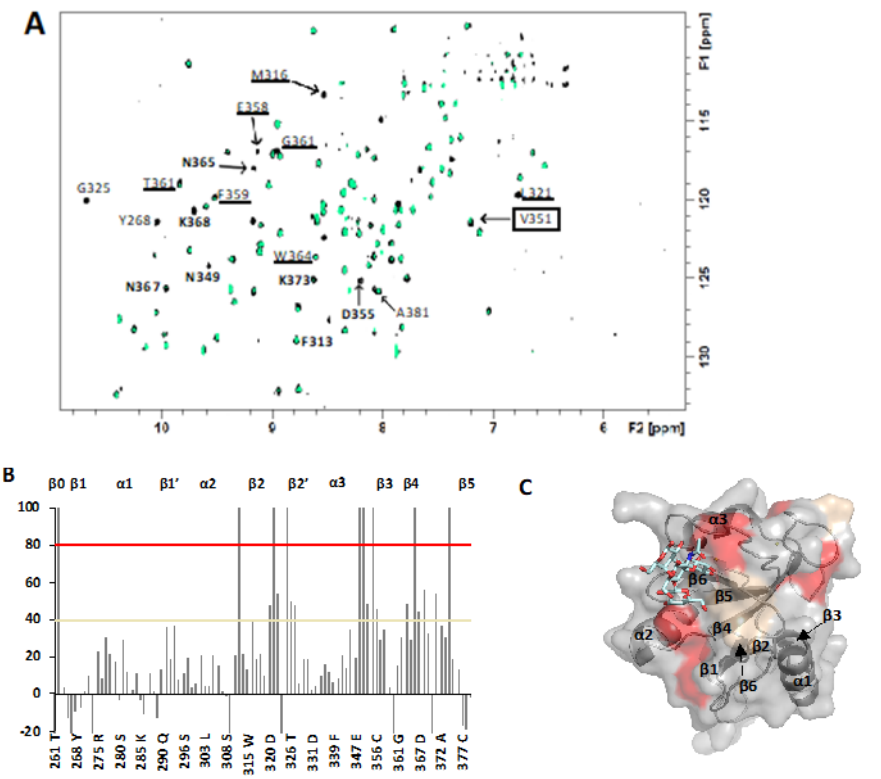


Figure 5.14: *Interaction between DC-SIGN CRD and the RBD as deduced by NMR from the lectin perspective. A. Superposition of the $^{1}H,^{15}N$-TROSY spectra of DC-SIGN (black) and in the presence of 1 equivalent of RBD (green). In bold, residues at the canonical glycan binding site and underlined, residues far from the binding site. The V351 residue at the edge of the long loop at the primary Ca^{+2} binding site is squared. B. Graphical bar representation of the % cross peak volume reduction on the $^{1}H,^{15}N$-TROSY spectrum of DC-SIGN upon the addition of RBD. β-strands and helices are depicted with grey boxes. The red and wheat horizontal lines, are baselines for NMR cross peaks suffering 80-100% signal reduction and 40-80% signal reduction, respectively. C. Cartoon and surface representation of DC-SIGN complexed with LeX (PDB ID 1SL5) according to the crystal structure. Amino acids are coloured based on their perturbation (% cross peak volume reduction) due to the presence of the RBD according to the thresholds established in B.*

The behaviour of MGL was completely different, reflecting the different dynamic properties of both lectins. The presence of 0.5 equivalents of RBD produced the homogeneous intensity reduction for most of the lectin, except the C-terminal fragment (Figure 5.15). Moreover, the addition of 1 RBD equivalent produced the complete disappearance of all the NMR signals in the $^{1}H,^{15}N$-HSQC (Figure 5.15C) and therefore, no direct information could be extracted. Thus, in order to confirm that the MGL glycan-binding site was indeed involved, a different strategy was devised and competition experiments, by

adding the simple interacting epitope: the GalNAc moiety, were performed (Figure 5.15D). Suitably, the addition of 1 equivalent of GalNAc produced the recovery of the NMR signals of the lectin, confirming that the RBD and GalNAc compete for the same binding site and that, therefore, the GalNAc epitopes at the RBD are directly involved in the binding event.

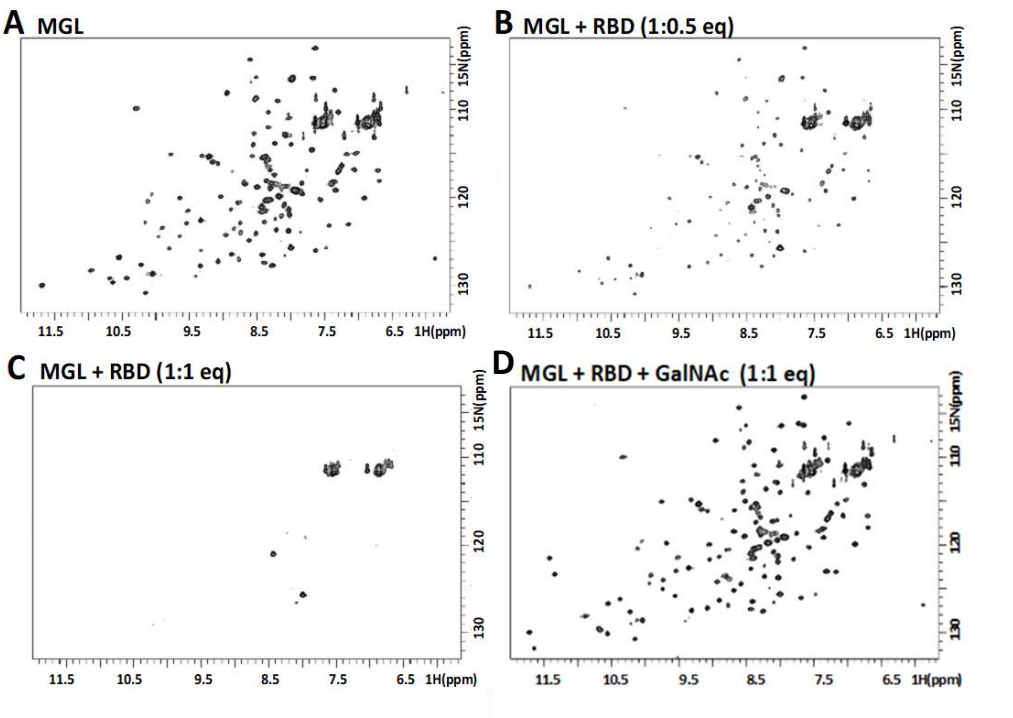


Figure 5.15: *A) $^{1}H,^{15}N$-HSQC spectrum of MGL CRD. B) $^{1}H,^{15}N$-HSQC spectrum of MGL CRD and 0.5 equivalent of RBD. C) $^{1}H,^{15}N$-HSQC spectrum of MGL CRD and 1 equivalent of RBD. D) Competition experiment. $^{1}H,^{15}N$-HSQC spectrum of MGL CRD with 1 equivalent of RBD and 1 equivalent of GalNAc.*

5.4 Conclusions

The RBD of the SARS-CoV-2 spike glycoprotein with ^{13}C glycan labeling has been generated. The great sensitivity provided by the ^{13}C atoms opens the door to significant opportunities for exhaustive NMR analysis of its glycoprofile and its molecular recognition features. Thus, by employing an NMR-based methodology, which avoids sample digestions and derivatizations, most of the

^{1}H and ^{13}C NMR glycan resonances of the intact (folded) glycoprotein in solution have been assigned, allowing to characterize the specific terminal glycan epitopes exposed on the antennae of the RBD N-glycans. Although the current analysis does not allow for fully quantitative occupancy determination and site-specific identification at N331 and N343, it has provided unprecedented structural details. Thus, besides the expected LN, 3´SLN, and 6´SLN terminal moieties, the presence of LDN and its fucosylated LDNF derivative have been assessed. Whereas the former has been detected in a trimer-stabilized version of the SARS-CoV-2 S protein, the presence of the LDNF epitope was unexpected. Indeed, LDN motifs have been found on several mammalian glycoproteins and observed in HEK293-produced glycoproteins [46].

In contrast, the LDNF epitope has been mainly related to pathogens [31]. Additionally, 4-O-sulfated and α2-6 sialylated LDN derivatives, not previously reported, together with the LeX antigen, have also been identified as terminal epitopes. Overall, the NMR analysis described herein highlights the presence of important levels of N-acetyl-galactosylation and hyper-fucosylation at the terminal chains of the RBD N-glycans, revealing glyco-epitopes not observed in previous MS-based analysis. Interestingly, the comparison with a different glycoprotein produced exactly under the same conditions suggests a relationship between the observed high levels of GalNAc and Fuc contents with the intrinsic protein structure. The exhaustive NMR analysis has also allowed disclosing the main N-glycan scaffold, being complex biantennary and core fucosylated, while lacking bisecting GlcNAc and a significant amaount of elongated antennas involving type I polyLacNAc sequences.

The interaction of the glycosylated RBD with a panel of human lectins has also been scrutinized. The ^{13}C-glycan labeling of the RBD has permitted to exploit the ^{1}H,^{13}C-HSQC spectrum of the RBD to report on the specific glycan epitopes recognized by each lectin, affording the corresponding glycan binding selectivity. Thus, while galectins-3 and -7 recognize the LN, LDN and 3´SLN motifs on the RBD, galectin-8-N seems to prefer exclusively the 3´SLN epitope.

Siglecs-8 and -10 demonstrated markedly differences, with Siglec-8 unable to recognize any of the glycan epitopes on the RBD, while Siglec-10 interacts with both 3´SLN and 6´SLN. For the C-type lectins, DC-SIGN exhibited selectivity for the two fucosylated terminal epitopes LeX and LDNF, while MGL showed exquisite selectivity for all GalNAc containing epitopes, except for the 4-O-sulfated derivative. The complementary information obtained from the ^{15}N-lectin based experiments permitted to assess that binding occurs through the canonical glycan binding site for each of the lectins (except for Siglec-10, which could not be analysed). Importantly, all the binding studies have been carried out by using the intact (folded) form of the RBD glycoprotein in solution, allowing to propose atomistic 3D models for the corresponding complexes.

This study paves the way to unveiling the interlaces roles of glycosylation patterns and cell receptors in SARS-CoV-2 infection mechanisms in the cell, particularly the recognition of tissue-dependent ACE2 by full-length glycosylated spike protein (S).

5.5 References

[1] B. Lin, X. Qing, J. Liao, and K. Zhuo, "Role of Protein Glycosylation in Host-Pathogen Interaction," *Cells*, vol. 9, no. 4, p. 1022,. 2020, doi: 10.3390/cells9041022.

[2] Z. T. Berndsen, S. Chakraborty, X. Wang, and C. A. Cottrell, "Visualization of the HIV-1 Env glycan shield across scales," 2020, doi: 10.1073/pnas.2000260117.

[3] S. Maurer-Stroh and P. C. Reading, "to Infection," pp. 1294–1316, 2014, doi: 10.3390/v6031294.

[4] Y. Watanabe *et al.*, "Vulnerabilities in coronavirus glycan shields despite extensive glycosylation," *Nat. Commun.*, vol. 11, no. 1, p. 2688. 2020, doi: 10.1038/s41467-020-16567-0.

[5] Y. Watanabe, T. A. Bowden, I. A. Wilson, and M. Crispin, "Exploitation of glycosylation in enveloped virus pathobiology," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1863, no. 10, pp. 1480–1497. 2019, doi: 10.1016/j.bbagen.2019.05.012.

[6] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, . 2020, doi: 10.1016/S0140-6736(20)30183-5.

[7] J. Shang *et al.*, "Structural basis of receptor recognition by SARS-CoV-2," *Nature*, vol. 581, no. 7807, pp. 221–224, 2020, doi: 10.1038/s41586-020-2179-y.

[8] I. Bagdonaite and H. H. Wandall, "Global aspects of viral glycosylation," *Glycobiology*, vol. 28, no. 7. Oxford University Press, pp. 443–467. 01, 2018, doi: 10.1093/glycob/cwy021.

[9] W. Van Breedam, S. Pöhlmann, H. W. Favoreel, R. J. de Groot, and H. J. Nauwynck, "Bitter-sweet symphony: Glycan-lectin interactions in virus biology," *FEMS Microbiol. Rev.*, vol. 38, no. 4, pp. 598–632, 2014, doi: 10.1111/1574-6976.12052.

[10] J. Lan *et al.*, "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor," *Nature*, vol. 581, no. 7807, pp. 215–220, 2020, doi: 10.1038/s41586-020-2180-5.

[11] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," *Cell*, vol. 181, no. 2, pp. 281-292.e6,. 2020, doi: 10.1016/j.cell.2020.02.058.

[12] L. Casalino *et al.*, "Beyond Shielding: The Roles of Glycans in the

SARS-CoV-2 Spike Protein," *ACS Cent. Sci.*, vol. 6, no. 10, pp. 1722–1734,. 2020, doi: 10.1021/acscentsci.0c01056.

[13] R. L. Schnaar, "Glycans and glycan binding proteins in immune regulation: A concise introduction to glycobiology for the allergist," *J Allergy Clin Immunol.*, vol. 135, no. 3, pp. 609–615, 2015, doi: 10.1016/j.jaci.2014.10.057.

[14] A. N. Baker *et al.*, "The SARS-COV - 2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device," 2020, doi: 10.1021/acscentsci.0c00855.

[15] L. Liu *et al.*, "Heparan sulfate proteoglycans as attachment factor for SARS-CoV-2," *bioRxiv Prepr. Serv. Biol.*, 2020, doi: 10.1101/2020.05.10.087288.

[16] F. Chiodo *et al.*, "Novel ACE2-Independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota," *bioRxiv*, p. 2020.05.13.092478, 2020, doi: 10.1101/2020.05.13.092478.

[17] M. Thépaut *et al.*, "DC / L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist," pp. 1–34.

[18] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, and M. Crispin, "Site-specific glycan analysis of the SARS-CoV-2 spike," *Science (80-.).*, vol. 369, no. 6501, pp. 330–333, 2020, doi: 10.1126/science.abb9983.

[19] J. D. Allen, H. Chawla, F. Samsudin, and L. Zuzic, "Site-specific glycosylation steric control of," pp. 1–37, 2021, doi: 10.1101/2021.03.08.433764.

[20] A. Shajahan, N. T. Supekar, A. S. Gleinich, and P. Azadi, "Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2," *Glycobiology*, vol. 30, no. 12, pp. 981–988. 2020, doi: 10.1093/glycob/cwaa042.

[21] Y. Zhang *et al.*, "Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins using High-Resolution Mass Spectrometry," *bioRxiv*, p. 2020.03.28.013276, 2020, doi: 10.1101/2020.03.28.013276.

[22] P. Zhao *et al.*, "Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor 3," *bioRxiv*, p. 2020.06.25.172403,. 2020, doi: 10.1101/2020.06.25.172403.

[23] L. Unione *et al.*, "Glycoprofile Analysis of an Intact Glycoprotein As Inferred by NMR Spectroscopy," *ACS Cent. Sci.*, vol. 5, no. 9, pp. 1554–1561, 2019, doi: 10.1021/acscentsci.9b00540.

[24] A. Gimeno *et al.*, "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.*, vol. 58, no. 22, pp. 7268–7272, 2019, doi: 10.1002/anie.201900723.

[25] J. Iwaki and J. Hirabayashi, "Carbohydrate-Binding Specificity of Human Galectins: An Overview by Frontal Affinity Chromatography," *Trends Glycosci. Glycotechnol.*, vol. 30, no. 172, pp. SE137–SE153,2018, doi: 10.4052/tigg.1728.1SE.

[26] J. M. Propster *et al.*, "Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 29, pp. E4170-9. 2016, doi: 10.1073/pnas.1602214113.

[27] L. D. Movsisyan and M. S. Macauley, "Structural advances of Siglecs: insight into synthetic glycan ligands for immunomodulation," *Org. Biomol. Chem.*, vol. 18, no. 30, pp. 5784–5797, 2020, doi: 10.1039/D0OB01116A.

[28] K. Pederson, D. A. Mitchell, and J. H. Prestegard, "Structural characterization of the DC-SIGN-LewisX complex," *Biochemistry*, vol. 53, no. 35, pp. 5700–5709. 2014, doi: 10.1021/bi5005014.

[29] F. Marcelo *et al.*, "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.*, vol. 20, no. 49, pp. 16147–16155,. 2014, doi: 10.1002/chem.201404566.

[30] S. Neelamegham *et al.*, "Updates to the Symbol Nomenclature for Glycans guidelines," *Glycobiology*, vol. 29, no. 9, pp. 620–624,. 2019, doi: 10.1093/glycob/cwz045.

[31] A. v. Remoortere, C. H.Hokke, G. J. v. Dam, I. v. Die, A. M.Deelder, and D. H. v. d. Eijnden, "Various stages of Schistosoma express Lewisx, LacdiNAc, GalNAc 1-4 (Fuc 1-3)GlcNAc and GalNAc 1-4(Fuc 1-2Fuc 1-3)GlcNAc carbohydrate epitopes: detection with monoclonal antibodies that are characterized by enzymatically synthesized neoglycoproteins," *Glycobiology*, vol. 10, no. 6, pp. 601–609. 2000, doi: 10.1093/glycob/10.6.601.

[32] A. Suga, M. Nagae, and Y. Yamaguchi, "Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation," *Glycobiology*, 2018, doi: 10.1093/glycob/cwy059.

[33] I. Hang *et al.*, "Analysis of site-specific N-glycan remodeling in the

endoplasmic reticulum and the Golgi," *Glycobiology*, vol. 25, no. 12, pp. 1335–1349, 2015, doi: 10.1093/glycob/cwv058.

[34] S. M. Dharmesh, T. P. Skelton, and J. U. Baenziger, "Co-ordinate and restricted expression of the ProXaaArg/Lys-specific GalNAc-transferase and the GalNAcβ1,4GlcNAcβ1,2Manα-4-sulfotransferase," *J. Biol. Chem.*, vol. 268, no. 23, pp. 17096–17102, 1993.

[35] D. Bonar and F. G. Hanisch, "Trefoil factor family domains represent highly efficient conformational determinants for N-linked N,N′-di-N-acetyllactosediamine (LacdiNAc) synthesis," *J. Biol. Chem.*, vol. 289, no. 43, pp. 29677–29690, 2014, doi: 10.1074/jbc.M114.596049.

[36] M. Gomez-Redondo *et al.*, "The two domains of human galectin-8 bind sialyl- and fucose-containing oligosaccharides in an independent manner. A 3D view by using NMR," *RSC Chem. Biol.*, 2021, doi: 10.1039/d1cb00051a.

[37] M. I. Nielsen *et al.*, "Galectin binding to cells and glycoproteins with genetically modified glycosylation reveals galectin-glycan specificities in a natural context.," *J. Biol. Chem.*, vol. 293, no. 52, pp. 20249–20262. 2018, doi: 10.1074/jbc.RA118.004636.

[38] J. Hirabayashi *et al.*, "Oligosaccharide specificity of galectins: A search by frontal affinity chromatography," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1572, no. 2–3, pp. 232–254, 2002, doi: 10.1016/S0304-4165(02)00311-2.

[39] H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko, and K. Yamashita, "Galectin-8-N-Domain Recognition Mechanism for Sialylated and Sulfated Glycans," *J. Biol. Chem.*, vol. 286, no. 13, pp. 11346–11355, 2011, doi: 10.1074/jbc.M110.195925.

[40] S. Carlsson *et al.*, "Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface," *Glycobiology*, vol. 17, no. 6, pp. 663–676. 2007, doi: 10.1093/glycob/cwm026.

[41] E. Ermakova *et al.*, "Lactose binding to human galectin-7 (p53-induced gene 1) induces long-range effects through the protein resulting in increased dimer stability and evidence for positive cooperativity," *Glycobiology*, vol. 23, no. 5, pp. 508–523, 2013, doi: 10.1093/glycob/cwt005.

[42] S. Bertuzzi *et al.*, "Unravelling the time scale of conformational plasticity and allostery in glycan recognition by human galectin-1.," *Chem. – A Eur. J.*. 2020, doi: 10.1002/chem.202003212.

[43] P. Valverde, J. D. Martínez, F. J. Cañada, A. Ardá, and J. Jiménez-

Barbero, “Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin,” *ChemBioChem*, vol. 21, no. 21, pp. 2999–3025. 2020, doi: 10.1002/cbic.202000238.

[44] C. D. L. Lima *et al.*, “Structural insights into the molecular recognition mechanism of the cancer and pathogenic epitope, LacdiNAc by immune-related lectins,” *Chem. – A Eur. J.*, 2021, doi: 10.1002/chem.202100800.

[45] P. Valverde *et al.*, “Molecular Insights into DC-SIGN Binding to Self-Antigens: The Interaction with the Blood Group A/B Antigens,” *ACS Chem. Biol.*, vol. 14, no. 7, pp. 1660–1671. 2019, doi: 10.1021/acschembio.9b00458.

[46] D. Sugahara, A. Tomioka, T. Sato, H. Narimatsu, and H. Kaji, “Large-scale identification of secretome glycoproteins recognized by Wisteria floribunda agglutinin: A glycoproteomic approach to biomarker discovery,” *Proteomics*, vol. 15, no. 17, pp. 2921–2933. 2015, doi: 10.1002/pmic.201400443.

Chapter VI

Structural insights of the interaction of Siglec-8 with sialic acid containing ligands: glycomimetics and FcεRI alpha

6.1 Introduction

As described in the general introduction, Siglec-8 is a CD33-related receptor expressed on the surface of eosinophils, mast cells and basophils.

It is composed of an extracellular domain (ECD) with one V-domain and two C-type Ig-like domains, a single transmembrane domain, and an intracellular domain with the presence of an ITIM domain [1]. The 3D structure of the V-domain has been solved by NMR [2], and it is composed of one β-sandwich of two antiparallel β-sheets (β-strands ABED and C′CFG) connected by one intra-sheet disulfide linkage. The canonical sialic acid-binding pocket has unique structural features at the G-G´ and C-C´ loops that confers specificity for 6′-sulfo sialyl LewisX (6′S sLeX: Neu5Acα2-3[6S] Gal β1-4[Fucα1–3] [6S]GlcNAc) [2] (Figure 6.1).

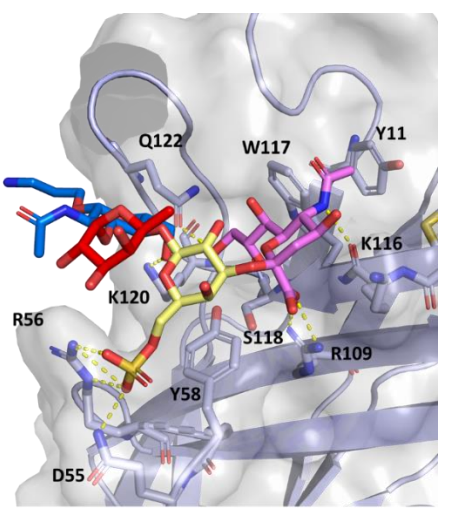


Figure 6.1: *The interaction model between Siglec-8 V-domain with with 6'-sulfo sialyl LewisX proposed by Propster et al [2].*

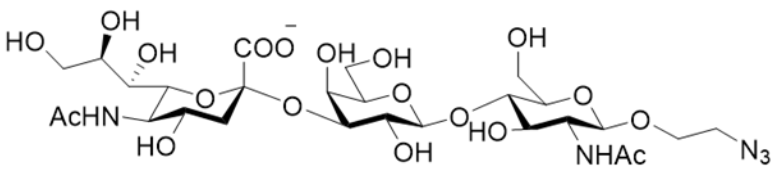
Currently, an anti-Siglec-8 mAb (AK002, Allakos Inc. Company) has been developed for the treatment of allergic diseases, for example chronic urticaria, gastroenteritis and duodenal eosinophilic [3]. However, antibodies have functional limitations such as inadequate pharmacokinetics and tissue accessibility, apart from harmful interactions with the immune system that can

cause serious collateral effects [4]. For these reasons, in certain cases, small molecule ligands could be used as alternatives to antibodies, thanks to their ability to dissociate from their target once endocytosis. On this basis, glycomimetics -chemically modified glycans- are drug-like compounds that mimic the structure and function of native glycans, improving their affinities, bioavailabilities and with longer serum half-lives. Due to this, in the last decade many efforts have been made for finding specific glycomimetics against Siglec-8.

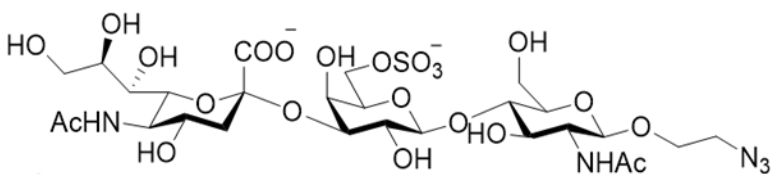
The laboratory of Prof. Paulson at Scripps has designed different glycan arrays of modified sialic acids against a variety of Siglecs. Among a library of 156 derived glycans, they identified a sulfonamide-based glycan with high affinity for Siglec-8 [5]. This high-affinity and selective glycan contains the Neu5Acα2-3[6S] Galβ1-4GlcNAc (6′-O-sulfo α2-3SLacNAc, **2**) scaffold decorated with the 9-N-(2-naphthyl-sulfonyl) moiety at C9 position (6′-O-sulfo NSA-α2-3SLacNAc, **3**). Furthermore, they generated liposomes bearing **3** on the surface, and tested them on a transgenic mouse line expressing Siglec-8 on mast cells [5]. The results showed that these liposomes could recruit Siglec-8 to the IgE-FcεRI complex, suppressing activation, and desensitizing mast cells antigen-induced mast cell degranulation [6].

In this work, we have studied the molecular recognition features of the interaction between the carbohydrate recognition V-domain of Siglec-8 with three different ligands (Figure 6.2): the Neu5Acα2-3-Galβ1-4GlcNAc (α2-3SLacNAc, **1**), the 6′-O-sulfo α2-3 SLacNAc (**2**) and the 6′-O-sulfo NSAα2-3SLacNAc (**3**).

Ligand 1



Ligand 2



Ligand 3

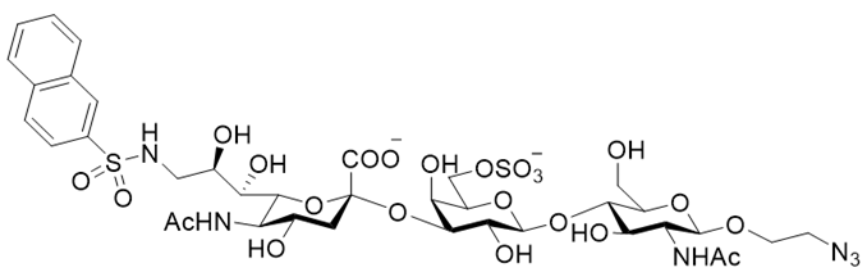


Figure 6.2: *The basic α2-3SLacNAc ligand (1), its Gal 4-O-Sulfated derivative (2) and the high affinity glycomimetic employed in this work. All ligands were provided by Prof. Paulson.*

Two different NMR-based approaches were employed, i) from the point of view of the protein, monitoring the chemical shift perturbation of the lectin cross peaks by titrating the protein with increasing amounts of the ligands; and ii) from the ligand point of view, using STD-NMR experiments. Additionally, to better understand the structural basis of the interaction between Siglec-8 and FcεRIα, we also studied their interaction by NMR. It is tempting to think that deciphering the molecular features of this interaction will help to better understand the mechanism of action of Siglec-8 under physiological and pathological conditions and to eventually develop therapeutic molecules.

Finally, X-ray crystallography was also used to decipher the three-dimensional structure of Siglec-8 in complex with the high affinity ligand **3** as well as the suprastructure formed by the full-length extracellular domain of Siglec-8 (Siglec-$8_{d1\text{-}d3}$) in complex with the AK002 Fab.

6.2 Purification of Siglec-8

Studies concerning the molecular basis of the interaction of Siglecs with their ligands are usually difficult due to the absence of suitable recombinant expression methods for producing these disulfide-containing proteins in sufficient quantities required for structural techniques. In this case, following a similar protocol that was developed by Propster et. Al [2], we produced the carbohydrate recognition domain (V Ig domain) of Siglec-8 from *E.coli* cells. On the other hand, we were able to produce the stable and well-folded complete extracellular domain of Siglec-8 using HEK293 cells (see Materials and Methods section).

6.2.1 Purification of Siglec-8 V domain from *E-coli* cells

Firstly, an expression test using *E.coli* Rosetta pLyss cells was carried out in order to find the best expression conditions for Siglec-8. These competent cells are generally used for the expression of eukaryotic proteins that are potentially toxic when over-expressed in *E.coli*. SDS-PAGE electrophoresis gel was used to check the presence of the protein in the soluble fraction. Unfortunately, the V domain of Siglec-8 remained in the insoluble portion (pellet), forming inclusion bodies. Therefore, the plasmid was transformed using *E.coli* Rosetta-gami B(DE3) competent cells (Novagen). In this case, Siglec-8 V domain was identified in the soluble fraction by observing a 16 kDa band in the SDS-PAGE gel (data not shown). Using these cells, Siglec-8 V was then overexpressed and purified using an affinity chromatography for the His-tag located at the C-terminus. The protein was eluted with 500 mM imidazole, concentrated and the monomer was separated from other contaminants and possible aggregates using

Superdex 75/26 600 gel filtration chromatography. The monomeric and pure protein eluted as a single peak in the mL 290, as expected for a globular protein with a molecular weight of 16 kDa. SDS-PAGE was used to check the protein purity (Figure 6.3).

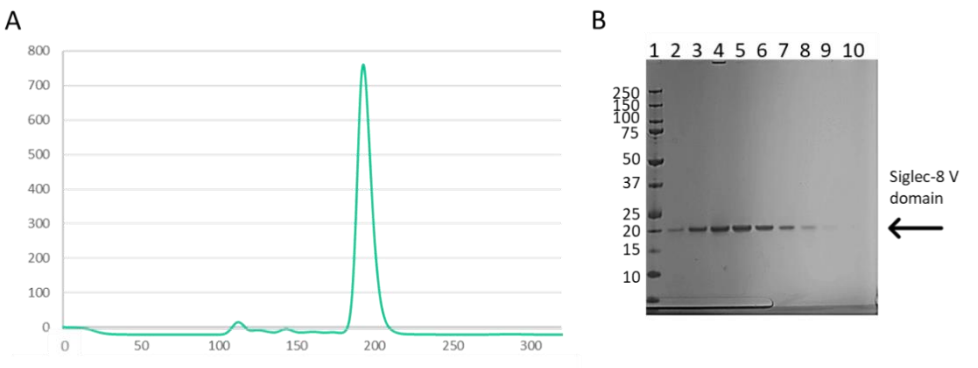


Figure 6.3: *Purification of Siglec 8 V domain. A) the gel filtration (Superdex 75 26 600) profile shows that Siglec-8 is in a monomeric form. B) SDS-PAGE gel for the eluted fraction of Siglec-8 from the gel filtration column : lane 1 molecular marker, lane 2-10 fractions corresponding to the elution peak.*

The final yield of the protein purification step was 1.75 mg/L, for the unlabeled and ^{15}N-labeled Siglec-8. Thus, the labelling conditions do not affect the protein expression levels or stability.

6.2.2 Purification of Siglec-8$_{d1\text{-}d3}$ from human cells

In order to decipher the three-dimensional structure of the full-length extracellular domain of Siglec-8, Siglec-8$_{d1\text{-}d3}$ was expressed in HEK293S cells, which incorporate high-mannose glycans. Expression on this type of cells yields a homogeneous sample that improves the possibility to get protein crystals for X-Ray diffraction. The DNA construct of Siglec-8$_{d1\text{-}d3}$ consists of the V domain and the following two Ig-like domains (C-type, amino acid residues 17-362) inserted in the mammalian expression vector pHLSec, between AgeI and KpnI restriction sites, to include a C-terminal 6 x His Tag. The first expression test using standard conditions provided a very low yield.

Thus, two approaches were followed to improve the protein expression yield:

i) Addition of m-VENUS protein for solubilization of Siglec-8_{d1d3}.
ii) Co-expression of Siglec-8 with the fragment antigen binding (Fab) portion of the AK002 antibody. The Fab, which is composed of one light chain (containing a variable domain and a constant domain) and one heavy chain (composed of a variable domain and a constant domain), is the portion of an antibody that binds to a specific epitope. Complexing proteins with other

The Fab is the portion of an antibody that binds a specific epitope. It is composed of two different chains: one light chain (containing a variable domain and a constant domain) and one heavy chain (composed of a variable domain and a constant domain).

Siglec-8-mVENUS construct yielded high levels of protein. As shown in Figure 6.4, after the gel filtration in a Superdex 10/300 200 Increase column, two different peaks were eluted. Although Siglec-8 forms aggregates (elution volume= 8 mL), it is primarily expressed as dimers/tetramers (elution volume= 12 mL) (Figure 6.4).

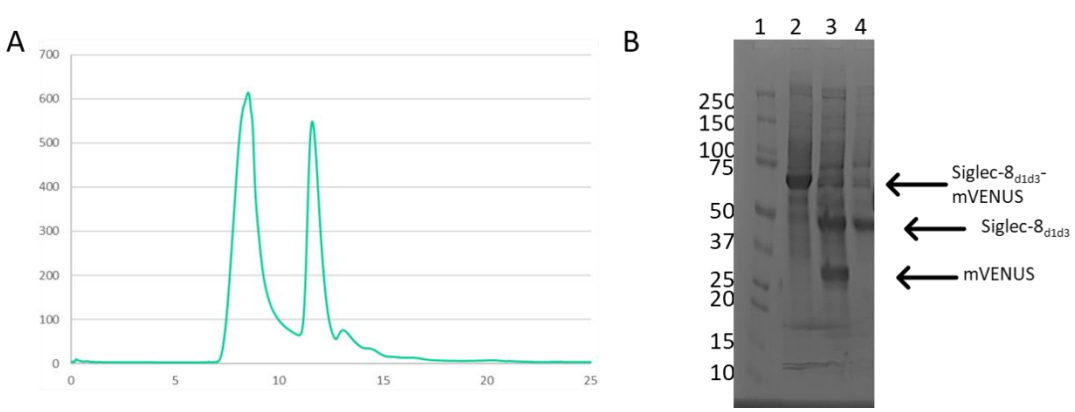


Figure 6.4: *Purification of Siglec 8_{d1d3}-mVENUS. A) the gel filtration profile (Superdex 200 10/300 Increase column) shows that Siglec 8_{d1d3}-mVENUS is in two oligomomeric forms. B) SDS-PAGE gel : lane 1 molecular marker (KDa), lane 2 sample after first step of purification, lane 3, Siglec-8_{d1d3} after TEV cleavage, lane 4 fraction correspondind to the second peak.*

.

However, protein glycosylation in HEK293 cells is hetereogenous and can confer high flexibility, two conditions that are detrimental for protein crystallization. Thus, to further increase the possibility of crystallization, the protein was incubated with Endo H, which is able to hydrolyze all the high mannose glycans. But in this case, the protein precipitated, indicating that the presence of glycans is essential for its stability.

Complexing a target protein with an auxiliary protein, which acts as a molecular chaperone, also represents an option for increasing crystallization chances [7], [8]. Fabs have previously shown to act as crystallization chaperones, by minimizing the target conformational heterogeneity through 'locking' the antigen in a particular conformation. Moreover, formation of the crystal lattice based on contacts between Fab molecules is also likely. In addition, a chaperone with a previously characterized structure can facilitate molecular replacement phasing. Therefore, we decided to co-express Siglec-$8_{d1\text{-}d3}$ with the anti-Siglec-8 AK002 Fab. After expression using the HEK293S cells, the protein complex was purified with a Ni affinity column and incubated with Endo H. The digested protein complex was finally purified using a gel filtration chromatography (Figure 6.5).

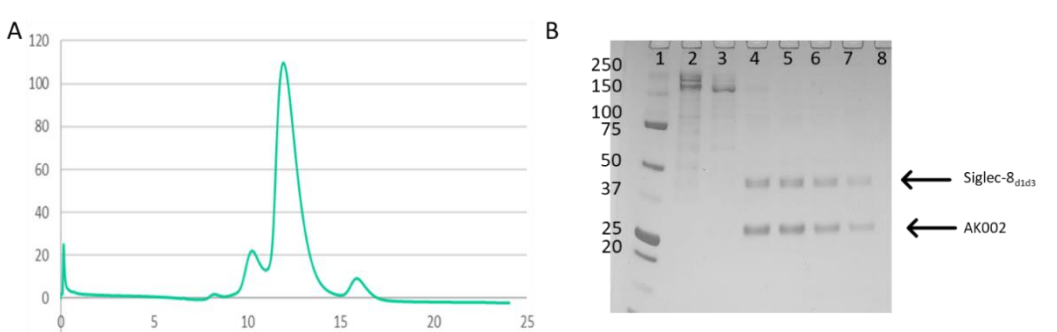


Figura 6.5: *Purification of Siglec-8_{d1d3}-AK002. A) Gel filtration chromatography (in a Superdex 200 10/300 Increase column) showing that the protein complex has a total molecular weight of 80 kDa. B) SDS-PAGE gel with fractions eluted from the gel filtration: lane 1 molecular marker (kDa), lane 4-fraction corresponding to the main peak eluted from the gel filtration.*

Fractions 12-19 were collected (corresponding to 80 KDa molecular weight, based on molecular weight standards) and the purity checked by SDS-PAGE gel. For the crystallization trilas, Siglec-8_{d1d3}-AK002 complex was concentrated at 10 mg/mL.

6.3 X-Ray crystallographic analysis of Siglec-8

In a first step to obtain details of the interaction between Siglec-8 with glycomimetics, different attempts to crystallize Siglec-8 (the V domain and full-length ECD) with the different ligands (Table 6.1). For crystallization, a range of 10-20 mg/mL of protein was used, employing different commercially available crystallization conditions (see Materials and Methods) and the vapor-diffusion technique in a sitting-drop format in 96-wells MRC crystallization plates. All experiments were carried out in the high-throughput crystallization facility at CIC bioGUNE.

Table 6.1: *Construct of Siglec-8 used for crystallyzation.*

Domains	Residues	Ligand
Siglec-8 V domain	17-155 (C42A)	-
Siglec-8 V domain	17-155 (C42A)	Ligand-3
Siglec-8 $_{d1-d3}$	17-362	AK002

For the crystallization of Siglec-8 V domain in complex with ligand **3**, soaking experiments were initially carried. The first crystals of the apo form of Siglec-8 V domain were obtained at 16 mg/mL concentration, in 25 % (w/v) PEG 3350 and 0.1 M Bis Tris-HCl, pH 6.5 (Figure 6.6A)

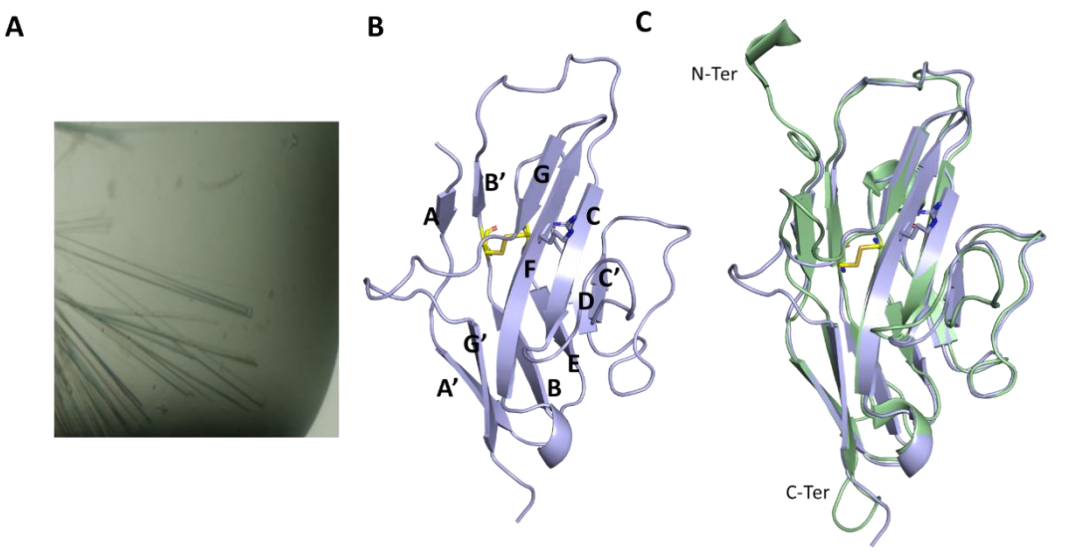


***Figure 6.6**: A) The obtained crystals of siglec-8 V-domain in 25 % (w/v) PEG 3350 and 0.1 M Bis Tris-HCl, pH 6.5 show a bar-shaped form B) the 3D Siglec-8 V-domain structure solved by X-Ray, The intrasheet disulfide bond between Cys31 and Cys91 is indicated by a yellow bar C) Superimposition of NMR (purple) and X-Ray (green) structures obtained for Siglec-8 V domain. The insets show the expansion of the key motifs for ligand binding. The main chain RMSD is 0.956 Angs.*

Cryoprotected crystals with 25% glycerol were tested for X-ray diffraction and collected on the Beamline X06DA - PXIII from Swiss Light Source (Villigen, Swiss). The apo form of Siglec-8 V domain crystal diffracted to 3.0 Å resolution and belong to the space group $P2_1\ 2_1\ 2_1$ with the unit cells parameters a= 35.992 Å, b= 69.166 Å, c= 166.513 Å, α=β=γ= 90.00º. The Matthew´s Volume coefficient (V_m) calculation, estimated three monomers in the asymmetric unit. The crystal structure was solved by molecular replacement using the NMR structure of Siglec-8 V domain discussed above in Phenix Phaser [9] (Table 6.2). As determined by the Vm, there are three molecules of Siglec-8 in the asymmetric unit.

Table 6.2: *Data collection and refinement for Siglec-8 V domain. One unique crystal was usef for thr data set. Values in parentheses are for the highest resolution shell*

Parameter	**Siglec-8 V domain**
Data collection	
Space group	$P2_12_12_1$
a, b, c (Å)	35.992, 69.166, 166.513
A, β, ɣ (Å)	90, 90, 90
Resolution	43.29-3.12
R_{meas}	33.80 (117.60)
I/σI	6.66 (1.83)
Completeness (%)	99.60 (100.00)
Redundancy	7.00 (6.90)
$CC_{1/2}$ (%)	98.20 (63.80)
Refinement	
Resolution	43.29-3.12
No. reflection	120116
R work/R free	0.26-0.29
No. atoms protein	3165
B factors Protein	68.09
RMS deviation	
Bond length (Å)	0.002
Bond angle (º)	0.525

However, in the X-Ray crystallographic structure (Figure 6.6B) no electron density could be detected for regions M1-D9 and T137-R144 in the N- and C-termini, respectively. The intrinsic flexibility of those regions in the apo form accounts for this fact. In contrast, very small differences were observed for the residues located at loops C-C' and D-E. Characteristic features for Siglec lectin domains were also observed in Siglec-8. The V-Ig like domain of Siglec-8 is formed by two antiparallel β-sheets formed by β-strands ABED and C′CFG. The conserved intra-sheet disulfide bond between adjacent β-strands B and E was also present. Moreover, the strictly conserved essential arginine (R109) on β-strand F, known to provide a key salt bridge interaction for sialic acid

recognition, as well as the splitting of the G-strand into two shorter β-strands (G and G′) were present (Figure 6.6B).

These crystals of Siglec-8 V domain were then used for soaking experiments with ligand **3**. For successful soaking experiments, the lattice of the crystal must contain channels containing the bulk solvent to provide a considerable access for the small ligands to reach the protein. In this case, the crystals were first reproduced in a plate of 24 wells at 25 % (w/v) PEG 3350 and 0.1 M BisTris-HCl, pH 6.5, and then soaked with **3** at a final concentration of 3 mM. The mix was incubated at different times (30 sec, 1 min, 5 min, 30 min and over-night) and then flash-frozen in liquid nitrogen after cryoprotection with glycerol.

The X-ray diffraction pattern was collected using the beamline MX XALOC BL-13 from ALBA (Barcelona, Spain). The crystal of the Siglec-8 V domain soaked with **3** diffracted to 2.0 Å resolution and belong to the space group P2 2 2, with three molecules in the asymmetric unit. The unit cells parameters a= 36.311 Å, b= 69.896 Å, c= 167.222 Å, α=β=γ= 90.00º. Unfortunately, the ligand was not present in the binding site of Siglec-8 V domain. After the crystal lattice was carefully analyzed, it was evident that the disposition of the molecules in the crystal lattice might preclude the access of the ligand to the binding site (Figure 6.7). Indeed, the binding sites in molecule A and molecule C are hindered by loop C'-D (residue 67-76) of the symmetry related molecules B and C, respectively. Additionally, the N-terminal of each Siglec-8 would need to rearrange (Figure 6.7) to allow the formation of the hydrophobic pocket that interacts with the naphthyl group according to the NMR studies described above.

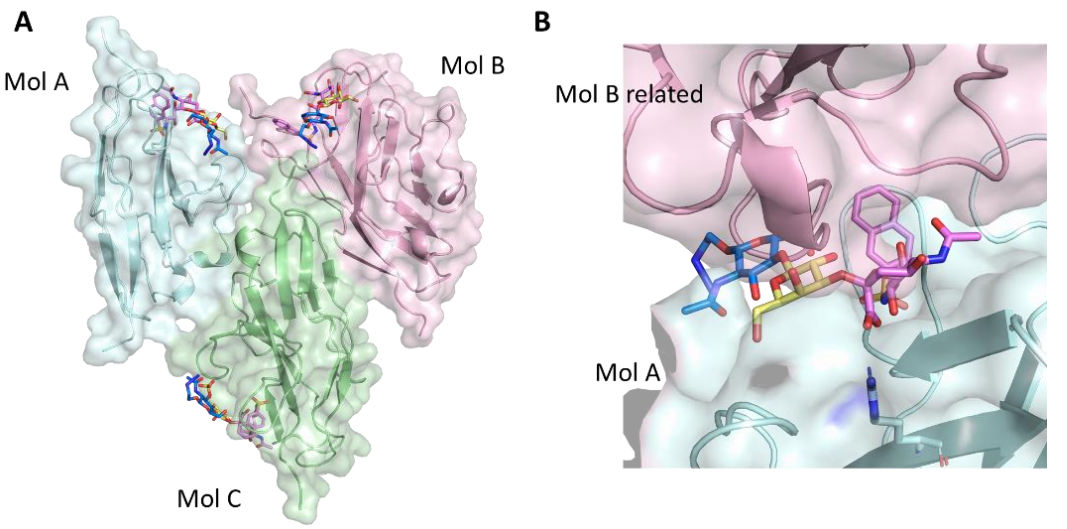


Figure 6.7: *Absence of ligand 3 bound in the binding pocket of Siglec-8 V domain crystals after soaking experiments. A) Surface and cartoon representation of the three molecules (mol A, B and C) in the asymmetric unit. Ligand 3 was manually docked into the binding pocket of each molecule of Siglec-8 V domain. B) Zoom view of molecule B blocking the sialic acid binding pocket on molecule A.*

For this reason, co-crystallization of Siglec-8 V protein with Ligand **3** was tried. For the co-crystallization trials, the lectin: ligand molar ratio was set to 10:1 (10 mM of **3** and 1 mM of Siglec-8). Crystals were obtained in two different conditions of Crystal screen Cryo 2 after incubation for 1 month at 277 K (Figure 6.8). These crystals were tested in Beamline X06DA - PXIII from Swiss Light Source (Villigen, Swiss) and diffracted to 12 and 8 Å resolution, respectively. Thus, further improvement of the quality of these crystals must be performed.

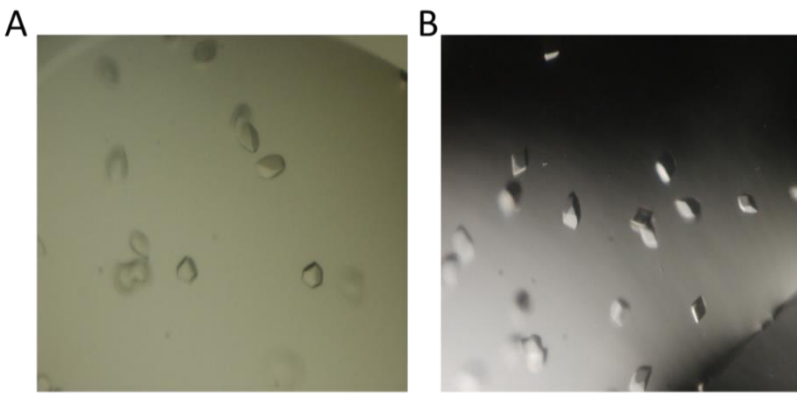


Figure 6.8: *Crystals of Siglec-8 V-domain co-crystallised with Ligand 3 in A) 0.075 M HEPES pH 7.5 15% w/v PEG 10,000 25% v/v Glycerol and B) 0.65 M Imidazole pH 7.0, 35% v/v Glycerol conditions.*

In parallel, additional crystallization trials were carried out with the full-length ECD of Siglec-8. In this case, the AK002 Fab was used as a crystallization chaperone [8] for the Siglec-8 ECD. Bar-like crystals were obtained at 20% (w/v) PEG 3350, 0.2 M ammonium Citrate tribasic pH 7.0 (Figure 6.8). Crystals were recently tested in the X06DA - PXIII beamline from Swiss Light Source beamline (Villigen, Swiss), diffracting to 2.7 Å resolution. Similarly, the AK002 Fab (Figure 6.9) was crystallized using 20 % (w/v) PEG 3350 and 0.2 M Ammonium Tartrate dibasic conditions, which diffracted to 2.7 Å resolution (Table 6.3).

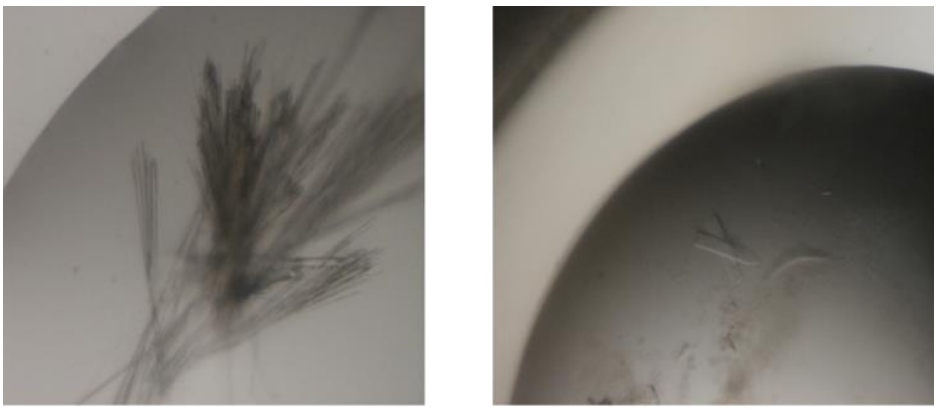


Figure 6.9: *The obtained crystals of AK002 (left) and of the complex Siglec 8_{d1d3}-AK002 8 (right) in 0.2 M Ammonium tartrate dibasic and 20% w/v PEG 3350, and 0.2 M Ammonium citrate tribasic pH 7 and 20% w/v PEG 3350, respectively. The analysis is currently in progress.*

Table 6.3: *Data collection for Siglec 8_{d1d3}-AK002 and AK002. One unique crystal was use for the data set. Values in parentheses are for the highest resolution shell*

Parameter	**Siglec 8_{d1d3}-AK002**	**AK002**
Data collection		
Space group	C121	P1
a, b, c (Å)	145.244, 4.512, 111.291	52.063, 146.176, 146.126
A, β, ɣ (Å)	90.00, 111.402, 90.00	59.969, 90.002, 90.00
Resolution		
R_{meas}	26.6 (137.4)	35.0 (169.5)
I/σI	6.07 (1.12)	3.72 (0.72)
Completeness (%)	99.4 (98.1)	98.2 (99.6)
Redudandy	3.49 (3.43)	6.48 (6.51)
$CC_{1/2}$ (%)	97.3 (42.9)	92.6 (14.0)

6.4 NMR-based studies to unravel the molecular details of the interaction of ligand 3 with Siglec-8: chemical shift perturbation analysis.

Since the X-Ray crystallographic analysis of the interaction of Siglec-8 with the glycomimetics remains elusive so far, the molecular recognition features were investigated by NMR. First, a protein-based analysis was carried out. The chemical shift perturbation (CSP) induced by a ligand on the lectin, as

determined in the ^{1}H-^{15}N TROSY NMR spectra provides information on the location of the binding site. In this case, to obtain a ^{1}H-^{15}N TROSY NMR spectra were recorded for the ^{15}N-labelled Siglec-8 V domain. In this experiment (Figure 6.10), each cross-peak represents a NH amide group for each amino acid (except for proline, obviously). The corresponding protein sequence and 3D structure provides a unique chemical environment for each aminoacid and therefore, the resulting TROSY spectrum can be considered as an ID, the protein fingerprint. Indeed, the dispersion and specific $^{1}H/^{15}N$ chemical shifts of the cross peaks are characteristic and unique for each protein. The assignment of the different cross peaks was based on the previous work by Propster et al [2].

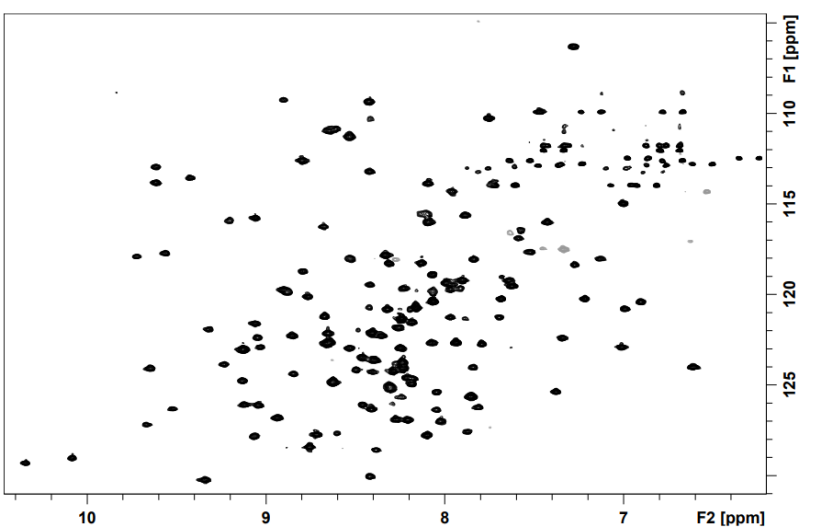


Figure 6.10: *^{1}H-^{15}N TROSY NMR spectrum of Siglec 8 V domain.*

Different ^{1}H,^{15}N-TROSY spectra were recorded for the apo form and upon addition of increasing amounts of the different ligands (Figure 6.10 & 6.11). The addition of a ligand modifies the chemical environment for those amino acid residues that are directly or indirectly involved in the binding event.

Therefore, the induced chemical shift perturbations (CSP) provide a direct hint on the existence of binding and on those residues that are more profoundly affected by the presence of the added ligand. The fitting of the observed chemical shift perturbations to a model based on the law of mass action for a chemical exchange process between a free and bound protein allows estimating the binding affinity [8].

The analysis of the data showed that the estimated affinity for the basic scaffold (ligand **1**, α2-3SLacNAc) was very low, in the mM range. Indeed, after addition of 300 equivalents of ligand, the protein was still not saturated with the ligand (Figure 6.11). Nevertheless, specific chemical shift perturbations at the C-terminus, in the F-G β-sheet at the binding site were clearly observed. The low affinity did not allow to quantitatively calculate the affinity, although it was estimated to be *ca*. 4 mM, as previously described [2].

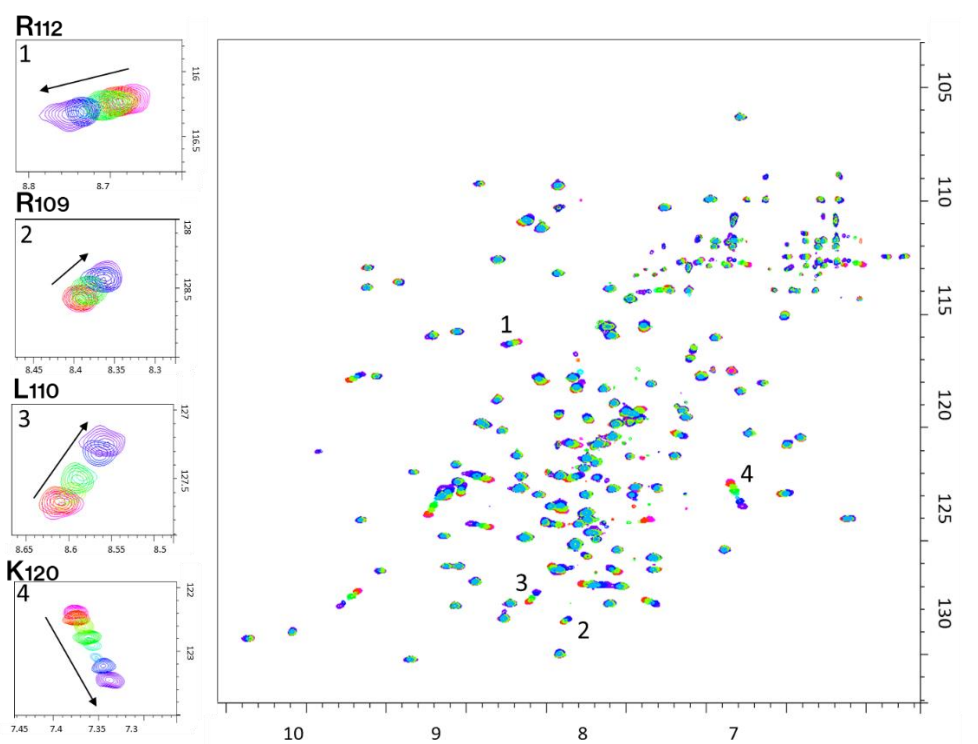


Figure 6.11: *CSP analysis of the interaction of Siglec-8 V domain with ligand 1. The specific perturbation of different key residues is shown in the insets.*

For ligand **2** (6′-O-sulfo α2-3SLacNAc), the analysis of the chemical shift perturbations of the lectin cross peaks showed that saturation was reached before the final addition of 50 molar equivalents of ligand versus the protein (Figure 6.12)

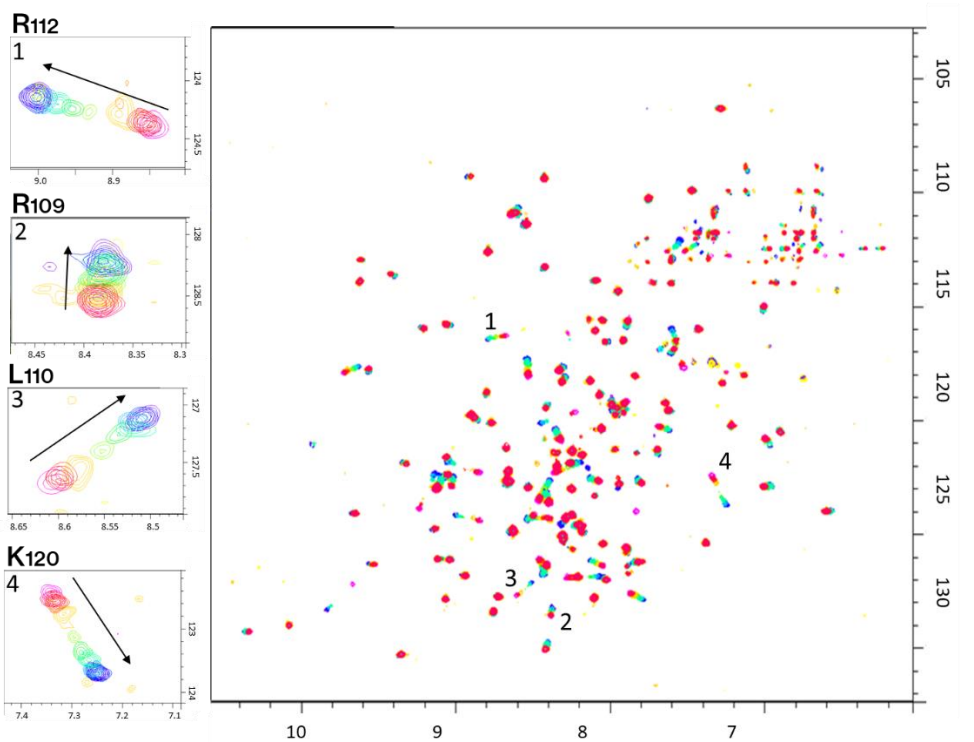


Figure 6.12: *CSP analysis of the interaction of Siglec-8 V domain with ligand 2. The specific perturbation of different key residues is shown in the insets.*

The analysis of the CSP showed that the more perturbed amino acids are located in the carbohydrate recognition domain (F and G β-sheet). Moreover, CSPs were also observed in the C-C' loop located in front of the binding site. In particular, changes in residues R109, K116, W117, S118 and K120 were observed (Figure 6.12 & 6.14). The analysis of the interaction of the GlcNAc 3-O-fucosylated analogue of **2** with Siglec-8 [2] has shown that these residues directly interact with the ligand. Indeed, R109 is the canonical arginine located in the binding site that interacts via a salt bridge with the carboxyl group in position 1 of the

sialic acid. Moreover, the same authors proposed that the rest of the sialic acid pyranose ring polar groups display hydrogen bonds with residues K116, S118 (in the G β-sheet). In contrast to the observations for ligand **1**, residues R56, Y58 and Q59 in the loop C-C' now suffer important CSP, suggesting that the presence of the sulphate group in position 6 of the Gal moiety is directly involved in the interaction with the lectin at this region. In accordance with the observations reported [2], a fast/intermediate exchange was noticed for most cross peaks along the titration. Based on these experimental data, the X-ray structure described above, and the previous model proposed by Propster et al. for sulfated sLeX analogues [2], a putative binding pose between Siglec-8 and **2** was proposed (see below in Figure 6.17).

The observed CSP were used to estimate the dissociation constant (k_D). The affinity of Siglec-8 V domain for **2** increases in more of one order of magnitude respect to that of **1**, to ca. 170 µM.

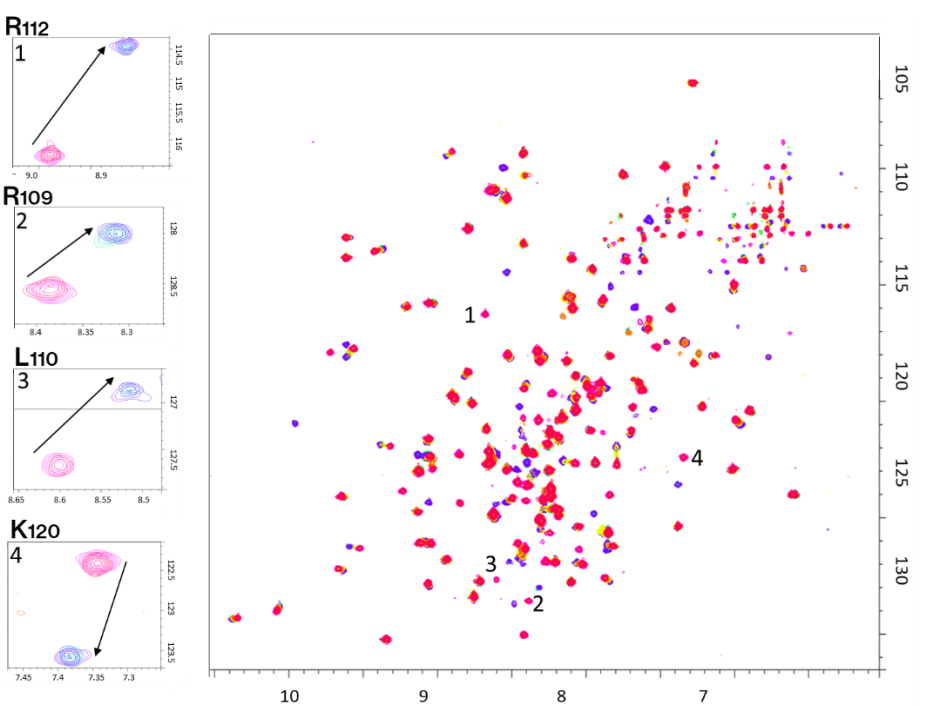


Figure 6.13: *CSP analysis of the interaction of Siglec-8 V domain with ligand 3. The specific perturbation of different key residues is shown in the insets*

Interestingly, similar CSPs in the above mentioned regions (F and G β-sheets and the C-C' loop) were deduced when Siglec-8 was titrated with ligand **3** (Figure 6.12 & 6.13). Additionally, CSPs in the N-terminal region of the protein were also observed. This fact strongly suggests that the sulfated sialic acid moiety of **3** is accommodated in a similar manner to that described for **2**, while the naphthyl group at C9 position is located close to the N-terminus. In this case, the saturation was reached using ca. 10 equivalents of the ligand. This feature suggests that this glycomimetic displays the highest affinity of the three ligands, in the low micromolar range.

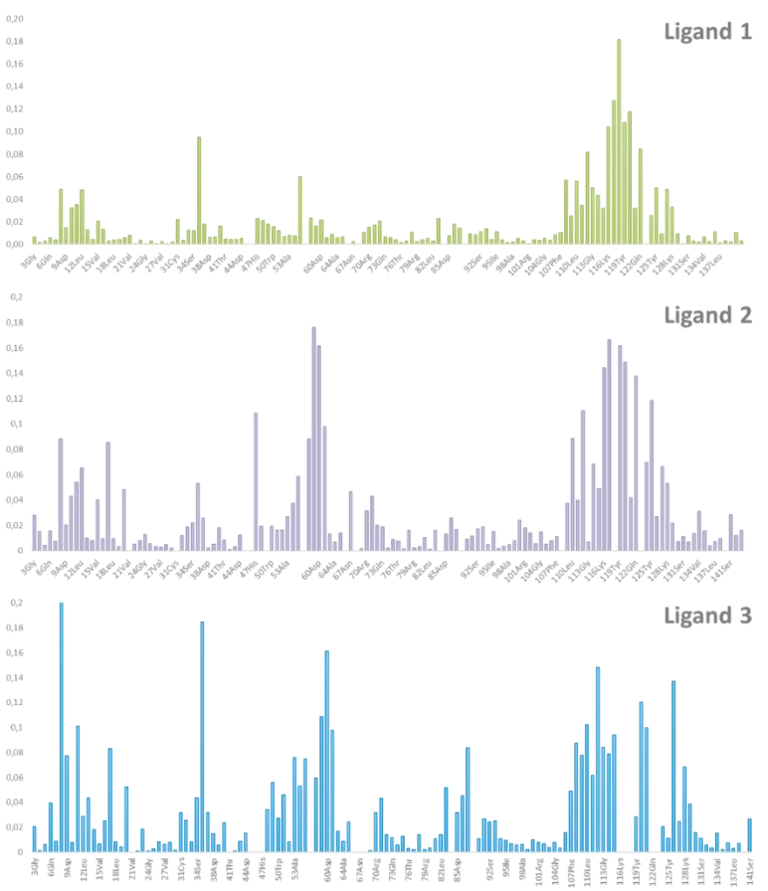


Figure 6.14: *CSP plots for the TROSY cross peaks of the amino acid residues of Siglec-8 V domain in the presence of the maximum amounts of ligands 1 (300 equivalents), 2 (50 equivalents), and 3 (10 equivalents).*

6.5 The ligand's perspective: STD NMR experiments to analyze the binding epitope of sialoside analogues

The molecular recognition study continued with the glycan-based NMR interaction analysis. As first step, the ^{1}H and ^{13}C NMR assignment of the signals of the corresponding nuclei of the ligands was performed using standard 1D and 2D NMR experiments. Tables 6.2 and 6.3 show the NMR assignment for ligand **2** and **3**.

Table 6.2: *^{1}H and ^{13}C NMR assignments for ligand 2.*

Ligand	position/ atom	^{1}H (δ, ppm)	^{13}C (δ, ppm)
GlcNAc	1	4.48	106.1
	2	3.64	60.0
	3	3.63	77.0
	4	3.52	84.0
	5	3.51	79.6
	$H6_1$	3.73	65.1
	$H6_2$	3.88	
	NAc	1.91	27.3
Gal	1	4.47	107.2
	2	3.45	74.1
	3	4.01	80.1
	4	3.89	72.3
	5	3.84	77.6
	$H6_{1,2}$	4.05	72.3
Sialic acid	$H3_{eq}$	2.62	44.5
	$H3_{ax}$	1.68	
	4	3.56	73.4
	5	3.57	56.7
	6	3.52	76.7
	NAc	1.91	27.0
	7	3.47	73.1
	8	3.76	76.6
	$H9_1$	3.51	67.48
	$H9_2$	3.75	

Table 6.3: *^{1}H and ^{13}C NMR assignments for ligand 3.*

Ligand	Position/atom	^{1}H (δ, ppm)	^{13}C (δ, ppm)
GlcNAc	1	4.42	85.9
	2	3.64	39.8
	3	3.54	57.7
	4	3.14	64.6
	5	3.40	59.6
	$H6_1$	3.71	45.3
	$H6_2$	3.49	
	NAc	1.91	7.2
Gal	1	3.89	87.6
	2	3.34	53.9
	3	3.79	60.2
	4	3.77	52.0
	H5	3.48	57.5
	$H6_1$	3.96	52.2
	$H6_2$	4.00	
Sialic acid	$H3_{eq}$	2.60	24.7
	$H3_{ax}$	1.61	
	4	3.53	53.1
	5	3.67	36.8
	6	3.53	57.0
	NAc	1.88	6.9
	7	3.42	54.1
	8	3.77	55.7
	$H9_1$	2.99	30.1
	$H9_2$	3.31	
Naphthyl group	N1	8.45	113.0
	N3	8.06	114.9
	N4	7.81	106.6
	N5	7.97	113.0
	N6	7.66	114.5
	N7	7.63	113.2
	N8	8.03	114.2

Then, STD-NMR experiments were used in order to reveal the binding epitope of the ligands to Siglec-8 V domain. For **2**, the obtained STD data corroborated the relevance of the sialic acid for binding to this lectin. In particular, H5, H6, H7, H8, $H9_1$ and $H9_2$ showed strong STD responses. Furthermore, H5 and H6

protons of the Gal moiety also showed STD effects (Figure 6.15A), suggesting that they are involved in the binding process.

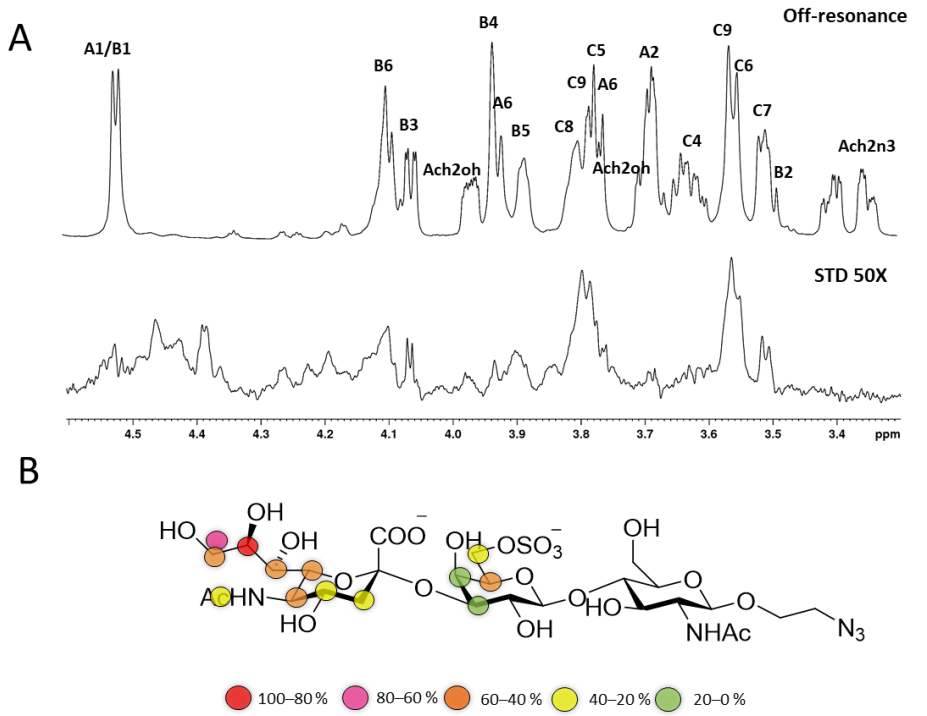


Figure 6.15: *A) Off resonance (top) and STD (bottom) NMR spectra recorded for the interaction of ligand 2 with Siglec-8 V domain. The inset shows the assignment of the different protons. B) The scheme at the right handside shows the deduced binding epitope from the observed STD intensities.*

For **3**, the experimental STD NMR data for the sugar protons showed a similar binding epitope to that described above for **2**. Fittingly, in this case, the naphthyl group also presented very high STD intensities. These observations strongly suggest that while the sialic acid and the Gal rings are located in the binding pocket of Siglec-8 in a similar manner respect to the pose of ligand 2, the naphthyl ring is also directly involved in the binding event (Figure 6.16A).

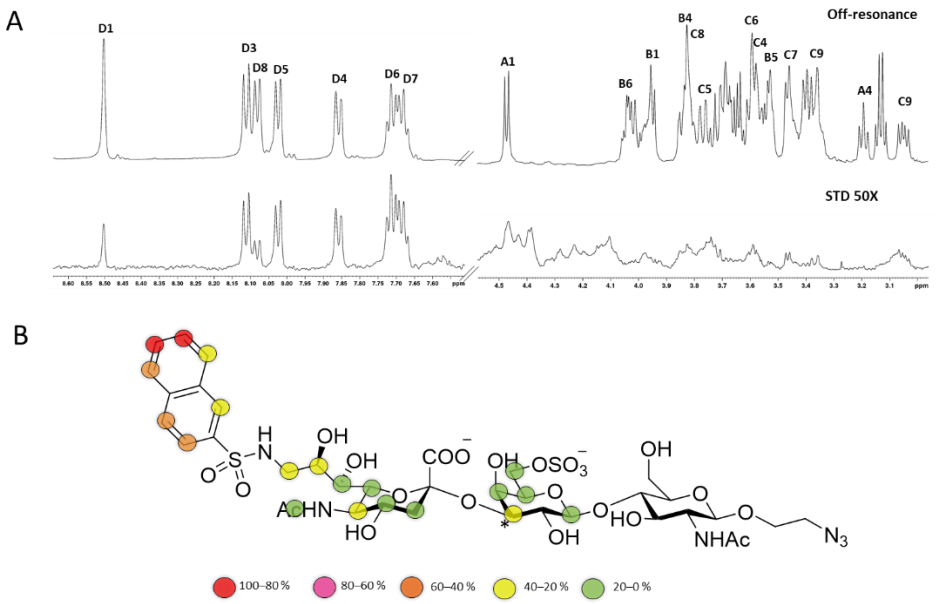


Figure 6.16: *A) Off resonance (top) and STD (bottom) NMR spectra recorded for the interaction of ligand 3 with Siglec-8 V domain. The inset shows the assignment of the different protons. B) The scheme at the right handside shows the deduced binding epitope from the observed STD intensities. The sugar binding epitope is basically identical for both 2 and 3. For 3, the naphtyl moiety is also heavily involved in the interaction with the lectin.*

6.6 Towards a 3D model of the sialosides/Siglec-8 complexes from the NMR data.

Based on these experimental observations from both perspectives, from the protein and from the ligand, 3D structures for the complexes formed by Siglec-8 with ligands **2** and **3** were built. The initial protein structure was taken either from our X-Ray crystallographic structure described above or from the deposited coordinates (PDB ID, 2N7A) of the complex between the sulfated analogue of sLeX and Siglec-8 deduced by NMR by Propster et al. [2], making the corresponding modifications to generate ligands **2** and **3** (Figure 6.17**Figure**). The results were basically identical, independently of the starting Siglec-8 coordinates. This task was performed by Mr. Unai Atxabal, a PhD candidate in the group. The generated starting geometries were submitted to a standard geometry optimization with the AMBER suite of programs followed by MD to

assess the conformational stability of the complexes. According to the obtained data, the formed complexes by Siglec-8 with **2** and **3** were fairly stable during the simulation and provided the rational for the observed affinities. Additional simulations were also carried out using the X-Ray crystallographic structure of apo Siglec-8 as starting geometry that will be discussed later in this chapter. Similar conclusions were drawn using either starting structures, which are indeed very similar (see below).

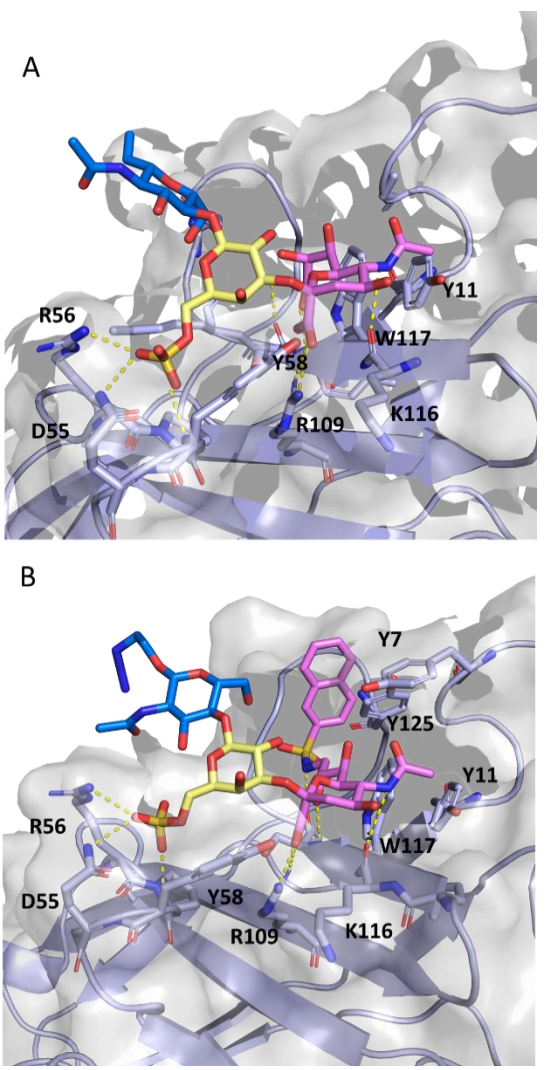


Figure 6.17: *Representation of the deduced models for the interaction of ligands 2 (A) and 3 (B) with Siglec-8 V domain. The models have been built using our crystallography structure of Siglec-8 V domain, the model proposed for ligand 2 and complied with the observed CSP during the titrations with 2 and 3.*

The generated geometries for the complexes comply with the different experimental observations. The Gal and the sialic acid moieties of both **2** and **3** are recognized by Siglec-8 in a similar manner (Figure 6.17). R109 that interacts with the carboxyl group in position 1 of the sialic acid, while the other sialic acid polar groups display hydrogen bonds with residues K116, S118 in the G β-sheet. R56, Y58 and Q59 bind to the sulfate group of the Gal ring, providing additional impetus to the interaction respect to the non sulfated analogue, **1**, thus explaining the large increase in affinity between **1** and **2**. Moreover, the non polar naphthyl group of **3** is further stabilized within a hydrophobic aromatic pocket composed by Y7, Y11 and Y125, thus explaining the further increase in binding affinity when passing from **2** to **3**.

6.7 The interaction of Siglec-8 with sialic acids on the N-linked glycans of FcεRI alpha subunit analyzed by NMR

The extracellular domain of FcεRIα contains 7 N-linked glycosylation sites and its glycoprofile has previously been described in chapter IV in this Thesis [10]. This analysis showed the presence of complex type glycans with different degrees of sialylation, including the presence of the Neu5Ac(α2−3)Gal and Neu5Ac(α2−6)Gal epitopes. Moreover, it was estimated that the amount of Neu5Ac(α2−3)Gal was 4 times higher than the Neu5Ac(α2−6)Gal. However, no evidence for the presence of 6-O-sulphated Gal epitopes was provided.

To evaluate if Siglec-8 interacts with the glycans on FcεRIα, $^{1}H–^{15}N$ TROSY spectra of isotopically ^{15}N-labelled Siglec-8 in the absence and presence of the extracellular domain of FcεRIα (expressed in HEK293F cells) were recorded. For the evaluation of the binding between the N-glycans on FcεRIα and Siglec-8, the CSP of the cross peaks and the changes in their linewidths in the $^{1}H,^{15}N$-TROSY spectra of Siglec-8 upon addition of FcεRIα were evaluated. There was a general signal decrease in the signal intensities of Siglec-8 when 1 equivalent

of FcεRIα was added, especially significant for several specific cross peaks (Figure 6.18).

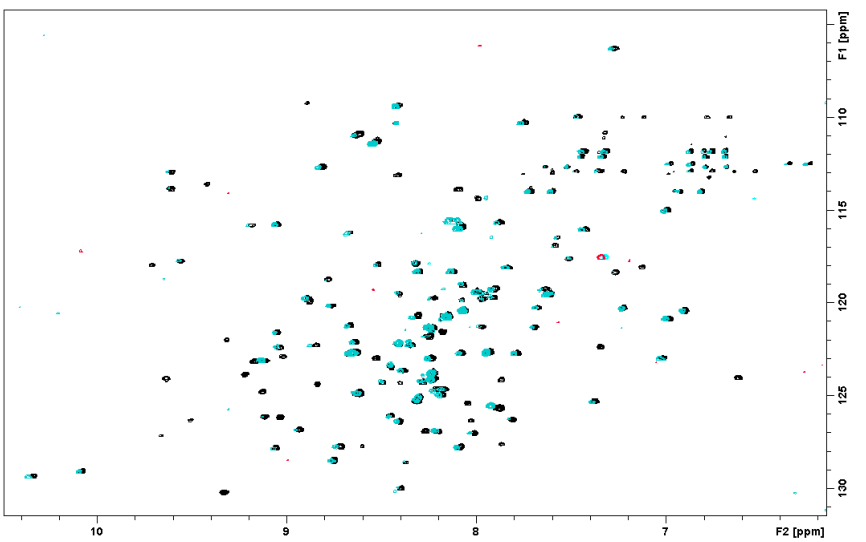


Figure 6.18: *Superimposition of the ^{1}H-^{15}N TROSY spectra recorded for Siglec-8 V domain in the absence (black) and presence (ligt blue) of FcεRIα expressed in HEK293F cells. The decrease in the intensities of many cross peaks is evident.*

Since many cross peaks beyond the sialic acid binding site described above show a significant decrease in intensity, a different protocol was employed to generate FcεRIα. Thus, FcεRIα was now expressed in HEK293S, which only provides high-Man N-linked glycans as post-translational glycosylation. Therefore, this FcεRIα does not display sialic acid residues. The analogous ^{1}H,^{15}N-TROSY spectra for Siglec-8 was then recorded in the presence of the new FcεRIα.

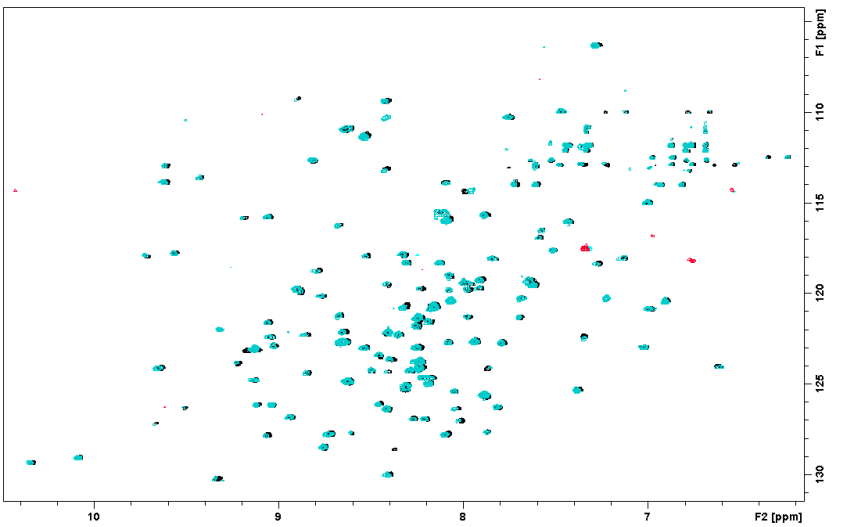


Figure 6.19: *Superimposition of the ^{1}H-^{15}N TROSY spectra recorded for Siglec-8 V domain in the absence (black) and presence (light blue) of FcεRIα expressed in HEK293S cells. No intensity changes are observed.*

As show in Figure 6.19, now there are no changes in the intensities of the cross peaks of Siglec-8 between both TROSY spectra, strongly suggesting that Siglec-8 does not interact with FcεRIα expressed in HEK293S cells. Therefore, these results demonstrate that the binding between Siglec-8 and FcεRIα is driven by complex N-glycans containing terminal sialic acids. The analysis of the intensities of the cross peaks of Siglec-8 in the presence of FcεRIα with all type of N-glycans (expressed in HEK293F) and with only high-Man (expressed in HEK293S) is shown in Figure 6.20.

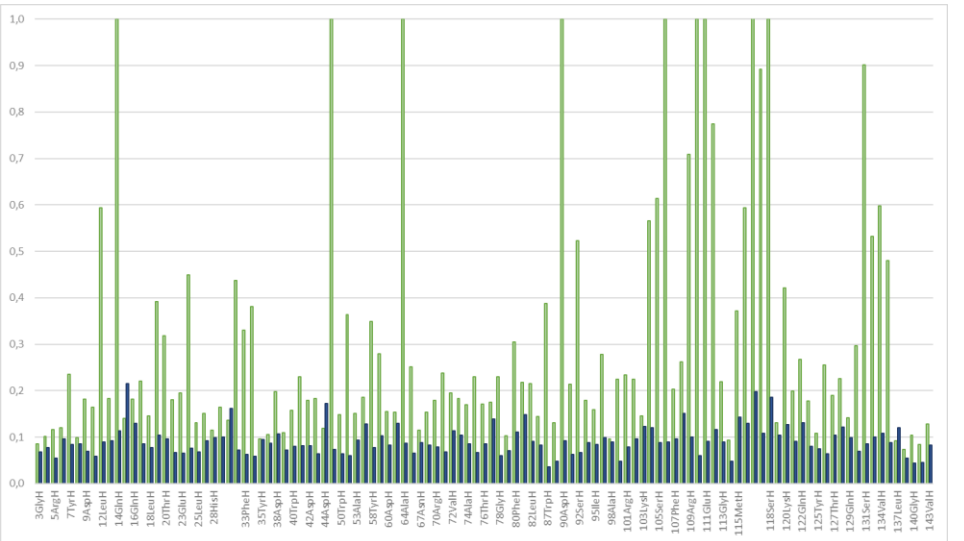


Figure 6.20: *FcεRIα expressed in HEK293F (green) and in HEK293S (blue). The analysis of the integrals indicate that there is a general decrease of the intensities upon addition of FcεRIα expressed in HEK293F cells, while no intensity changes are observed upon addition of FcεRIα expressed in HEK293S cells*

Upon addition of the FcεRIα expressed in HEK293F cells, the largest intensity decreases are observed for amino acids located in the canonical binding site (F-G β-sheet) and in the C-C´ loop. Based on these results, it is highly likely that the interaction between Siglec-8 and FcεRIα is driven by complex sialylated N-glycans.

6.8 Conclusions

Siglec-8 downregulates eosinophil- and mast cell-mediated inflammatory responses upon engagement by specific sialylated glycans. This immune receptor is considered a promising target for novel anti-inflammatory treatment strategies for asthma and other disease conditions in which inappropriate and/or prolonged inflammatory responses of these cell types contributes to pathology. Thus, there is an increasing number of antibodies (such as AK002) and modified sialic acids under development to target specifically Siglec-8.

Although the NMR structure of Siglec-8 V domain was already solved in the apo form and bound to its preferable glycan (6′-sulfo sialyl LewisX), here we

have analyzed its interaction with a glycomimetic (6′-O-sulfo NSAα2-3SLacNAc) using solution NMR spectroscopy. Siglec-8 V domain has been successfully expressed from *E.coli* Rosetta-gami cells and purified in labeled and unlabeled conditions.

$^{1}H,^{15}N$-TROSY experiments for the apo ^{15}N-labeled protein and after addition of increasing amounts of specific ligands (the α2-3SLacNAc **1**, 6′-O-sulfo α2-3 SLacNAc **2** and 6′-O-sulfo NSAα2-3SLacNAc **3**) allowed analyzing the chemical shift perturbation of the amino acids involved in the binding events. In particular, the three ligands showed a similar CSP profile, showing the perturbation of the F-G β-sheet, which corresponds to the binding pocket, with the involvement of the R109 essential for the binding with the sialic acid. Moreover, the presence of the sulfate group in Gal O6 in **2** and **3** caused a perturbation of a new region, the loop C-C', that is located in front of the binding site, suggesting that residues R56, Y58 and Q59 directly interact with this Gal moiety. Our data are in accordance with the results previously published by Propster et al. [2]. The analysis of ligand **3**, which is substituted at C9 by a naphthyl group, produced a CSP in residues located in the N-termini. These chemical shifts can be explained by the rearrangement of some aromatic amino acids (such as Y7, Y11, and Y125), forming a hydrophobic pocket where the naphthyl group is stabilized.

Additionally, the use of STD NMR experiments provided information on the binding epitope from the ligand point of view. For ligand **2**, the primary epitope binding corresponds to H5, H6, H7, H8, $H9_{1,}$ and $H9_{2}$ of the sialic acid moiety; besides, an STD effect was also observed for Gal H5 and H6. On the other hand, for ligand **3**, a strong STD effect was also observed for the protons of the naphthyl group. According to the experimental data from the CSP and STD, a putative 3D structure for the complex has been proposed.

The biological role of the Siglec-8 on mast cells and eosinophils is to modulate the allergic response activated by FcεRI. In particular, the V domain of Siglec-8 can bind glycans on the extracellular domain of FcεRIα and inhibits its function. To evaluate if this binding event is mediated by glycans, this interaction was studied from the Siglec-point of view. In particular, FcεRIα was expressed into two different cell lines to obtain two glycoforms, with all types of glycans (with sialic acid) and with only high-mannose types. The analysis of the $^{1}H,^{15}N$-TROSY spectra of the two different complexes showed a decrease of the intensities of the cross-peaks of siglec-8 only when complex-type glycans were present, suggesting that the binding between Siglec-8 and FcεRIα is mediated by glycan interactions. Moreover, upon the addiction of FcεRIα, most significant decrease in the signal intensities of the amino acids located in the F-G β-sheet and the C-C loop, in the binding pocket, was observed.

Moreover, different constructs of Siglec-8 have been crystallized. Crystals for the Siglec-8 V domain in the apo form and complex with ligand **3** have been obtained and the X-Ray structure of the apo form has been solved. However, for the crystals obtained for the complex with ligand **3**, the crystallization condition needs to be improved. Additionally, the crystals structure of the AK002, the antibody used in clinical trials specific for Siglec-8, and for the complex of the entire extracellular domain of Siglec-8 with AK002 have also been obtained. In this case, the structure refinement and analysis is still under investigation.

6.9 References

[1] S. Duan and J. C. Paulson, “Siglecs as Immune Cell Checkpoints in Disease,” *Annual Review of Immunology*, vol. 38. pp. 365–395, 2020, doi: 10.1146/annurev-immunol-102419-035900.

[2] J. M. Propster *et al.*, “Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 29, pp. E4170-9, Jul. 2016, doi: 10.1073/pnas.1602214113.

[3] B. S. Bochner *et al.*, “Glycan array screening reveals a candidate ligand for Siglec-8,” *J. Biol. Chem.*, vol. 280, no. 6, pp. 4307–4312, 2005, doi: 10.1074/jbc.M412378200.

[4] T. T. Hansel, H. Kropshofer, T. Singer, J. A. Mitchell, and A. J. T. George, “The safety and side effects of monoclonal antibodies,” *Nat. Rev. Drug Discov.*, vol. 9, no. 4, pp. 325–338, 2010, doi: 10.1038/nrd3003.

[5] C. M. Nycholat *et al.*, “A Sulfonamide Sialoside Analogue for Targeting Siglec-8 and-F on Immune Cells,” *J. Am. Chem. Soc.*, vol. 141, no. 36, pp. 14032–14037, Sep. 2019, doi: 10.1021/jacs.9b05769.

[6] S. Duan *et al.*, “Nanoparticles Displaying Allergen and Siglec-8 Ligands Suppress IgE-Fc ε RI − Mediated Anaphylaxis and Desensitize Mast Cells to Subsequent Antigen Challenge,” 2021, doi: 10.4049/jimmunol.1901212.

[7] R. Lieu *et al.*, “Rapid and robust antibody Fab fragment crystallization utilizing edge-to-edge beta-sheet packing,” *PLoS One*, vol. 15, no. 9 September, pp. 1–16, 2020, doi: 10.1371/journal.pone.0232311.

[8] S. J. Stahl, N. R. Watts, and P. T. Wingfield, “Generation and use of antibody fragments for structural studies of proteins refractory to crystallization,” *Methods Mol. Biol.*, vol. 1131, pp. 549–561, 2014, doi: 10.1007/978-1-62703-992-5_35.

[9] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007, doi: 10.1107/S0021889807021206.

[10] L. Unione *et al.*, “Glycoprofile Analysis of an Intact Glycoprotein As Inferred by NMR Spectroscopy,” *ACS Cent. Sci.*, vol. 5, no. 9, pp. 1554–1561, 2019, doi: 10.1021/acscentsci.9b00540.

Chapter VII

General Conclusions

7.1 General Conclusions

The specific conclusions for every chapter have been already described. Herein, I gather the key general conclusions of this Thesis.

- A novel NMR methodology has been presented to determine the glycan composition on an intact glycoprotein using the FcεRIα glycoprotein as proof of concept.
- As application of this NMR-based strategy, the glycan composition of the RBD of the spike glycoprotein of SARS CoV-2 has been deduced. Interestingly, glyco-epitopes not observed in previous MS-based analysis have been found.
- The molecular recognition features of the interaction of the glycosylated RBD with a panel of human lectins have been deduced by using a combined (glycan-based and lectin-based) NMR strategy. The interacting glycan epitopes for galectin-3, -7, -8-N, siglec-8, -10, DC-SIGN and MGL have been elucidated.
- Different variants of Siglec-8 (the V domain and the full D1-D3 ectodomain have been expressed and purified using diverse strategies and protocols, and their interactions with small molecules (glycomimetics), glycoproteins (FcεRI), and antibodies (AK002) have been studied using a combination of X-Ray crystallography and NMR methods.

Thus, this Thesis represents the first stage towards the development of an integrated chemical biology approach, using a variety of techniques, methods, protocols, and strategies to elucidate glycan-protein interactions and to apply the acquired knowledge to the development of new molecules (either chemicals or biologicals) that can interfere the key processes in which glycans are involved.

7.2 Scientific publications during this dissertation

1 Unione L, Lenza MP, Ardá A, Urquiza P, Laín A, Falcón-Pérez JM, Jiménez-Barbero J, Millet O. Glycoprofile Analysis of an Intact Glycoprotein As Inferred by NMR Spectroscopy. ACS Cent Sci. 2019 Sep 25;5(9):1554-1561. doi: 10.1021/acscentsci.9b00540. Epub 2019 Jul 24. PMID: 31572782; PMCID: PMC6764210.

2 Lenza MP, Oyenarte I, Diercks T, Quintana JI, Gimeno A, Coelho H, Diniz A, Peccati F, Delgado S, Bosch A, Valle M, Millet O, Abrescia NGA, Palazón A, Marcelo F, Jiménez-Osés G, Jiménez-Barbero J, Ardá A, Ereño-Orbea J. Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and their Interactions with Human Lectins. Angew Chem Int Ed Engl. 2020 Dec 21;59(52):23763-23771. doi: 10.1002/anie.202011015.

3 Lenza MP, Atxabal U, Oyenarte I, Jiménez-Barbero J, Ereño-Orbea J. Current Status on Therapeutic Molecules Targeting Siglec Receptors. Cells. 2020;9(12):2691. doi:10.3390/cells9122691

7.3 Book chapter

1 A Ardá, S Bertuzzi, I Calloni, A Canales, FJ Cañada, J Ereño-Orbea, B Fernández de Toro, A Gimeno, H Coelho, MP Lenza, MG Lete, J López-Ogalla, F Marcelo, JD Martínez, A Poveda, JI Quintana, P Valverde, J Jiménez-Barbero; NMR in Chemical Glycobiology, In NMR in Chemical Biology. (E. Cabrita, Ed). Royal Society of Chemistry.

2 A. Arda, U. Atxabal, S. Bertuzzi, A. Canales, F. J. Cañada, A. Gimeno, J. Ereño-Orbea, B. Fernández de Toro, A. Franconetti, M. Gómez-Redondo, M. G. Lete, M. P. Lenza, J. D. Martínez, M. J. Moure, A. Poveda, J. I. Quintana, P. Valverde and J. Jiménez-Barbero; Recent advances in the application of NMR methodologies to analyze the conformation, dynamics, and interactions of saccharides. Carbohydrate Chemistry: Chemical and Biological Approaches Volume 44.

7.4 Contribution to congress during this dissertation

Poster awards

1. "NMR structural characterization of the N-linked glycans in the Receptor Binding Domain of the SARS-CoV-2 spike protein and their interactions with human lectins" Euromar. 7-8 December (Virtual congress, **2020**) Journal of Magnetic Resonance - EUROMAR **2020** Young Scientist Award 2020

Flash presentation

1. "Glycoprofile analysis of FcεRIα using NMR Spectroscopy" 2nd Glycobasque Meeting. (Derio, Spain, March 22 **2019**). **[Author]** (presented by Maria Pia Lenza)
2. "An NMR approach to evaluate The glycoprofile analysis of FcεRIα" Introductory Workshop on biomedical Glycoscience. (Donostia, Spain, June 03-05, **2019)**. **[Author]** (presented by Maria Pia Lenza)
3. "Deciphering the N-glycan profile of FcεRIα by NMR" Workshop at CICbioGUNE. (Derio, Spain, December 2-3, **2019)**. **[Author]** (presented by Maria Pia Lenza)
4. "Deciphering the N-glycan profile and interactions of FcεRIα by NMR" V GEQB. (Granada, Spain February 19-22, **2020**). **[Author]** (presented by Maria Pia Lenza)
5. "Deciphering the N-glycan profile of FcεRIα by NMR" 3nd Glycobasque Meeting. CIC bioGUNE, (Derio, Spain, March 11-13, **2020**). **[Author]** (presented by Maria Pia Lenza)
6. "NMR structural characterization of the N-linked glycans in the Receptor Binding Domain of the SARS-CoV-2 spike protein and their interactions with human lectins" Euromar. 7-8 December (Virtual congress, **2020**)

Poster communication

1. "Glycoprofile analysis of FcεRIα using NMR Spectroscopy" 2nd Glycobasque Meeting. (Derio, Spain, March 22 **2019**). **[Author]** (presented by Maria Pia Lenza)

2. XXXVII Reunión Bienal de la Real Sociedad Española de Química. (Donostia, Spain May 26-30, **2019**). **[Author]** (presented by Maria Pia Lenza)

3. "An NMR approach to evaluate The glycoprofile analysis of FcεRIα" Introductory Workshop on biomedical Glycoscience. (Donostia, Spain, June 03-05, **2019)**. **[Author]** (presented by Maria Pia Lenza)

4. "Glycoprofile analysis of FcεRIα using NMR Spectroscopy" Eurocarb XX. (Leiden, The Netherlands, 30 June-5 July, **2019**). **[Author]** (presented by Maria Pia Lenza)

5. "Deciphering the N-glycan profile of FcεRIα by NMR" Workshop at CICbioGUNE. (Derio, Spain, December 2-3, **2019)**. **[Author]** (presented by Maria Pia Lenza)

6. "Deciphering the N-glycan profile and interactions of FcεRIα by NMR" V GEQB. (Granada, Spain February 19-22, **2020**). **[Author]** (presented by Maria Pia Lenza)

7. "Structural characterization of the N-glycans on RBD SARS-CoV-2 Spike protein and their interactions with human lectins" 9th Ibero-American NMR meeting/7th Iberian NMR meeting. (Virtual meeting, April 26-29, **2021**). **Author]** (presented by Maria Pia Lenza)