Università degli Studi di Napoli Federico II

Ph.D. Program in
Computational and Quantitative Biology

XXXVII Cycle

Thesis for the Degree of Doctor of Philosophy

Physico-chemical Analysis of Siglec-Glycan Complexes

by

Tania Gerpe Amor

Advisor: Prof. Alba Silipo

Co-advisor: Prof. Roberta Marchetti

Scuola Politecnica e delle Scienze di Base
Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

A ti, mamá.

"Porque nadie muere mientras
permanece vivo en los recuerdos"

IMPERATOR ROMANORVM SEMP AVGVST FRIDERICVS

Ph.D Thesis presented

for the fulfilment of the Degree of doctor of

Philosophy in Computational and Quantitative

Biology

Physico Chemical Analysis of Siglec-Glycan Complexes

by

Tania Gerpe Amor

October 2024

Approved as to style and content by

Prof. Alba Silipo, Advisor

Prof. Robert Marchetti, Coadvisor

Università degli Studi di Napoli Federico II

Ph.D Program in Computational and Quantitative Biology

XXXVII cycle – Chairman; Prof. Michele Ceccarelli

Candidate's declaration

I hereby declare that this thesis submitted to obtain the academic degree of Philosophiæ Doctor (Ph.D.) in Computational and Quantitative Biology is my own unaided work, that I have not used other than the sources indicated, and that all direct and indirect sources area acknowledged as references.
Part of this dissertation is currently being published in international journals and/or conference articles.

Napoli, October 14th, 2024

Tania Gerpe Amor

FRIDERICVS SECVNDVS ROMANORVM IMPERATOR SEMPER AVGVSTVS

Contents

Funding

The research presented in this dissertation has been supported by the Marie Skłodowska-Curie Actions (MSCA), supporting a PhD scholarship within an Early-Stage Research (ESR) enrolment under the coordination and supervision of Prof. Alba Silipo.

The author's work has been fully carried out in the framework of the European Union project of the Horizon program "GLYTUNES", H-2020 grant agreement N. 956758

List of Acronyms

The following acronyms are used throughout the thesis.

NMR	Nuclear Magnetic Resonance
MD	Molecular Dynamics
MM	Molecular Mechanics
QM	Quantum Mechanics
ECM	Extracellular Matrix
MAMP	Micro-associated molecular pattern
DAMP	Danger-associated molecular pattern
CDAM	Cell-death associated molecular pattern
SAMP	Self-associated molecular pattern
PRR	Pattern Recognition Receptor
ER	Endoplasmatic reticulum
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif
GBS	Group B *Streptococcus*
SLE	Systemic Lupus Erythematosus
ITC	Isothermal Calorimetry
SPR	Surface Plasmon Resonance
ELISA	Ensyme-linked immunoassay
STD	Saturation Transfer Difference
HSQC	Heteronuclear Single Quantum Correlation
HMBC	Heteronuclear Multiple Bond Correlation
COSY	Correlation Spectroscopy

TOCSY	Total Correlation Spectroscopy
NOESY	Nuclear Overhauser Spectroscopy
ROESY	Rotating-frame Overhausesr Effect Spectroscopy
CSP	Chemical Shift Perturbation
CSM	Chemical Shift Mapping
TROSY	Transverse Relaxation Optimized Spectroscopy
HF	Hartree-Fock
PHF	Post-Hartree-Fock
MC	Monte Carlo
AD	AutoDock 4.2
GA	Genetic algorithm
EP	Evolutionary programming
LS	Local search
gg	gauche-gauche
gt	gauche-trans
tg	trans-gauche
Nm	*Neisseria meningitidis*
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
OMV	Outer membrane vesicle
MoDC	Monocyte-derived dentritic cell
CRC	Colorectal cancer
PDB	Protein Data Bank
ECD	Extracellular domain
TM	Transmembrane Domain

List of Abbreviations

The following abbreviations are used throughout the thesis.

Glc	Glucose
GlcNAc	N-Acetyl Glucosamine
Man	Mannose
Gal	Galactose
Rha	Rhamnose
GlcA	Glucuronic acid
GalNAc	N-Acetyl Galactosamine
Neu5Ac	5-N-acetylneuraminic acid
Sia	Sialic acid
Neu5Gc	N-glycolylneuraminic acid
KDN	2-keto-3-deoxy-D-glycero-D-galacto-nononic acid
Siglec	Sialic acid immunoglobulin-type lectin
Asn	Asparagine
Ser	Serine
Thr	Threonine
Arg	Arginine
Lys	Lysine
Tyr	Tyrosine
Glu	Glutamic acid
Trp	Triptophan
Asp	Aspartic acid
Ala	Alanine
Ile	Isoleucine

LPS	Lipopolysaccharide
LOS	Lipooligosaccharide
OS	Oligosaccharide
TLR	Toll-Like Receptor
GBPs	Glycan binding proteins
Ig	Immunoglobulin
NK	Natural Killer
Hz	Herz
MenY	Capsular polysaccharide meningococcus Y
MenW	Capsular polysaccharide meningococcus W
MenB	Capsular polysaccharide meningococcus B
OAc	O-Acetylated
SleX	Sialyl Lewis X
MUC	Mucine
3'SLN	3'-sialyllactosamine
SNFG	Symbol nomenclature for glycans

Abstract

Cells possess glycans at the interface between the environment and the cell membrane. The specific recognition of glycans amongst glycan binding proteins (GBP) triggers many crucial functions, especially in infectious processes, immune responses and inflammation. Siglecs, sialic immunogloblin type lectins, are GBPSs that hold critical functions, acting as immune checkpoints through its specific binding with sialic acid.

This PhD dissertation tackles some inhibitory Siglecs that downregulates the immune cell activation. In detail, Siglec-7, Siglec-9 and Siglec-10, which all contain a V-set domain, one or several C2-Ig domains and their critical tyrosine-based inhibition motifs (ITIMs), have been investigated in the context of disease and therapeutic targets, focusing on their ability to interact with both endogenous and exogenous ligands. Siglec-7, expressed on NK cells, was studied through a multidisciplinary approach, including ligand and protein NMR experiments, fluorescence, ELISA and computational techniques to successfully elucidate the structural and mechanistic aspects of the interactions with pathogenic exogenous ligands. Siglec-9, expressed on leukocytes, primarily engages with endogenous ligands, holding an important role in inflammation and autoimmune diseases. It has been studied through a first comprehensive approach based on computational analysis, to set the basis for a novel therapeutic target for the control of autoimmunity, by engaging immune tolerance. Finally, Siglec-10, expressed on B and dendritic cells, was analyzed in the context of a potential therapeutic target through the study with synthetic and glycomimetics, mimicking natural sialic acids by a first computational approach. Therefore, the molecular details of different Siglecs amongst both endogenous and exogenous ligands have been detailed provide novel insights for the understanding and design of future therapeutic approaches targeting different human diseases.

Keywords: Computational chemistry, sialoglycans, Siglecs, Molecular recognition, NMR.

Chapter

I

Introduction

AVGVST FRIDERICVS TO IMPERATOR ROMANORVM SEMP

I. Introduction

1.1 Glycans and their biological roles

Glycans, often called carbohydrates, are complex molecules composed of monosaccharide units linked by glycosidic bonds. They are one of the four fundamental macromolecules of life, alongside proteins, lipids, and nucleic acids. Glycans can be simple monosaccharides or complex polysaccharides, often existing as glycoconjugates—molecules where glycans are covalently attached to proteins (glycoproteins) or lipids (glycolipids).[1] They hold an enormous structural diversity, arising from the variety of monosaccharide building blocks where, differently to proteins and nucleic acids, they can be linear or even branched.[2] Another characteristic that contributes to their heterogeneicity is the different types of glycosidic linkages that connect the different aglycon substitutions they may have such as: acetyl, sulphate groups or even with peptides and lipids. This high structural diversity creates a plethora of different functions where they can be involved, playing indeed crucial roles in several biological processes. [3]

Glycans and glycoconjugates are often found in the cell-wall/extracellular matrix (ECM), or in vesicles in transit to/from the cell surface. Those present on the cell surface are mainly involved in cell recognition and signaling processes with glycoproteins and glycolipids on the plasma membrane serving as recognition sites for other cells, pathogens, and molecules.[4] This recognition is critical for the immune response and cellular interactions.[5,6] Glycosylation, the process of adding glycans to proteins, is essential for proper protein folding and stability, affecting protein solubility, resistance to proteases, and overall functionality through N-linked and O-linked glycosylation. Therefore, glycans play a pivotal role in the immune system, modulating the effector functions of antibodies, and recognizing unique glycan patterns on pathogens, thereby triggering an

immune response.[7] Several glycan-protein interactions play a crucial role in maintaining the delicate balance between immune tolerance and a robust immune response. Throughout evolution, the glycocalyx has developed the ability to distinguish between endogenous (*self*) and exogenous (*non-self*) components. The innate non-specific response serves as the first line of defense against non-self, foreign microorganisms.[8] In this process, proteins expressed by innate immune cells—such as macrophages, dendritic cells, neutrophils, monocytes, and epithelial cells—known as Pattern Recognition Receptors (PRRs), are activated against microbial organisms. These include microbe-associated molecular patterns (MAMPs), mainly composed of bacterial glycoconjugates, danger-associated molecular patterns (DAMPs), and cell death-associated molecular patterns (CDAMs). These interactions activate cellular mechanisms that trigger proinflammatory cytokines and stimulate adaptive immunity.[9] During development, glycan structures can also change dynamically, influencing cell differentiation and organogenesis, acting as markers for different cell types and developmental stages.[10]

In addition to their role in immune responses, glycans are also implicated in tumor progression. Tumor cells often exhibit altered glycosylation patterns, which can promote cancer cell proliferation, invasion, and metastasis. These altered glycan structures can modulate the immune response, helping tumor cells in evading immune detection and destruction. By interacting with various glycan-binding proteins, such as lectins, these glycans can influence cell signaling pathways and the tumor microenvironment, further contributing to tumor progression and immune escape.[11]

1.2 The structure of glycoconjugates and their classification

Glyconjugates are found at the extracellular surface of eukaryotic and prokaryotic cells conforming, together with different glycans, the glycocalyx. The composition and structure of the glycocalyx is heterogeneous among cell types and species. It forms a gel-like meshwork

that has several roles in the immune system and among them, some remarkable ones are the control of the vascular permeability, since it creates a barrier between the cell and its surrounding.[12] It also serves to inhibit leukocyte-endothelial interactions, additionally capable of dampen coagulation factors at the endothelial surface as a mediator for cell-cell interactions.[13]

Glycosylation is largely present on lipids and proteins on the surface of cell, and several studies have demonstrated that glycans can also be present in nucleic acids. The glycosylation process occurs between the endoplasmatic reticulum (ER) and the Golgi apparatus through the combination of a set of glycosyl transferases and glycosidases of a given biological system.[14,15]

Among all the previously mentioned biological functions involving glycans and glycoconjugates, they can be divided into two categories: i) intrinsic functions and ii) extrinsic functions. The intrinsic functions include structural and modulatory activities of glycans. In contrast, the extrinsic functions encompass interactions of glycans with other receptors, including lectins, and even involve immune evasion mechanisms developed by pathogens.

Lectins, a type of proteins that specifically bind to glycans, play a crucial role in mediating these extrinsic interactions, influencing various cellular processes and pathogen recognition. This thesis focused on the extrinsic functions of glycans and particularly their interactions with lectins and their implications for immune evasion by pathogens. [16]

Understanding conformation and structural properties of glycoconjugates [17] is crucial to acknowledge information at the fundamental level to be able to further develop new promising applications in the prevention, diagnosis and target therapies for the future. Typically, they consist in glycolipids and glycoproteins. Glycolipids form stable micelles in aqueous solutions due to their amphiphilic nature. There are different types, but mammals have two big groups:

i) Glycosphingolipids: they consist of a glycan moiety linked to a ceramide, which is a lipid composed of sphingosine and fatty acids. Glycosphingolipids are primarily involved in cell-cell interactions and cell adhesion events. (Figure 1.1)

ii) Glycoglycerolipids: comprising glycerol, lipids and carbohydrates, these glycolipids play different structural roles, such as the maintenance of the membrane bilayer stability, serving as the precursors for complex membrane components, therefore serving also as the anchorage of different proteins to cell membranes.

In glycoproteins, glycans are linked to different proteins as a critical post-translational modification produced in mammals.[18] They are formed by the attachment of the carbohydrate portion through an N-glycosidic linkage to an asparagine residue, forming N-glycans, or also to a serine of threonine residue through an O-glycosidic linkage, thus forming O-glycans.

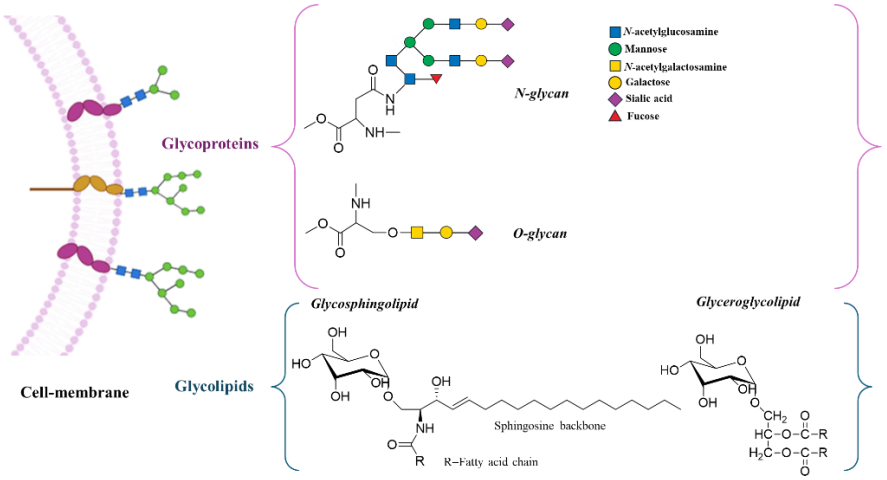

Figure 1.1 Schematic representation of glycolipids and glycoproteins in the cell membrane. Glygoproteins can be synthesized as N-glycans and/or O-glycans. Glycolipids are divided into glycosphingolipids and glycerolipids.

N-glycosylation initiates with the addition of *N*-acetyl glucosamine (GlcNAc) to the asparagine side, this attachment occurs through recognition of a specific sequence motif, Asn-X-Ser/Thr, where X represents any amino acid except proline. *N*-glycans structures hold a common $(GlcNAc)_2(Man)_3$ core to which different glycans can be linked afterwards, including Mannose (Man), GlcNAc (GlcNAc), Galactose (Gal), Fucose (Fuc) and Sialic acid (Neu5Ac)[19] through different type of glycosidic linkages, arriving to even bi- tri or tetra-antennary structures. In a subsequent stage of glycosylation, referred to as trimming, the core region of the *N*-glycan undergoes modifications involving the removal of three glucose residues and one mannose[20] from the chain by specific glycosidases. This process facilitates the translocation of the glycoprotein to the Golgi apparatus. Within the Golgi, the glycoprotein undergoes further modifications mediated by specific glycosidases and glycosyltransferases tailored to its intended biological functions.[21] These modifications contribute significantly to the heterogeneity observed in N-glycan structures.

Therefore, N-glycans are classified into three main groups, all share a chitobiose core composed of three Man and two GlcNAc, one of which is covalently linked to the Asn residue. These groups are known as: i) high-mannose, ii) complex-type, and iii) hybrid. (Figure 1.2A)

In the O-glycosylation process, the initial step involves adding N-acetylgalactosamine (GalNAc) to the free hydroxyl group of a serine or threonine side chain. This modification can lead to the formation of various structures, classified into four primary glycan core structures, with an additional alternative core structure based on the presence of the sialyl-T antigen (Figure 1.2B). These O-glycans exhibit significant heterogeneity. The attachment of GalNAc requires the involvement of several enzymes and constitutes a post-translational modification. Despite protein synthesis occurring in the endoplasmic reticulum, the attachment of GalNAc residues occurs in the Golgi apparatus.[22] The subsequent

elongation and termination is carried out by several glycosyltransferases. Its cellular distribution determines the outcome of the O-glycan biosynthesis. O-linked glycans termination often includes Gal, GlcNAc, GalNAc, Fucose or sialic acid. O-glycans are commonly known as mucin-type O-glycans due to their prevalent occurrence in mucins, which are extensively studied glycoproteins. These mucin-type O-glycans are predominantly linear and play crucial roles in inflammation processes and cell signalling events, as mucin glycoproteins are primary components of mucus, serving as a barrier against pathogen invasion.

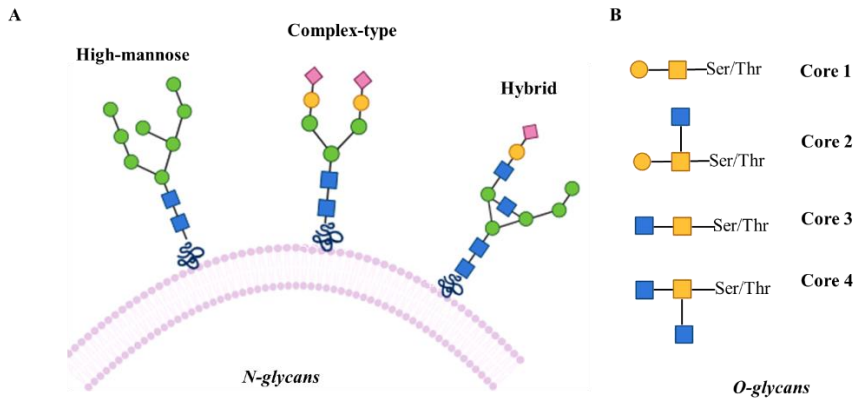

Figure 1.2 Classification of *N*-glycan structures and *O*-glycans cores.

1.3 Sialic acid and Molecular mimicry of host glycans

Sialic acids (Sias) belong to the α-keto aldonic acids. They hold a nine-carbon backbone, and they decorate the outermost part of the N-glycans, O-glycans and glycosphingolipids. (Figure 1.3) Hundreds of sialylated glycan structures have been identified and among them, the N-acetylneuraminic acid (Neu5Ac, 5-(Acetylamino)-3,5-Dideoxy-D-glycero-D-galacto-2-nonulosonic Acid), N-glycolylneuraminic acid (Neu5Gc) and 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid, (KDN) are, by far, the most abundant naturally occurring Sias.[23],[24] Their anomeric position is located at the C-2 and they are usually linked with an α-(2,3), or α-(2,6) linkages to the hydroxyl group to galactose or *N*-

acetylgalactosamine residues, forming typical sialylated glycans. Some enzymes called sialyltranferases can also form disialyl core structures that are commonly known as gangliosides. They are characterized by the existence of at least two sialic acid residues that can be attached to each other by α-(2,3) α-(2,6) or α-(2,8) glycosidic linkages.[25]

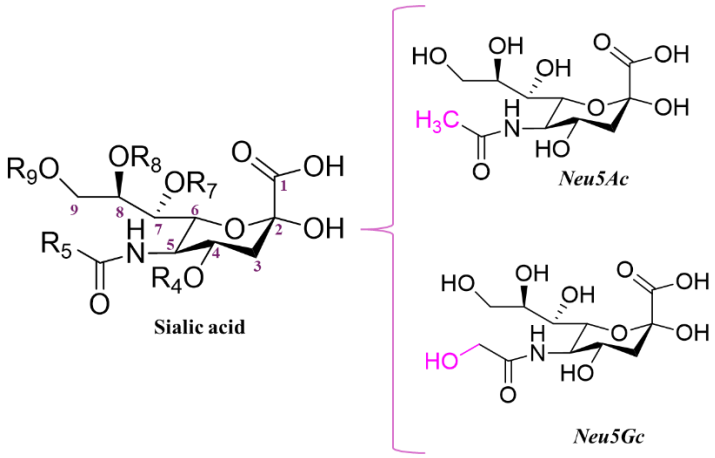

Figure 1.3 Sialic acid structure. As shown, this nine-carbon sugar possess α-configuration. On the right, there are represented the most common sialic acid structures in mammals: the acetylated (Neu5Ac) and the glycolylated (Neu5Gc) neuraminic acids.

This unit is biosynthesized by an enzymatic addition of a hydroxyl group to an N-acetyl moiety at the 5'-position of Neu5Ac, being catalyzed by a hydroxylase enzyme, commonly known as cytidine monophospho-N-acetyl neuraminic acid hydroxylase. Differently to other mammals like chimpazees, humans have lost the glycolylated form of sialic acid, Neu5Gc due to the inactivation of the genomic mutation in the CMAH (Figure 1.4), that provoked the gene loss in the human-lineage. [26] However, even though humans have the inability to synthetize the Neu5Gc, it can be assumed by diet, i.e. from red meat and cow's milk, and then metabolically introduce it following the Neu5Ac biochemical pathway (Figure 1.4). Different studies often associate the presence of Neu5Gc with carcinoma.

It is present in tumours like melanoma, retinoblastoma, colon cancer and breast cancer. This makes sialylated glycans as target for the study and design of innovative therapeutic strategy for cancer and autoimmune diseases. [27]

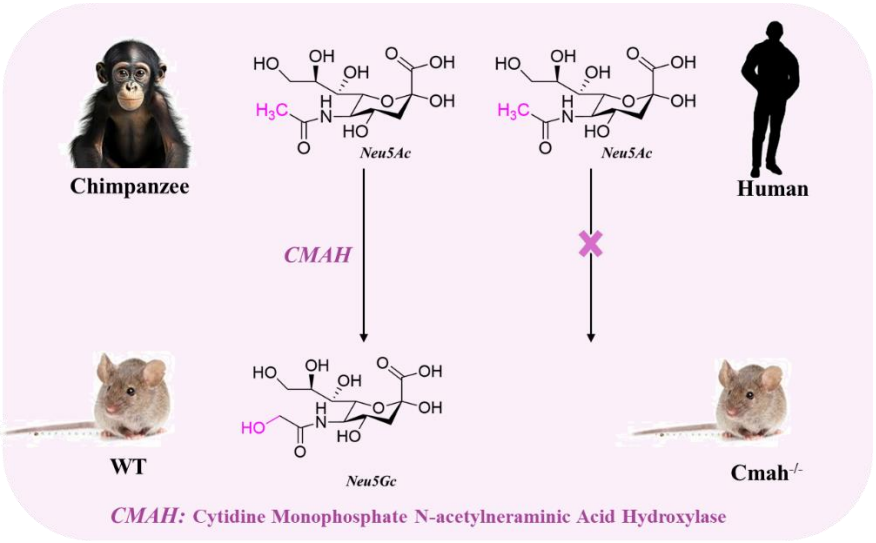

Figure 1.4 Loss of the glycolylated neuraminic acid, Neu5Gc, in humans.

5'-Acetylated sialic acids, Neu5Ac, are highly expressed on the outer counterparts of cell membranes. Their negative charge and hydrophilic properties permitted them to have roles in both normal and pathological processes. With their attachment to different glycans, they affect the structure of glycoconjugates being the most suitable ligands for lectins, antibodies and different enzymes. Sialic acids (Sias) are linked to macromolecules by labile ketosidic bonds, and they are usually found attached to galactose, N-galactosamine, N-acetylgalactosamine, via α-2,3 or α-2,6 linkage or as homopolymers, with α-2,8 or α-2,9 linkages. They can also contribute to biological functions, as ligand for different molecules like hormones, immune proteins, but can also form part of viruses and bacteria as part of the molecular mimicry process to evade immune mechanisms.[28]

1.4 Glyconconjugates on bacterial cell wall from gram negative bacteria.

Microbial cell wall glycoconjugates are essential in mediating host-guest interactions, significantly impacting both health and disease states.[29]

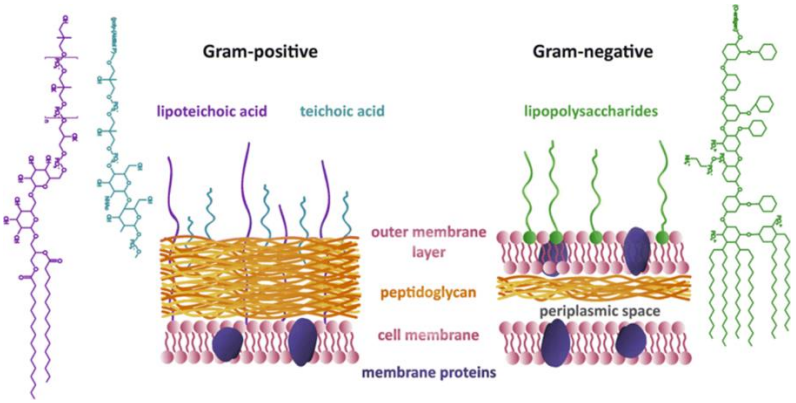

Figure 1.5 Structural arrangement of cell membranes from Gram-positive (left) and Gram-negative bacteria (right). Difference is based on the thickness of the peptidoglycans.

In bacteria, cell wall components such as lipopolysaccharides, peptidoglycans, and teichoic acids serve as crucial virulence factors. These glycoconjugates, recognized as PAMPs[30] (Pathogen-Associated Molecular Patterns), have the unique ability to trigger the host's immune response due to their inflammatory properties. By activating innate immunity, they play a pivotal role in the body's defense mechanisms. However, these interactions also facilitate immune evasion and suppression strategies employed by pathogens, complicating infection outcomes and contributing to chronic disease states. Understanding these processes is vital for developing targeted therapies and improving infection management.

Apart from the peptidoglycan contained in the bacterial cell wall we can encounter different components divided by:

I) Gram negative bacterial cell-wall glycoconjugates

The primary component of the outer membrane in Gram-negative bacteria is undoubtedly, the lipopolysaccharides (LPS). These are highly

acylated glycolipids, and they play a crucial role in maintaining the membrane permeability, acting as a protective barrier. They also have the ability to stimulate innate or natural immunity due to their significant involvement in host-pathogen interactions. (Figure 1.6) The structure of LPS varies across bacterial strains, defining their immunological function. LPS is an amphiphilic (both hydrophobic and hydrophilic) macromolecule composed of three distinct regions: lipid A, the core oligosaccharide (core OS), and a polysaccharide portion (O-chain). The complete form of LPS is known as smooth-type LPS (S-LPS), while some LPS terminate at the core oligosaccharide and are referred to as rough-type LPS (R-LPS) or lipooligosaccharide (LOS) (Figure 1.7). The hydrophobic lipid A is the anchor of the LPS to the outer bacterial membrane and serves as the endotoxin portion, responsible for its immunostimulatory potential. It can trigger the release of proinflammatory mediators through Toll-like receptor 4 (TLR4)-dependent signaling in macrophages[31] and endothelial cells, leading to various biological effects, including sepsis and septic shock. Lipid A[32] consists of two β-(1→6)-linked D-glucosamine units, which are variably acylated by fatty acids of different lengths and phosphorylated at the anomeric position of the reducing (GlcNI) unit and the 4' position of the non-reducing (GlcNII) residue. The inner core OS contains characteristic monosaccharides such as heptoses (L-glycero-D-manno heptose or D-glycero-D-manno heptose) and Kdo (3-deoxy-D-manno-octulosonic acid), the latter being a marker of all Gram-negative bacteria and covalently linked to the GlcN II of lipid A.[33] The outer core OS is usually composed of hexoses, deoxyhexoses, uronic acids, and aminosugars. The external portion of LPS consists of the hydrophilic O-polysaccharide, a polymer of saccharide repeating units (up to eight sugars) which can be repeated up to 50 times. The O-chain is highly variable and heterogeneous, being linear or branched, homo- or heteropolymeric, and may include non-carbohydrate substituents.

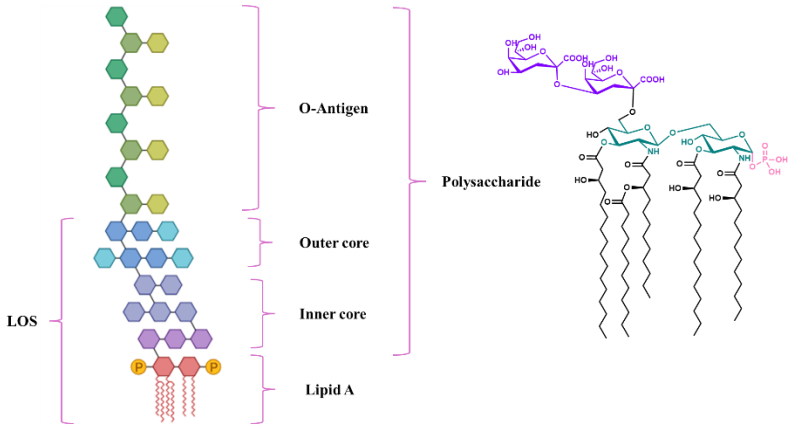

Figure 1.7 Representation of LPS structure on Gram-negative membrane. The main components of LPS are: lipid A, core and O-antigen. Structures lacking O-antigen consist of LOS. On the right, an example of Kdo-α-(2→4)-Kdo (in purple) connected to β-(1→6)-linked D-glucosamine residues (GlcN I and II, in acquamarine) of a general LPS structure.

1.5 Bacterial capsular polysaccharides (CPS)

Capsular polysaccharides (CPS) are complex carbohydrates structures that can form a protective layer, called capsule, that surrounds the cell wall of certain bacteria. This capsule plays a crucial role in the pathogenicity and immune evasion mechanism of these bacteria. The understanding of structure and functions of CPS is key in microbiology, immunology, and vaccine development.[34,35]

They are composed of long polysaccharides chains; their specific composition can vary widely among different bacteria species and strains. The most common sugars found in CPS include Glc, Gal, Man, Rha, glucuronic acid (GlcA) and sialic acid (Sia).[36]

They play different roles in biological systems. On the one hand, they serve as protection from host defence, since the capsule can act as a physical barrier by protecting the bacterial cell form the phagocytosis of immune cells like macrophages and neutrophils. This allows pathogens to evade the host's immune response, improve resistance to antimicrobial

peptides by the penetration of these molecules, enhancing the bacterial survival. Moreover, there are capsules able to prevent dehydrating processes, by retaing water.

The pathogenic role of capsules is one of the most studied nowadays in this field.[37] They are considered virulence factors due to the immune evasion and survival mechanisms in the host. The bacterial capsules are often more virulent than the species and strains that lack the capsule counterparts. They are abundant in species like: *Streptococcus pneumoniae*, causing pneumonia and meningitidis, *Haemophilus influenzae type b* leading to meningitis, epiglottitis and other serious infections often seen in children; *Neisseria meningitidis*, as responsible of bacterial meningitis and septicemia and even some strains of *Escherichia coli*, causing urinary tract infections.[38]

-*Sialylated Capsular Polysaccharides*

The sialylated capsular polysaccharides are a subset of CPS that contain the sialic acid attached to their polysaccharide chains[39],[40]

Host receptors that are able to recognize sialylated capsular polysaccharides are typically involved in cellular communication and immune regulation. These are receptors that bind specifically sialic acids, so bacterial capsules facilitate the pathogen to perform adhesion and immune evasion mechanisms.[41] Within these receptors are found different proteins like: Siglecs (sialic acic-binding immunoglobulin-type lectins), selectins, Siglec-like, complement factor H and DC-SIGN, among others.

1.6 Immune regulators – lectins: Siglecs and their role

The interactions between glycans and receptors are vital for cellular mechanisms crucial for both health and disease. Glycan-binding proteins (GBPs) are the specific receptors that interact with glycans, including lectins, toxins, microbial adhesins, antibodies, and enzymes. [42] Each GBP has a unique glycan recognition domain that ensures selective glycan

binding, paired with other domains that translate this binding into a functional biological response. These receptors can be either intrinsic or extrinsic. Intrinsic GBPs are involved in direct cell-to-cell interactions and glycan binding within the same cell. The most representative lectins are classified into C-type lectins, I-type lectins, P-type lectins, and S-type lectins or galectins, which are present on the surface of immune cells and classified according to their structure.[43] They play crucial roles in immunity, including inflammation and immunomodulation, by initiating intracellular signaling and binding to cell surface targets to trigger specific immune responses (Figure 1.8).[44,45]

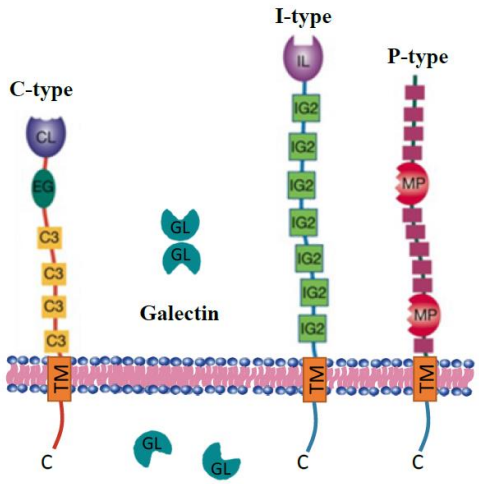

Figure 1.8. Classifications of lectins. C-type lectins bind sialoglycans by being calcium-dependent; I-type lectins contain an immunoglobulin-like carbohydrate recognition domain (CRD); P-type lectins are specific to glycoproteins containing mannose 6-phosphate; galectins are thiol-dependt soluble proteins and specific to B-galactosides; C-type lectin CRD (CL), galectin CRD (GL), P-type lectin CRD (MP), I-type lectin CRD (IL), EFG-like domain (EG), immunoglobulin C2-set domain (IG2), complement regulatory repeat (C3) transmembrane region (TM)

Siglecs: Structure and Function

Siglecs are transmembrane I-type receptors that exhibit diversity in their sequence length and specificity for various sialic acid-containing ligands.[46] To date, 15 human Siglecs and 9 murine Siglecs have been identified, as shown in Figure 1.9. Human Siglecs are categorized based on sequence homology and conservation across orthologs into "evolutionary conserved" and "CD33/Siglec-3 related" sub-categories.[47] Siglecs-1, -2, -4, and -15, which display low sequence similarity but a high degree of gene preservation in mammalian species, belong to the evolutionary conserved group. Siglecs with high homology (above 50%) to Siglec-3, including Siglecs 5 to 16, belong to the CD33 related group.[48]

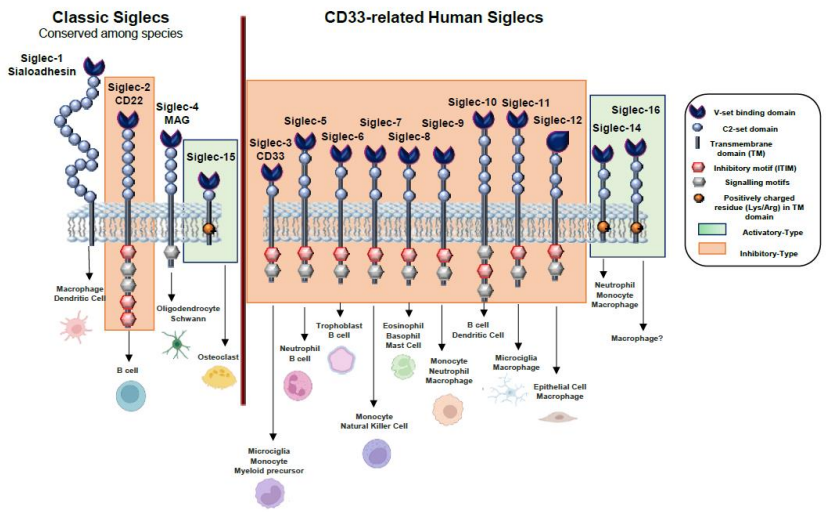

Figure 1.9 Scheme of human Siglecs. Siglecs family divided in "evolutionary conserved" and CD33 related. The biological functions of Siglecs depend on the nature of the cytoplasmic tails and the transmembrane domain features that are able to distuinguish between activatory and inhibitory proteins.

Structurally, these receptors are composed of an extracellular N-terminal V-set (Ig-V) domain responsible for binding different sialoglycans. This V-set domain is directly connected to one to 16 C2-set Ig domains through disulfide bridges (Figure 1.10). The extracellular portion is followed by a

single-pass transmembrane domain that connects to a variable number of cytosolic tails, strongly influencing the biological function of Siglecs[49]. In their cytoplasmic region, Siglecs contain one or multiple immunoreceptor tyrosine-based inhibitory motifs (ITIMs) capable of triggering cellular signaling. These ITIMs can inhibit immune cell stimulation through binding and activation of SH-2 domain-containing protein tyrosine phosphatases such as SHP-1 and SHP-2. Interestingly, a few Siglecs, including Siglec-14, Siglec-15, and Siglec-16, display activating signalling properties.

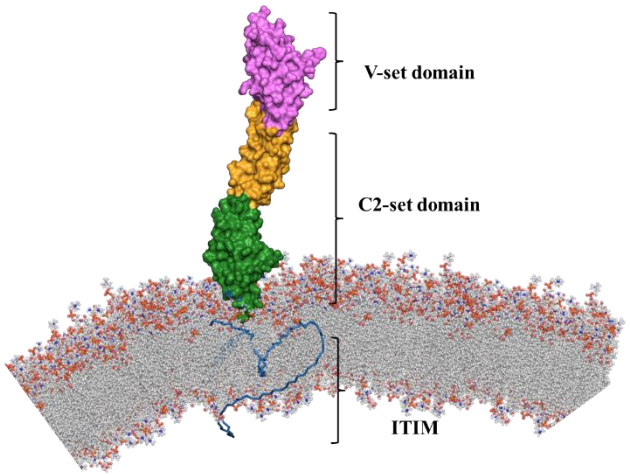

Figure 1.10. 3D representation on the details of a general Siglec structure. Represented in grey stick and balls is the membrane. In blue cartoon the ITIM; surfaced in lightorange and green the C2-set domain, and in pink surface the V-set domain.

The extracellular N-terminal V-set domain (Ig-like) consists of nine β-strands (named A to G) arranged in a distinctive topology.[50] This domain shows close sequence homology to the Ig V regions and specifically accommodates sialic acid from sialoglycans in the F-β strand. The primary interaction with sialic acid is through a salt bridge formed between its carboxylate group and the guanidinium moiety of a conserved arginine

residue, common to all Siglecs. A structural feature is the presence of an intra-sheet disulfide bond between the B and E β-strands, which increases the separation between the β-sheets and allows greater exposure of two aromatic residues in conserved positions on the A and G strands. These residues interact with the protons of the lateral glycerol chain and the N-acetyl group of the sialic acid. Recognition also involves the CC-loop of Siglecs, which outlines the F-G strand of the binding site. This loop displays variable sequences among Siglecs and plays a significant role in the specificity of recognition of different sialylated ligands. This specificity is achieved through a conformational change (from open to closed) upon interaction with sialic, as observed on the complex of the GT1b ganglioside containing Neu5Ac-α-2,8-Neu5Ac and Siglec-7.

Siglecs and sialic acid interactions can regulate immune signalling, thereby contributing to fundamental mechanisms in cell signalling, cell-cell, and host-pathogen interactions.[51,]

Recognition between Siglecs and sialogylcans

Siglecs are able to recognize glycans exposed on the surface of the same cells (*cis* interactions) or with sugars from different cells (*trans* interactions) interactions, as depicted in Figure 1.11.

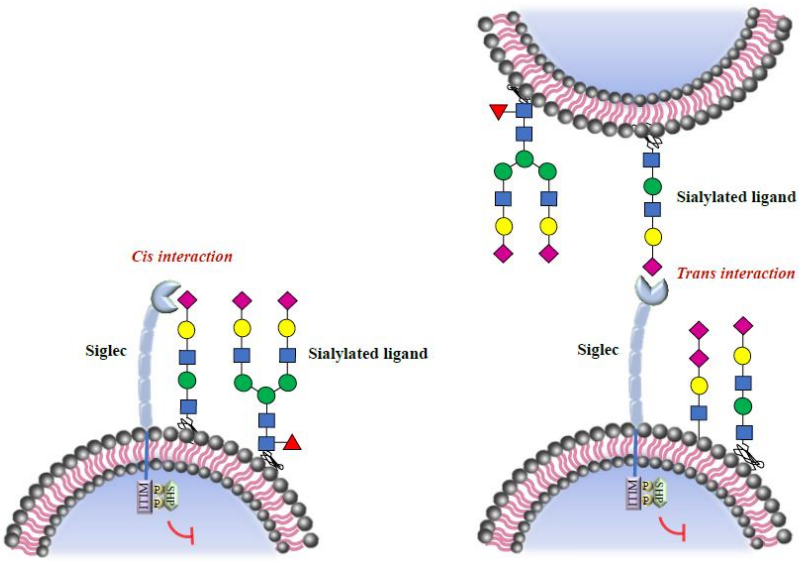

Figure 1.11 Siglec-sialoglycan binding mode. Siglecs are able to interact with a ligand express on the same cell (*cis*) or a different cell surface (*trans*)

There is an equilibrium between *cis* and *trans* mechanisms of Siglec interactions, depending on ligand affinity and the availability of Siglecs. Due to the high local concentration of sialylated glycans on immune cell surfaces, Siglec binding pockets are generally masked by *cis* interactions, outcompeting *trans* ones (Figure 1.12).[52] Siglec-1 is an exception, as it possesses a long and extended structure that avoids *cis* interactions because its V-set domain is too far from the cellular membrane. Additionally, the action of membrane sialidases, cellular activation, or plasma membrane domain reorganization can facilitate interactions with *trans* ligands by "unmasking" Siglecs. It is noteworthy that "self" endogenous ligands (*cis* interactions) may not harness the *trans* interactions on adjacent cells due to their relatively low affinity compared to the high avidity of multivalent sialylated probes or higher-density trans ligands on sialylated pathogens and tumors. As a result, this can trigger stronger inhibitory signals compared to those (mainly in *cis*) delivered by self-endogenous ligands, shifting from *cis* to *trans* interactions. Overall, the binding between Siglecs and *cis* ligands has a significant impact on the regulation of main cellular functions, preventing non-specific cell-cell interactions that might stimulate unnecessary signaling. This type of *cis* signaling may be produced by inhibitory Siglecs that are able to promote or exclude the association with an activating receptor.

The modulation of the immune response by Siglecs is directly correlated with their ability to discriminate between "self" (endogenous) and "non-self" (exogenous) ligands. Due to this, some human pathogens like Group B *Streptococcus* (GBS), certain *Neisseria* species and *Campylobacter jejuni*, have matured the capability to evade immune mechanisms by covering their outer cell surfaces with sialic acid. These exogenous molecules mimic the self-associated molecular pattern (SAMP) structures

such as sialylated capsular polysaccharides (CPS) or lipooligosaccharides (LOS) and are mistakenly recognized as "self" host molecules, hence avoiding immune response activation and promoting host colonization. [53]

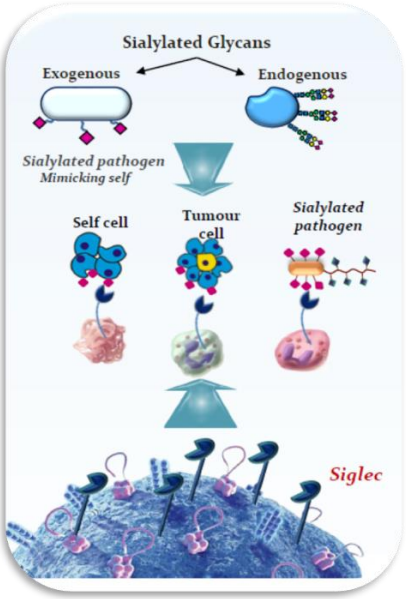

Figure 1.12 Schematic representation on the Siglec axis interacting with exogenous ligands.

In addition, some overexpressed tumoral cells promote the tumorigenesis thus, enhancing the tumor growth by covering their outer surface with sialylated glycans. For these reasons, both pathogenic and tumoral progressions, the Siglecs have been considered as targets for the design of therapeutic agents, like antibodies or glycomimetics for the treatment of several inflammatory, autoimmune, pathogenic and tumoral diseases.

1.6.1 Siglec - 7

Siglec-7 is a member of the human CD-33 related inhibitory Siglec receptors, which means that it acts as negative regulator of the innate immune system. It is mainly expressed on Natural Killer cells, and also found in T cells, eosinophils, monocytes and dendritic cells. The extracellular domain of Siglec-7 is mainly characterized by an N-terminal

V-set domain common of all Siglecs and two C2 Ig domains.[54] It has a preference for α-2,8 linked disialylated ligands that are present in many gangliosides as their terminal portions. The crystal structure of the Siglec-7 V-set domain was the first to be solved among the CD33-related Siglecs and it has been widely studied in complex with different sialylated ligands. It has been demonstrated through mutagenesis studies, that the key residue through the recognition process is the conserved Arg124 (Figure 1.13), establishing always a stable salt-bridge with the carboxylate moieties of sialylated ligands. As mentioned before, Siglec-7 shows a preference for α-2,8 sialic acid linkages, and it provokes a significant conformational change of the CC' (Arg67-Trp78) upon the binding of the GT1b ganglioside, that contains this Neu5Ac linkage.[55] Moreover, recent studies have shown that Siglec-7 contains actually a second binding site, Yamawaka et al have proven comprising Arg67 as the anchorage to a binding event. Indeed, the conformational changes upon the CC-loop due to its flexibility seem to be the potential control of binding on these different pockets.[56]

Siglec-7 is mainly involved in the suppression of the NK cell activation, therefore preventing unwanted damage to normal healthy cells. Its function is key in the maintenance of the immune homeostasis and preventing abnormal autoimmune responses. Some recent studies have demonstrated the potential involvement in cancer and infectious diseases. The inhibitory signaling is exploited by sialylated tumorigenic and pathogen cells and through the recognition with Siglecs are able to escape and evade immune mechanisms. For this reason, it is an important target of interest for immunotherapy strategies. Even though it has a central role in cancer, Siglec-7 is also involved in other diseases and pathologies, such as HIV-1, obesity, hepatitis, as emerged over the last years. Interestingly, GQ1b-like epitopes, containing LOS of *Campylobacter jejuni,*[57] involving Guillain-Barré syndrome (GBS) can be recognized by Siglec-7, leading to the modulation of host-pathogen binding. It has been shown to interact

with capsular polysaccharides (CPS) from pathogens like *Neisseria meningitidis*, *E. coli* and Group B *Streptococcus*. Moreover, the presence of sialylated lipopolysaccharide on certain *Fusobacterium nucleatum* [58]strains, oncogenic pathogen in different human tissues, may also induce the activation of Siglec-7, causing immunosuppression that may promote its carbinogenic behavior.

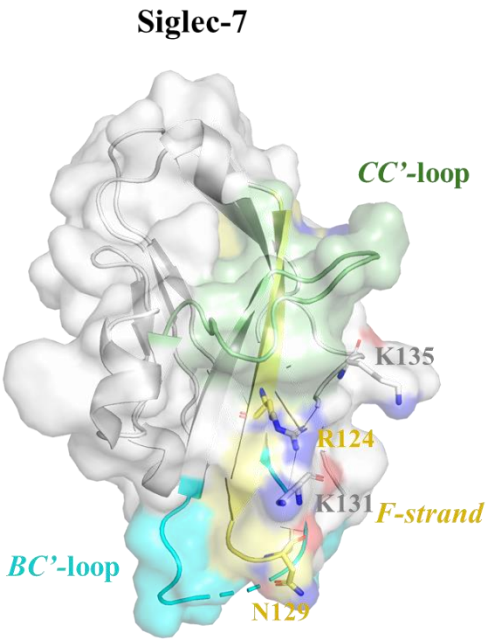

Figure 1.13 3D representation of Siglec-7 with protein surface colored by strand and loops, as indicated in the figure: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. Conserved residues on the binding site are highlighted with sticks and tagged with its residue number.

1.6.2 Siglec - 9

Siglec-9 is a member of the human CD33-related inhibitory Siglec receptors, acting as a negative regulator of the innate immune system. It is mainly expressed on neutrophils, monocytes, and subsets of T cells and B

cells. The extracellular domain of Siglec-9 is characterized by an N-terminal V-set domain, common to all Siglecs, and two C2 Ig domains. [59]Siglec-9 shows a prevalence for α-2,3 and α-2,6-linked sialylated glycans, which are commonly present in different cell surface glycoproteins and glycolipids.[60] The crystal structure of the Siglec-9 V-set domainreveals important insights into its ligand interactions. (Figure 1.14) Similar to other Siglecs, Siglec-9 recognizes sialylated ligands through a conserved arginine residue, Arg120, that forms a crucial salt bridge with the carboxylate group of sialic acid.[61] This interaction is essential for its ability to bind to sialylated structures on host cells and pathogens. Studies have shown that the binding of Siglec-9 to its ligands can induce conformational changes in its extracellular domain, particularly in the CC loop region, which play a role in fine-tuning its binding specificity and affinity. Siglec-9 shows not only preference for α-2,3 and α-2,6 ligands, but also for sulphated glycans, which act as an anchorage through the interaction with different lysines and arginines present on the Siglec-9's binding site.

Siglec-9 plays a critical role in modulating immune responses by inhibiting the activation of immune cells, thus preventing excessive inflammation and autoimmune reactions. It is particularly involved in the regulation of neutrophil lifespan and the resolution of inflammation. The inhibitory signaling mediated by Siglec-9 is exploited by sialylated tumor cells and pathogens to evade immune detection. Tumor cells often overexpress sialylated ligands that engage Siglec-9, leading to the suppression of anti-tumor immune responses and promoting tumor growth and metastasis. As a result, Siglec-9 has emerged as a potential target for cancer immunotherapy.[62]

In addition to its role in cancer, Siglec-9 is also involved in other diseases. For instance, its interaction with sialylated structures on certain bacteria and viruses like HBV, HIV-1 and even COVID-19 can modulate immune responses during infections when the interaction with Siglec-9 takes place.

Some researchers have also linked Siglec-9 to chronic inflammatory conditions and autoimmune diseases like rheumatoid arthritis, where dysregulation of its signaling pathways may contribute to disease pathogenesis.[63] Understanding the role of Siglec-9 in these contexts continues to be an area of active research, with implications for developing novel therapeutic strategies targeting this receptor.[64]

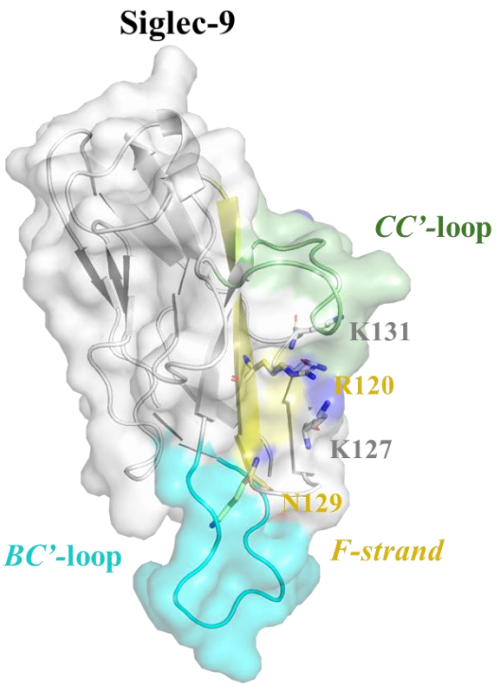

Figure 1.14 3D representation of Siglec-9 with protein surface colored by strand and loops, as indicated in the figure: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. Conserved residues on the binding site are highlighted with sticks and tagged with its residue number.

1.6.3 Siglec – 10

Siglec-10 possess one ITIM domain in its cytosolic tail that defines it as a inhibitory siglec from the CD-33 related Siglec family like Siglec-7 ad

Sigec-9, mentioned above.[65] Is mainly expressed on B cells surface, but can also be found on dendritic cells and on subsets of human leukocytes, like neutrophils and macrophages.[66] In a similar way as Siglec-9, Siglec-10 can recognize both α-2,6 and α-2,3 sialylated ligands. The crystal structure of this inhibitory Siglec has not been published yet, but we've performed homology modelling from the Siglec-10 published sequence to obtain a 3D model of human Siglec-10. [67] Structurally, it was observed that differently from other CD-33 related Siglecs, the CC-loop from Siglec-10 tends to point outside to the binding residues, elongating the binding site so different shapes and lengths of sialylated glycans can be accommodated. It has been demonstrated that Arg119 is the key residue to establish the salt-bridge in the recognition events. (Figure 1.15)

Siglec-10 is associated to several patho-physiological processes. It can bind CD24 cells.[68] In fact, high expression of Siglec-10/CD24 has been associated with poor prognosis in ovarian and breast cancers, suggesting that Siglec-10/CD24 axis is a key mechanism through which cells promote tumor evasion. Siglec-10 has also been associated with autoimmune diseases like Systemic Lupus Erythematosus (SLE); neurodegenerative diseases like Alzheimer's, where Siglec-10 is highly expressed on microglia and inflammatory processe like sepsis.[69,70]

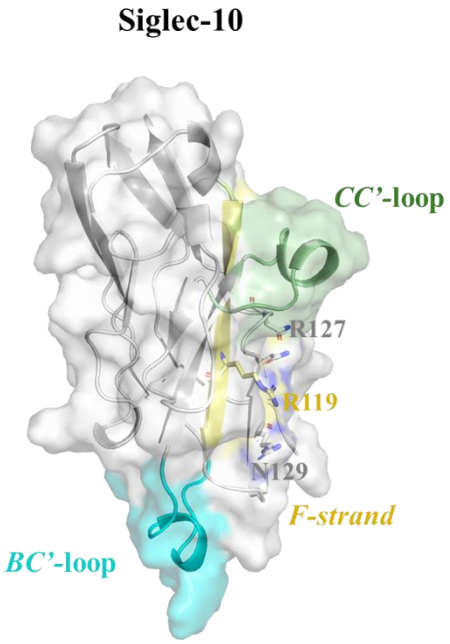

Figure 1.15. 3D representation of Siglec-10 with protein surface colored by strand and loops, as indicated in the figure: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. Conserved residues on the binding site are highlighted with sticks and tagged with its residue number.

Chapter II:

Deciphering the molecular intricacies on the protein –ligand interactions: different advanced techniques

II. Deciphering the molecular intricacies on the protein –ligand interactions: different advanced techniques.

As discussed previously, many of the roles that glycoconjugates display are crucial in health and disease and they are mediated by binding proteins that are able to decode the glycome's information content through the recognition with specific glycans. Therefore, the determination of the three-dimensional structure of protein-ligand complexes that are present in biological systems is an essential precondition for the research and development of new therapeutic strategies for the structure-based drug design. The plethora of non-covalent interactions that may occur in solution form the basis of the recognition events defining the protein-ligand interface and contribute to the complex formation. The low affinity and the high degree of complexity that is present in glycans makes the unravelling of the interactions a challenging task. Thus, employing a combination of a wide range of complementary techniques can be the most effective strategy to elucidate the interactions that take place.

There are many methods that are used to evaluate the protein-glycan interaction affinity and/or thermodynamics, including isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), enzyme-linked immunoassays (ELISA), fluorescence titrations. These methods, in combination with structural data, provide structure-activity relationships. The depiction of the structural biology techniques mainly involves X-ray crystallography, nuclear magnetic resonance (NMR) and the emerging cryogenic electron microscopy (cryo-EM), and allow the characterization of the three-dimensional structure and physicochemical properties of glycan-receptor complexes. The intricacies surrounding these structural

techniques come from the challenges on the preparation of appropriate systems to analyse, such as difficult crystallizations processes and problems related to obtain isotopically labelled proteins for NMR. So, complementary computational techniques are employed to partially overcome these limitations and gather molecular insights on the glycan-protein interactions. The present chapter is focused on the description of the principal techniques that have been used or are important for the thesis scopes, thus putting special focused on the implementation of molecular modelling together with ligand-based NMR techniques.[71]

2.1 Nuclear Magnetic Resonance NMR Spectroscopy

Nuclear Magnetic Resonance spectroscopy is an effective tool for the characterization of structure, conformation, geometry and dynamic of organic molecules and more specifically carbohydrate ligands. This dissertation will not get only into the NMR fundamentals, but most focused on the main methodologies that are currently used to dissect glycans structure and behavior in solution as well as upon receptor binding with different proteins.[72,73]

2.2 Ligand-Based NMR approaches

NMR spectroscopy is a widely used technique for elucidating the recognition processes between ligands and receptors[74], with the recent application on the dissection of protein-glycoconjugate interactions. The equilibrium between a small molecule (L) and a protein (P), leading to the formation of a complex (PL), follows a bimolecular association reaction with second-order kinetics. [75]

$$L + P \leftrightarrow PL \qquad [2.1]$$

As observed in the equation above, k_{on} and k_{off} represent the association and dissociation constants, respectively, whose ratio (k_{off}/k_{on}) determines the rate constant, K_D, defined as follows:

$$K_D = \frac{[L][P]}{[LP]} = \frac{k_{off}}{k_{on}} \qquad [2.2]$$

Depending on the K_D value, two cases may occur: slow or fast chemical exchange between the free and bound states. In the case of slow chemical exchange ($K_D \approx 10^{-5}$ M), the lifetime of the complex is longer than the time scale of the chemical shift difference between the free and bound states, resulting in two distinct NMR signals in the spectrum. On the other hand, in the fast exchange regime (K_D around 10^{-8} M), the exchange process is faster than the time scale of the chemical shift difference, causing the signals to merge into a single peak that represents the average chemical shift of the free and bound forms.

Certain NMR techniques depend on the chemical exchange regime of a protein-ligand complex, so the selection of the appropriate NMR experiment to decipher the protein-glycoconjugate interaction becomes crucial. Table 2.1 outlines the applicability of some NMR techniques that are used to describe these interactions and will be employed in the following paragraphs.

Table 2.1 Range of the applicability of the main NMR techniques that study protein-ligand interactions.

	K_D [M]	**Target MW [kDa]**	**Protein ligand ratio**	**Target labelled is required**	**Description of target binding site**	**Ligand epitope mapping**	**Ligand selectivity in a mixture**
Tr-NOE	10^{-6} -10^{-3}	No limit	1:5/1:50	No	No	No	Yes
STD NMR	10^{-6}-10^{-3}	>15	1:50/1:200	No	No	Yes	Yes
Water LOGSY	10^{-6}-10^{-3}	No limit	1:5/1:50	No	No	Yes	Yes
Diffusion experiments	10^{-6}-10^{-3}	No limit	1:1/!:20	No	No	Yes	Yes
CSP	10^{-9}-10^{-3}	<100	1:1/1:10	Yes	Yes	No	No

In the ligand-based NMR techniques commonly used for protein-glyconconjugate interactions, the first step is the ligand NMR assignment, and can be accomplished through a combination of mono and bidimensional NMR spectra. ^{1}H and ^{13}C NMR experiments are useful to define the monosaccharide composition, the linkage α or β configuration, the possible substitution pattern and the nature of its non-sugar substituents. Specifically, ^{1}H NMR experiments give valuable information about signal multiplicity, scalar interactions between vicinal and geminal nuclei, which are propagated through the electrons that are involved in the bond, measured by coupling constants (J). These characteristics are essential for obtaining detailed structural information about the glycan.

Table 2.2 Typical ^{1}H and ^{13}C chemical shifts of glycan molecules.

δ (ppm)	**^{1}H**
8.5 – 7.5	Amide resonance
5.5 – 4.2	Anomeric protons
4.5 – 2.8	Sugar ring protons
2.6 – 1.8	α- methylene protons of deoxy sugars
1.0 – 2.0	Methyl protons of 6-deoxy sugars and also acetyl groups.
δ (ppm)	**^{13}C**
160 – 180	Carbonyl carbons
95 – 105	Anomeric carbons
60 – 80	Sugar ring carbons
45 – 60	Nitrogen bearing carbon signals
~ 30	Aliphatic methylene carbons of deoxy sugars
20 – 17	Methyl carbon of deoxy sugars and acetyl groups

Additionally, the distinctive values of the coupling constants $^{1}J_{C1,\,H1}$ and $^{3}J_{H1,\,H2}$ can discriminate between the anomeric orientation of the sugar units. In fact, for pyranose rings with *gluco* or *galacto* configuration (H-2 Axial), a $^{3}J_{H1,\,H2}$ above 8 Hz is indicative of a β configuration at the anomeric carbon, and below 3 Hz we face an α- configured anomer. In contrast, for *manno* configuration (H2- equatorial) a $^{3}J_{H1,\,H3}$ below 3 Hz is

typical. The configuration at the anomeric position can also be analyzed through the $^{1}J_{C1,\,H1}$, giving typical couplings at 165 Hz for β and 170 Hz for α, respectively.

Homonuclear through-bond 2D experiments, such as COSY (Correlation Spetroscopy) and TOCSY (Total Correlation Spectroscopy) spectra, measure the correlation of the protons within the same monosaccharide unit and do not provide details about the monosaccharide connection. Nuclear Overhauser Spectroscopy (NOESY)[76] is a homonuclear through-space experiment that identifies spins undergoing cross-relaxation by measuring the rates of the cross-relaxation. This process allows for the transfer of magnetization to protons that are spacially close to one another. With respect to the heteronuclear 2D carbon-proton correlation, HSQC (Heteronuclear Single Quantum Correlation) relates directly the coupled ^{13}C and ^{1}H, HMBC (Heteronuclear Multiple Bond Correlation) uses multiple-bond couplings over two or three bonds (J=2-15 Hz) for the determination of the long-range ^{1}H-^{13}C connectivity. (Figure 2.1)

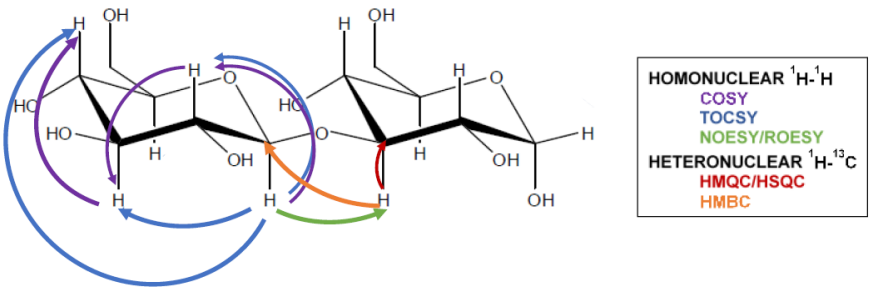

Figure 2.1 Representation of the correlations observable by different 2D NMR experiments.

Following ligand assignment, ligand-based NMR methods are then performedLigand-based NMR experiments require a fast chemical exchange transfer of the bound state to the free state. This regime is characterized by a range of $K_D > 100$ μM in the medium-low affinity range,

dissociation rate constants in the range of $1000 < k_{off} < 100\ 000\ s^{-1}$ and use of large ligand molar excesses.

2.3 Transferred – NOESY

NOE (Nuclear Overhauser Effect) effects are fundamental for determining the 3D structure of molecules in solution. The transferred NOE effect employs the different behaviors of a ligand in its free versus bound states, allowing the detection and characterization of ligand binding and its bioactive conformation. For low to medium molecular weight moieties (less than 2 kDa), such as disaccharides or trisaccharides, fast tumbling in solution leads to short correlation times. As a result, NOE intensities may be close to zero, making ROESY (Rotating-frame Overhauser Effect Spectroscopy) experiments, where NOEs are measured under spin-locked conditions, a suitable alternative to NOESY. NOE correlations provide valuable information about the sugar units present and the configuration of the anomeric center. For example, in β-configured sugars like Glc, Gal, or Man, the H-1 proton shows strong intra-residue NOE correlations with H-3 and H-5. In contrast, in α-configured sugars, H-1 primarily correlates with H-2.

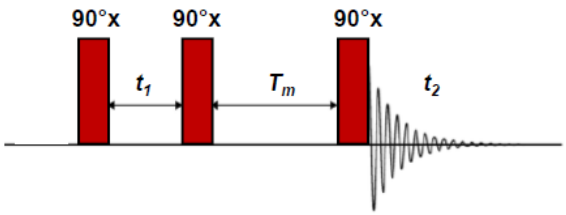

Figure 2.2 The NOESY experiment pulse sequence

Additionally, NOE spectra can reveal correlations between the anomeric proton and protons in adjacent sugar moieties. These *inter*-residual contacts are crucial for determining the sugar.

These effects are due to homonuclear dipolar interactions, which decrease rapidly with increasing inter-nuclear separation. Consequently, NOE interactions are detectable between nuclei that are within 5-6 Å of distance

of each other. The intensity and sign of NOE contacts depend on the molecular tumbling rates in solution, which are characterized by the parameter $\omega_0 t_c$, where t_c is the correlation time and ω_0 is the Larmor precession frequency. As showed in Figure 2.3, small molecules with rapid and random rotation in solution exhibit positive NOE contacts with maximum values around 0.5, whereas molecules with slower tumbling show negative NOE contacts with maximum values around -1.

.

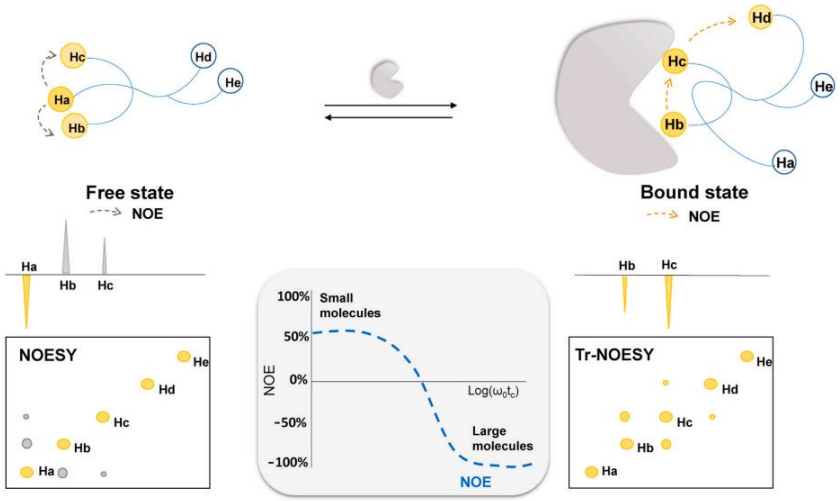

Figure 2.3. Schematic representation of NOE effects. In the free state, small molecules exhibit positive NOEs (cross-peaks with opposite sign to the diagonal peaks in the NOESY spectrum); in the bound state, small molecules adopt negative NOEs, behaving as the large protein, as shown in the tr-NOESY spectrum.

It is noteworthy that the small molecules binding to a receptor protein can exhibit transferred NOE (tr-NOE), behaving as part of the macromolecule. The ligand, freely and rapidly exchanges between the bound and free states, thereby retaining the NMR properties of the protein and providing information about the bound state. Discrimination of NOEs between the free and bound state can be also achieved by the analysis of the build-up

rate, which is the time required to achieve maximum intensity of NOE. The maximum NOE is a function of molecular tumbling rates, defined by ω_0 and t_c, ω_0 being the spectrometer observation frequency and t_c the rotational correlation time, connected directly to the molecular size. The maximum enhancement for tr-NOEs is observed at significantly shorter mixing times, t_{mix} in range of 50 to 100 ms) than for unbound ligands (4 to 10 times longer).

The analysis of NOE-derived inter-atomic protons distances allows us to detect conformational changes upon binding, enabling the determination of the bioactive conformation adopted by the ligand in the presence of the receptor while it gets in the bound state. The construction of the NOE build-up accurately determines ^{1}H-^{1}H nuclear distances. By integrating of the NOEs of spectra acquired at different mixing times, the build-up curves can be fitted to a double exponential function:

$$f = a\,(e^{-ct})(1 - e^{-bt}) \qquad [2.3]$$

Where f is the cross-peaks integral, a b and c are adjustable parameters and t is the mixing time. The initial slope is determined from the first derivative at time t=0.

$$f(0) = a\,x\,b \qquad [2.4]$$

From the initial slope, the inter-atomic proton distances are derived and calculated by the use of the isolated spin-pair approximation:

$$r_{ij} = r_{ref} \sqrt[6]{\frac{\sigma_{ref}}{\sigma_{ij}}} \qquad [2.5]$$

Where r_{ij} is the distance we want to calculate, r_{ref} is a known reference distance, σ_{ref} is the cross-relaxation rate and σ_{ij} is the cross-relaxation time that gives the desired distances. In conclusion, the tr-NOE allows fast screening of possible ligands respect to a specific target receptor and, at the same time, the knowledge of the recognized conformation of the ligand bound to the receptor, which has considerable implication for a rational structure-based drug design.[77]

2.4 Saturation Transfer Difference

Saturation transfer difference (STD) NMR spectroscopy is a powerful technique for detecting and characterizing epitope mapping on ligand when is complexed with a receptor in solution.[78] Based on magnetization transfer from the protein to the ligand, the protons by spin diffusion and intermolecular NOE, the STD allows to verify the interaction process and investigate it at a molecular level, providing the ligand epitope mapping. STD NMR is acquired as a pseudo- 2D experiment, resulting from the subtraction of two mono-dimensional NMR spectra.[79] 1) the off-resonance; in this step a region far from protein and ligand signals (usually at around 40 ppm) is irradiated to detect the reference spectrum, and 2) the on-resonance; here, the protein signals are saturated (typically in the aromatic or aliphatic spectral region) by applying a selective low power radio frequency pulse train in the range of seconds (this is called saturation time).[80] The saturation is transferred from the receptor to the interacting ligand during its residence time in the protein binding pocket due to intermolecular saturation transfer and a rapid chemical exchange. During the on-resonance, the enthalpic relaxation (R1) of the free ligand is slower than the k_{off}.

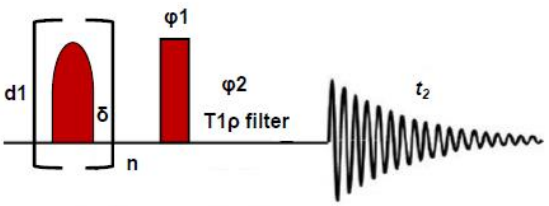

Figure 2.4 The 1H STD-NMR pulse sequence.

This leads to an accumulation of saturated free ligand in the bulk of the solution, causing a decrease in signal intensities of the protons that receive magnetization from the receptor. Since the STD is obtained by subtracting of the off and on-resonance spectra, ligand protons closer to the protein binding site will receive a higher degree of magnetization through intermolecular ^{1}H-^{1}H cross relaxation pathways, resulting in higher STD

signals. On the other hand, ligand protons that are found at a higher distance from the protein, will receive little to no saturation, giving as a result a lower or absence on STD signals. The STD intensities can be calculated as follows:

$$I_{STD} = \frac{I_0 - I_{sat}}{I_0} \qquad [2.6]$$

Where I_0 is the intensity of a signal in the off-resonance experiment, I_{sat} is the intensity of the signal in the on-resonance experiment. Depending on the degree of saturation of the protons, it is possible to map the interacting epitope of the ligand. Once set the most intense STD signals as 100%, all the other protons can be normalized to obtain the % of STD values.

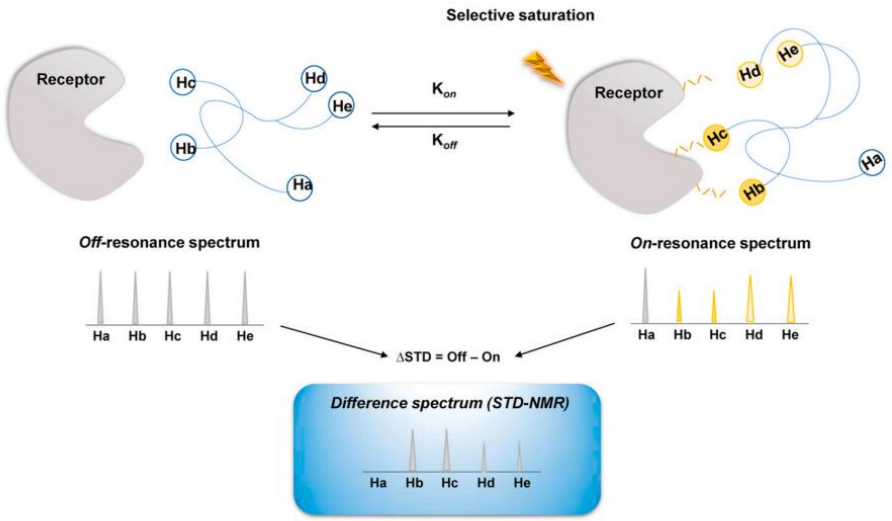

Figure 2.5. Schematic representation of STD NMR technique. In the off-resonance spectrum, acquired by irradiating far from both protein and ligand signals, ligand signals (Ha-Hb-Hc-Hd-He) do not show a decrease in their intensities. In the on-resonance spectrum, the receptor is selectively saturated with RF pulses and the magnetization is transferred to the ligandm protons by intermolecular NOE. Only the ligand signals involved in the binding process will show STD signals in the spectrum.

To overcome potential artifacts related with the relaxation times, STD NMR spectra can be recorded at different saturation times (tsat is usually from 0.5 to 5s) to generate the STD build-up curves. [81]

The initial growth rates are obtained by the derivation of the STD intensities close at the limit of zero saturation time, where no ligand-rebinding or relaxation takes place. STD-build up curves are calculated from the STD amplification factor (ASTD) following the equation

$$A_{STD} = \frac{I_0 - I_{sat}}{I_0} * \frac{[L]_0}{[P]_0} = \frac{I_{STD}}{I_0} * \frac{[L]_0}{[P]_0} \qquad [2.7]$$

Where I_{STD}/I_0 is the relative STD effect at total ligand ($[L]_0$) and protein $[P]_0$ concentrations. A_{STD} is calculated for each proton involved in the interaction at each saturation time. Data is fitted to the following mono-exponential function:

$$STD\ (t_{sat}) = STD_{max} * (1 - \exp(-k_{sat} * t_{sat})) \quad [2.8]$$

Where STD (t_{sat}) is the observed STD intensity, STD_{max} represents the asymptotic maximum of the build-up curve, t_{sat} is the saturation time and k_{sat} is the rate constante related to the relaxation properties of a given proton, measuring the speed of the STD build-up. The parameter STD_{fit} represents the slope of the curve when the time zero and depends on the proximity of the ligand to the protein. To obtain the ligand's epitope map , the STD fit values are normalized with respect to the highest one that is set to 100%.

Therefore, STD NMR is a useful technique to obtain information on the binding epitope of a ligand when is interacting with a macromolecule, which is particularly useful, for example, in the development of new therapeutic approaches, such as small drugs or mimetics (Figure 2.5). Notably, a sample containing a low concentration of macromolecule (in the μM range) is enough, but a large molar excess of the ligand is a requirement (tipiycally an excess from 1:10 ratio to 1:1000). As mentioned above, STD NMR technique is only applicable to protein-ligand systems in rapid exchange, with a medium to weak affinity to the receptor,

exhibiting dissociation constants generally in the millimolar to micromolar range (K_D 10^{-6} – 10^{-3} M).

2.5 Protein-Based NMR Approaches

In previous sections, we discussed ligand-based NMR methods that elucidate a ligand's structure both in its free state and when complexed with receptors (bound state). These methods also describe the ligand's bioactive conformation, and the epitope mapping of the protons involved in the recognition process. Protein-based NMR methods involve several techniques that allow the direct observation of receptor signals, which can characterize interactions with cognate ligands and help identify the receptor binding site, even when it is unknown.[82,83]

For these NMR experiments, prior assignment of the protein resonances in the absence of the ligand is required. This involves using a soluble, non-aggregated, isotope-labeled protein (e.g., ^{13}C, ^{15}N, ^{2}H). The protein must be expressed, purified, and characterized through specific 3D NMR experiments to assign the protein backbone and side chains. (Figuree 2.6)

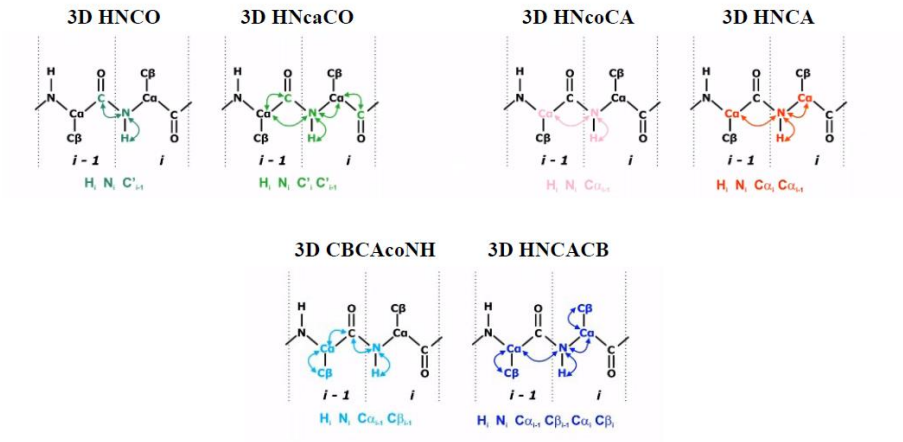

Figure 2.6. Representation of 3D NMR experiment used for the backbone resonances assignment. Each 3D experiment correlates HN resonances of each amino acid to the chemical shifts of CO, Cα and Cβ belonging to the same and preceding amino acid.

A limitation of protein-based techniques is that the macromolecule size should not be less than 50 kDa. Molecular weight affects relaxation times, which can result in low quality ^{15}N-HSQC spectra for large molecules. To address it, several methologies have been developed in the last decades. One approach is the minimization of signals overlap, which can be accomplished by the reduction of spin-spin relaxation, using the triple labelling protein expression (^{15}N, ^{13}C and ^{2}H), where deuteration considerable extends the T2 relaxation times, resulting in narrower lines in NMR spectra. Another solution is the TROSY (Transverse Relaxation Optimixed SpectroscopY) implementation of triple resonance experiments, which selects a single component of different relaxation T2 mechanisms (due to dipole-dipole interactions and chemical shift anisotropy) leading to a single, sharp peak in the spectrum.

Chemical Shift Perturbation

Chemical shift perturbation (CSP), also known as chemical shift mapping (CSM), is a common protein-based NMR experiment employed to map the interaction between a ^{15}N labelled protein when and a ligand.[84] This technique involves the monitorization of the changes on the chemical shifts of the protein's aminoacids when it is titrated with the ligand.[85] If the interaction occurs, the chemical shifts of the aminoacids involved in the complex formation with the ligand are subjected to perturbation. To perform CSP analysis of a protein-/ligand complex, the acquisition of the ^{15}N-HSQDC spectrum of the uncomplexed protein, or apoprotein as reference is required. Then, it is followed by sequential ^{15}N-HSQC spectra that are performed upon the addition of increasing amounts of ligand concentrations until the protein binding site is completely saturated.

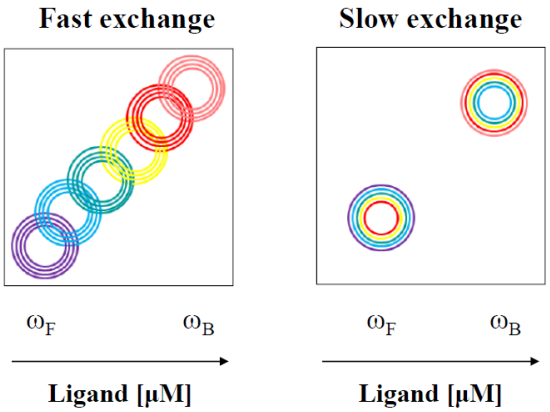

Figure 2.7. Chemical shift perturbation in protein-based NMR experiments. By adding amounts of ligand to the protein, the chemical shift can vary in the ^{15}N HSQC spectrum, depending on the protein-ligand exchange regime.

For the CSP analysis, it is essential is that both the protein and the ligand are dissolved in the same buffer, and that all the measurements during the titration are acquired under identical conditions. This is crucial because the chemical shifts, especially amide protons, are very sensitive to differences in temperature pH and buffer composition.

The chemical shift perturbation depends on the protein-ligand exchange regime: in the rapid exchange limit, the chemical shift represents the population-averaged value between the free and the bound forms of the protein, and the peak moves linearly by adding different amounts of ligand (Figure 2.7). On the contrary, in the slow exchange limit, a decrease in intensity of the peaks are affected by the interaction observed and in some cases and at large excess of ligand it can be possible to see the appearance of new cross peaks. When the ligand binds to multiple protein binding sites with different affinities, the chemical shift does not move linearly in the ^{15}N-HSQC spectra.

Although most peaks in a CSP experiment are affected by a variation of the chemical shift that helps in the determination of the binding site of the protein which is able to accommodate the ligand, conformational changes of the aminoacids can also lead to differences in the resonance frequencies.

Therefore, a shift in a signal is not always indicative of the vicinity to the binding interface. It can also give information about allosteric changes in the protein structure when a ligand is bound. Such conformational changes often happen to protein residues that are buried in the 3D structure or located distant from the binding pocket.

In section II, we described the advanced techniques that employ nuclear magnetic resonance (NMR) to unravel the intricacies between ligand-protein interactions. These techniques include elucidating the structure of the ligand, mapping the protons involved in the recognition process and depicting the bioactive conformation through ligand-based NMR methods. Additionally, to acknowledge the aminoacids involved in the interaction, the description of the binding pocket and conformational changes that may occur with the protein, can be revealed through protein-NMR based techniques. The following paragrpahs will explore complementary techniques that are used to confirm these experimental data and to add new interesting information about the recognition processes.

2.6 Computational Methods

During the last decade, computational methods applied in chemistry have gained success due to the several hardware and software improvements and the creation of high-performance computing and cloud computing systems. Computational chemistry has become one of the principal techniques in drug design, allowing for the rapid identification of promising drug candidates and the design of novel materials with tailored properties. Through the solving of several calculations, it prooved to be also key in the elucidation of structural and conformational features of complex and simple carbohydrate structures. Furthermore, the combination of molecular modelling with experimental techniques such as NMR spectroscopy methods have demonstrated to be extremely valuable in the determination of the conformational and dynamic properties of the free and bound state glycan structures. [86]

Typical techniques on computational chemistry involve:

A) *Ab initio*, Quantum Mechanics (QM) (2.6.1)

B) Molecular Mechanics (MM) (2.6.2)

C) Molecular Dynamics (MD) (2.6.3)

This thesis dissertation will be focused on the application of molecular mechanics and molecular dynamics to compliment with experimental data in the study of protein-glycan recognition processes. The following techniques will be described in this section.

2.6.1 Quantum Mechanics and Molecular Mechanics QM/MM

From the latin *ab initio*: "from the start". These are methods that are based on the resolution of the Schrodinger equation.[87,88] A base on the extension of the molecular orbital theory where electron-electron repulsions are not specifically taken into account is employed; but electron's average effect is used on the calculation. It gives as a result energy and a wavefunction, which is a mathematical function that can be used to calculate the electron distribution. This electron distribution gives useful information acknowledging the places on a molecule which are likely to be attacked by nucleophiles or electrophiles. However, solving the Schrodinger equation is noy an easy task, and can't solve molecules with more than one electron.

$$\hat{\mathrm{H}}\Psi = E\Psi$$

Where $\hat{\mathrm{H}}$ is the Hamiltonian (specifies its total energy—i.e., the sum of its kinetic energy (that of motion) and its potential energy (that of position), Ψ is the wave function and E the energy of the system.

Obviously, approximations were derived from the main equation to calculate the energy of the molecule, like Born-Oppenheimer or Hückel, but of course, it increases the computer power and time. Thus, *ab initio* calculations are relatively slow, and they are mainly used for the development of the commonly known force fields[89], that will be

afterwards applied in different calculations or simulations. [90]Forcefields contain information about the molecular geometries, energies, vibrational frequencies, spectra, ionization energies and electron affinities, and properties like dipole moments which are connected with electron distribution.

The different approaches of ab initio methods is divided into: Hartree Fock (HF), Post-Hartree Fock (PHF) and Quantum Monte Carlo.

2.6.2 Molecular Docking

From the application of the Schrodinger equation,by the employment of classical mechanics (harmonic oscillator) in approximations like the Born-Oppenheimer one, Molecular Mechanics was born. The combination of the calculations previously taken from Quantum Mechanics[91],and applied to classical mechanics has given the development of the force fields (Figure 2.8).They contain the energies of bonded (stretching, bending, torsions) and non-bonded contributions of the molecules and it is widely used for Molecular Dynamic simulations, and it is part of the scoring of the here explained Molecular Docking.

Molecular Mechanics is a useful part of the computational chemistry where the minimum of the potential energies of certain compounds are calculated. It is used to optimize known structures, to minimize systems before running molecular dynamics, to search conformational analysis in glycans or peptides and even to predict protein folding.

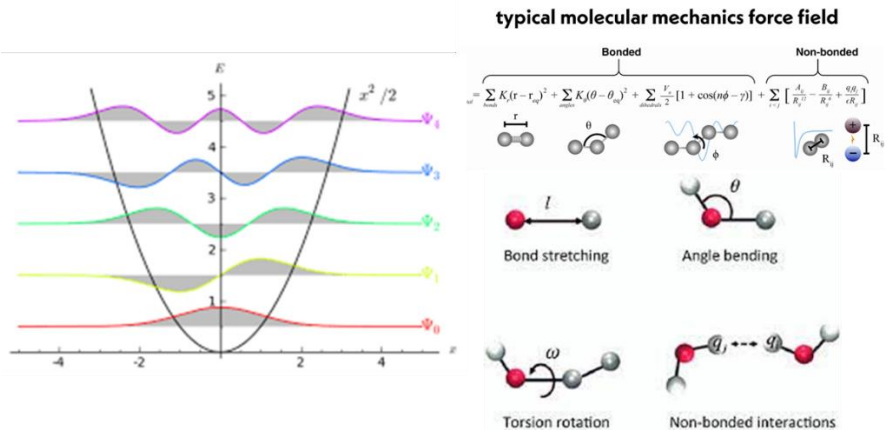

Figure 2.8 Classic mechanics where the fundamentals of Molecular Mechanics is developed: from the harmonic potential, the bonded and non-bonded molecule contributions are calculated.

Conformational search of glycans

In the context of carbohydrates molecules as ligands interacting with certain receptors, MM calculations are applied to get insights into the conformational behavior in solution of glycan molecules, alone and in complex with the biological target. The conformational analysis is crucial to determine the three-dimensional features, especially in the carbohydrate field as it allows to correlate a molecule with its real biological behavior.[92]

The conformational search aims to find the minimum local energy for a given molecule by rotating single bonds causing the rearrangement of atoms in space. The relative energies are then evaluated to find the lower energy conformer. The resulting structure can be afterwards compared to experimental data such the provided through NMR. For example, the NOE derived proton distances, to obtain a reliable 3D topology of glycans of interest. Indeed, the combination of modelling/NMR protocols is extremely useful to deduce conformational and dynamic properties of free and bound carbohydrate molecules and thus obtain the 3D complex of the system under study.[93]

The 3D structure of carbohydrates is characterized by the sugar composition and the glycosidic linkage, parameters that allow to describe shape. The ring shapes can be defined in terms of reference conformations: chair (C), twist (T), boat (B), envelope (F) and skew (S). Glycans are not only complex and diverse in their composition and substitution patterns, but the variety in linkage types and the spatial distribution of their monosaccharide units generate numerous conformers. These conformers can rearrange in three-dimensional space, leading to a wide array of structural possibilities that significantly influence their biological functions. The glycan conformations are mainly described by the glycosidic torsion angles called φ (H1-C1-Ox'-Cx'), ψ (C1-Ox'-Cx'-Hx') and for some linkages a higher degree of freedom, like 1-6 type, there is also the ω (Cx'-Ox'-O5'-C5') torsion, which is considered as the one providing the highest flexibility on the molecule, as depicted in Figure 2.9. Sampling the values of the w torsion has been described by means of the populations of gauche-gauche (*gg*), gauche-trans (*gt*) and trans-gauche (*tg*) rotamers (Figure 2.10).

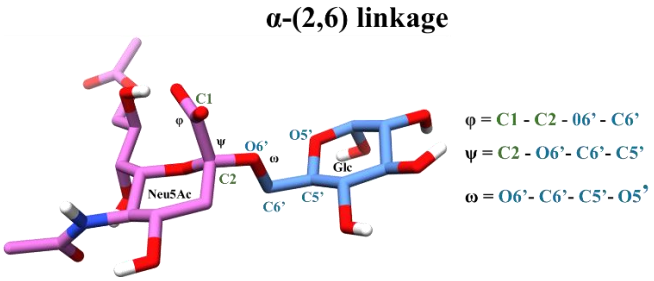

Figure 2.9. Representation of a sugar torsion angles

This additional flexibility of α-1,6, α-2,6 in case of sialylated glycans, makes it a challenging task to determine the preferential conformation in solution of the oligosaccharides containing theses type of linkages. The three cases of staggered rotamers are represented here below as Newman projections.

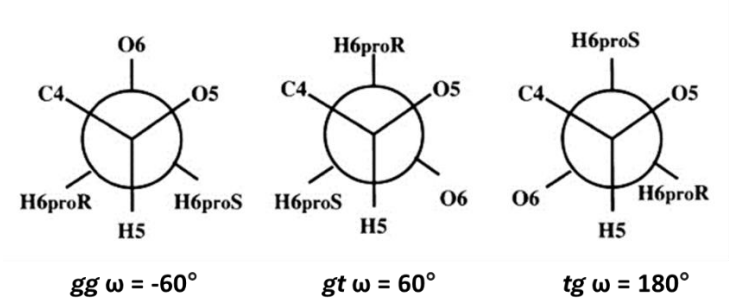

Figure 2.10 Newman projections of staggered rotameric states of the omega torsion anglealong the C6-C5 bond: gg (gauche-gauche), gt (gauche-trans) and tg (trans-gauche).

Several methods exist on the calculation of the principal torsion angles. These calculations define the energy profiles of the linkages, and it is generally applied grouping by disaccharides containing in the whole glycan chain. To run it, it is applied first a molecular mechanics calculation, MM3* force field,[94] to calculate local minima and flexibility of glycosidic torsions. Therefore, the potential energy surface shows the conformational energy for each torsion and is plotted on adiabatic energy maps, representing energy graphs similar to the Ramachandran plots used for the peptidic linkages, that show the energetically favorable conformations of a carbohydrate dimer.

Molecular Docking for protein-glycan interactions

Molecular Docking leverages the principles of Molecular Mechanics to predict how a small molecule like a glycan or a drug candidate interact with larger biological molecules, typically a protein.[95] In the docking, the ligand's ability to fit into the receptor's binding site is evaluated by generating different possible orientations and conformations (known as poses) of the ligand within the binding site. The success of this prediction largely depends on the scoring functions to assess these poses. The scoring functions often incorporates force fields that have been derived from Molecular Mechanics, through the evaluation of the bonded and

non-bonded interaction, as explained above, between the ligand an the receptor, Molecular Docking can estimate the binding affinity and stability of the complex. Essentially the force fields calculate the interaction energies ensuring that the process accurately reflects the molecular behavior as described by classical mechanics (Figure 2.11). The main diversity between the Docking programs relies on the algorithm used for the computational search and the nature of the scoring function applied to rank the docked conformations. [96]

Moreover, if the ligands are glycans, the main diversity relies on the torsions of the linkages, which entails different numbers of rotatable bonds on this ligand, and therefore its accuracy and computational cost. There are three types of algorithms for this conformational ligand search: shape matching, systematic search and stochastic algorithms.

The shape matching considers the geometrical overlapping between two molecules identifying the possible binding sites of a protein by a macromolecular surface search. The systematic search is often used for flexible-ñigand docking, and all the possible binding conformations are generated by the exploration of all the degrees of freedom of the ligand. The most time efficient are the stochastic algorithms, such as Monte Carlo (MC) method and evolutionary programming (EP), where random changes on the ligand bonded and non-bonded contributions are executed. Regarding the energy evaluation, three scoring functions can be classified: force-field based, knowledge-based and empirical-based scoring functions. As previously explained, and more used, the force fields employ the non-bonded terms of the classical mechanism to compute direct interaction energies, considering van der Waals and electro-static energies as well as the stretching, bending and torsional energies. The knowledge-based are based on a statistical analysis of the interacting contacts from protein-ligand complexes identified from structure databases. And last, the empirical scoring functions make use of several intermolecular interaction terms, such as hydrogen bond and

hydrophobic interactions for pose and affinity prediction so that fitted theoretical values are as close as possible to the experimental data.

To perform a docking calculation, the 3D structure of the protein is required, and it can be provided through experimental techniques such as X-ray crystallography or NMR spectroscopy. It can also be derived from homology modelling methods or prediction such as AlphaFold. It is noteworthy that the more we know about the binding site of our receptor, the more accurate the calculation will be. In AutoDock4, which is the software used for the Docking calculations in this thesis, the information on the binding site helps to build the AutoGrid grid box, which is the grid volume where the ligand has been providen free rotation.[97] If the binding pocket is unknown, a bigger grid volume is required so the grid can cover the protein surface and ligand rotates freely around it. Of course, this implicates a bigger computational cost and less accuracy. In this cases, preliminary docking experiments can be performed to investigate if some regions of the receptor are preferred by the ligand, in a process that is known as "blind docking". [98]In this way, a second round of docking calculations allows to choose a smaller grid where the ligand can move around the protein.[99]

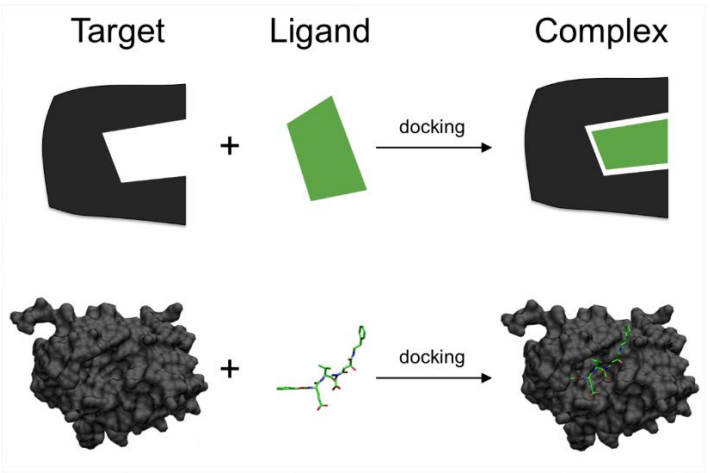

Figure 2.11 Schematic illustration of docking a small molecule ligand to a protein target.

AutoDock[100] uses evolutionary programming (EP) algorithm, where computational models are treated as evolutionary biological processes. [101]In particular, Autodock uses the Lamarckian Genetic Algorithm, an EP based on the combination of genetic algorithm (GA) and local search (LS) method that tries to find the closest conformation of the global energy minimum.[102],[103] In this hybrid method, the ligand variables (translation, orientation and conformation) are considered as "gene" that characterizes a "genotype" (ligand's state), while the atomic coordinates correspond to "phenotype". Genotypic space is characterized by mutation and crossover, whereas phenotypic space is established by the energy function to be optimized. Individual conformations search their local minima, inheriting the information to later generations (Lamarckian aspect). As force-field, this software currently employs semi-empirical free energy to score the docked binding conformations of small molecules to macromolecular targets, creating a set of docked conformations with the sum of their intermolecular and internal energy components. The pose with the lowest docked energy could be considered as "best" docking result.

2.6.3 Molecular Dynamic Simulation

Molecular Dynamic (MD) simulations often naturally follow Molecular Docking, to refine and validate of the previously predicted binding poses. While Molecular Docking only provides static snapshots of the different conformations on the ligand-receptor interactions, the MD simulations allow the exploration of the dynamic behavior of the formed complex over time (Figure 2.12). [104] This is done by the solving of Newton's equations of motion for the atoms involved; given the atom "I", with mas "mi", and cartesian position "xi", Fxi represents the force acting on the atom during the deriving time t:

$$\frac{d^2x_i}{dt^2} = \frac{F_{xi}}{m_i}$$

MD simulation is a widely spread technique since it allows and enables the whole complex (receptor + ligand) to explore the potential landscape, which in this case is defined by both classical and non-classical interactions within the molecular system. The energy profile is characterized by the possession of multiple minima, each corresponding to different conformational states of the system. As the simulation progresses in a time-dependant manner, this system can move from one minimum to other, exploring in this way several conformations that might not be accessible through Molecular Docking calculations. It is based on the hand-glove model, where both ligand and protein suffer conformational changes when the interaction takes place. The classical MD contains several s teps. The first step is an energy minimization in which, from a higher initial energy state, the search for the first minimum in the energy profile of a system is performed. The most common method applied in MD is the steepest gradient, by executing the geometry optimization until it reaches the local minimum. This method depends on the starting geometry and is suitable for a quicker optimization. During this step, the solvent environment is key. There are three different methods that can be applied to treat it. In this in vacu simulation ,a distance-dependent solvent is used, and they are firstly divided into explicit or implicit solvents. If it is considered implicit, two models can be applied: Generalized Born (GB) and Poisson-Boltzmann (PB). If it is explicit instead, the simulation takes place into a box of a defined shape and dimension where the solvent molecules surroung the simulated complex. In this case, periodic boundary conditions (PBC) may be applied to avoid surface artifacts. Different water models have been developed, but the most common and that will also be used during this thesis is the TIP3P model.

After the first step on solvation and minimization of the complex is completed, the following step consists in heating up the system, aiming to remove the unforable contqacts, such as steric clashes and donor/donor

or acceptor/acceptor bonds, between solvent and solute. Thus enhancing the atoms spped through a defined scheme of temperature increase. The resulting velocities are calculated by means of standard temperature-dependant Maxwell-Boltzmann distributions. A third step consists in the equilibration of the system, by controlling the energy through the temperature, pressure and volume conditions. Last, but not least, we face the production step, which actually starts and produces the trajectories of the simulation, that will be time-dependent.

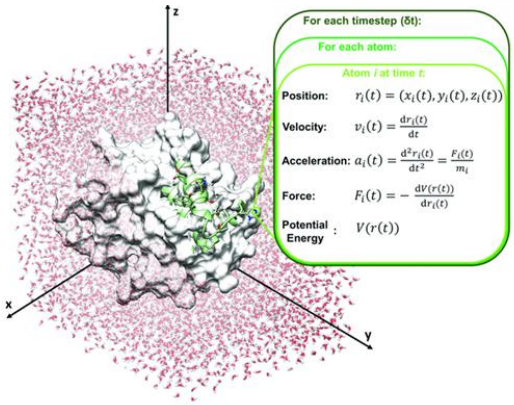

Figure 2.12 Graphical representation of the Newton's equation used for the production of timesteps on Molecular Dynamic Simumlation. Centered in the coordinate axis (x, y, and z) is found the protein complexed with a ligand.

As described, the simulation provides the energy profile of the system, and it is is given by the sum of the energies provided by the already known explained in this dissertation, *force fields*, that can be obtained through *ab initio* calculations or semi -empirical quantum mechanics (QM). The choice of the force field is, again, the key point on the validation of any Molecular Dynamic simulation.[105] So, prior to the actual performance of the MD, the coice of the correct force field is the crucial step. This election relies on the nature of the molecules to simulate as highly accurate force fields for specific systems have already been developed. AMBER[106,107] package, our chosen softwares, contains

force field[108] optimized for lipids (to date Lipid21[109]), for sugars (GLYCAM06)[110], for proteins, nucleic acids and water molecules (ff14SB)[111],[112] and for organic molecules (GAFF and GAFF2).

Focusing on carbohydrates structures, there are known two major force fields: from Amber/GLYCAM as mentioned above, and from CHARMM-GUI.[113] Both of these cover the majority of common monosaccharides and contain different glycosidic linkages. GLYCAM06 is suitable for D and L enantiomers, mono and oligosaccharides and for all glydosidic linkages, including the N-glycosilation.[114] It is the only force field with the same atom type (Cg at the anomeric carbon (C1) is assigned in both anomers, α and β, facilitating the simulation on ring-flipping and having equilibrium between conformers with axial and equatorial substituents at the anomeric center.

Chapter III:

Results and Discussion

III. Results and Discussion

3. Depiction of the recognition pattern of Siglec – 7 and Exogenous Ligands

The recognition pattern of Siglec-7 and its interactions with exogenous ligands are fundamental to understanding immune regulation and cellular interactions. Siglec-7, a member of the sialic acid-binding immunoglobulin-like lectins (Siglecs), is crucial in modulating immune responses and maintaining self-tolerance. Its binding to sialylated exogenous ligands can influence a range of immune processes, from controlling inflammation to regulating immune cell activation. Exploring these interactions provides deep insights into immune system function. The next sections will discuss different binding events between Siglec-7 and exogenour glycans from pathogens, aiming to enhance our comprehension of immune evasion mechanisms employed by various pathogens. This investigation not only advances our fundamental knowledge of immune regulation but also offers potential tools and pathways for developing targeted therapeutic strategies against diseases where immune modulation is key.

3.1 Molecular Recognition of Siglec – 7 and Serogroup Y CPS from *Neisseria meningitidis*

3.1.1 Introduction

As mentioned above, Siglecs represent an important family of immunomodulatory receptors primarily present on immune cells. The human genome encodes multiple Siglec members, possessing different ligand-binding specificties and intracellular signaling motifs; for this reason, there are Siglecs that have inhibitory or activatory immune responses. They are known for their ability to bind sialic acid residues that can be found on glycoproteins and glycolipids on the cell surface. In the introduction chapter, in the section 1.3, sialic acids have been introduced as a group of nine-carbon sugars, that decorate the ends of glycan chains on cell surfaces and secreted proteins, playing pivotal

roles in cellular functions like intercellular communication, signalling, and immune recognition.

In fact, this can be exploited by both tumour and pathogens including bacteria and viruses, with the aim to evade immune detection and clearance. Indeed, these pathogens can mimic host cell glycans by expressing sialic acid containing molecules on their surfaces, thereby interacting with Siglecs and evading immune surveillance: this evasion mechanism is crucial since it leads to the infectious diseases spreading as well contribute to cancer development.

Siglec-7, is a member of human C33-related inhibitory Siglecs, so it works as a negative regulator of the innate immune system. It is mainly expressed on Natural Killer (NK) cells[115] and structurally consists of an extracellular N-terminal V-set sialic acid binding domain which is followed by two C2-type Ig-like domains, that outdistance the carbohydrate binding region from the cellular surface.[116] It contains an immunoreceptor tyrosine-based inhibiton motif (ITIM) in its cytosolic region (Figure 3.1). This receptor is considered a novel glyco-immune checkpoint and occurs in the context of several diseases including cancer, infection and autoimmune disorders.

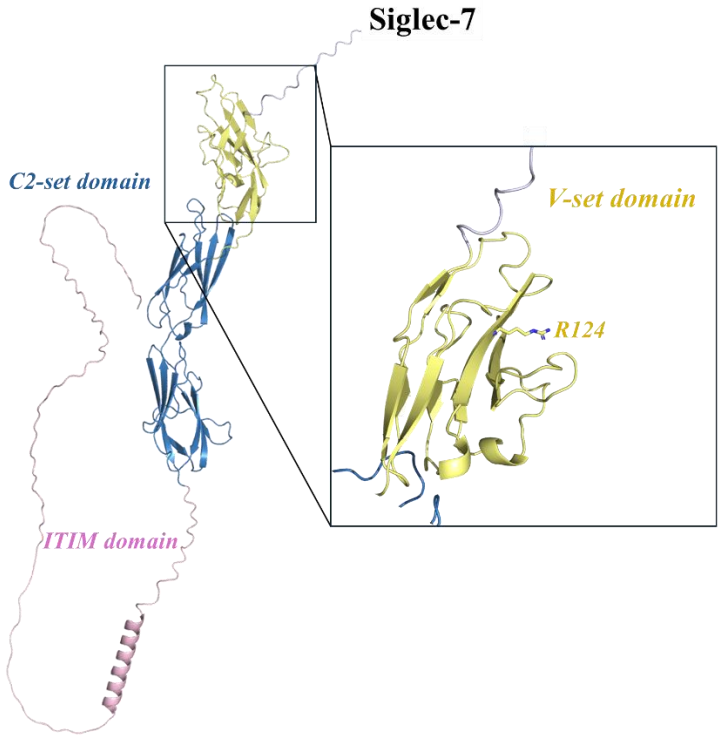

Figure 3.1. 3D representation of the full length of Siglec-7. V-sest domain is represented as yellow cartoon; C2-set domain in blue cartoon; ITIM domain in pink cartoon.

Neisseria meningitidis (Nm) is a Gram-negative human pathogen and etiological agent of the meningococcal meningitis, which can lead to potentially deadly infections, affecting membranes surrounding the brain and spinal cord in human host and can produce large epidemics.[117] The capsular polysaccharides (CPS) surrounding *Nm* cell envelope represent their major virulence factors, playing a crucial role in *N. meningitidis* pathogenicity, making them the main component of meningococcal vaccines as CPS or as protein-conjugate vaccines. [118, 119]

Nm holds thirteen serogroups, categorized on the basis of their CPSs into three pairs: Men-A and Men-X, whose CPSs are composed of phosphodiester polymers of aminosugars; Men-B and Men-B, whose CPSs consist of α-2,8 and α-2,9 sialic acid homopolymers, respectively; Men-Y and Men-W, containing

CPSs build-up of sialic acid (Neu5Ac) containing with 4)-α-Neu5Ac-(2→6)-α-Glc-(1→]$_n$ and [4)-α-Neu5Ac-(2→6)-α-Gal-(1→]$_n$ repeating units respectively.[120]

Table 1. Virulence factors as serogroups divided by repeating unit from *Neisseria meningitidis*.

Serogroup	Structure of the repeating unit
A	→6)-α–D-ManNAc(3OAc)(1-PO_4→
B	→8)-α–D-Neu5Ac(2→
C	→9)-α–D-Neu5Ac(7/8OAc)(2→
H	→4)-α–D-Gal(1→2)-glycerol-3-PO_4→
L	→3)-α–D-GlcNAc(1→3)-β-D-GlcNAc(1→3)-α-D-GlcNAc-1-PO_4→
X	→4)-α–D-GlcNAc-1-PO_4→
Y	→4)-α–D-Neu5Ac(7/9OAc)(2→6)-α-D-Glc-(1→
W135	→4)-α–D-Neu5Ac(7/9OAc)(2→6)-α-D-Gal-(1→
Z	→4)-α–D-GalNAc(1→2)-glycerol-3-PO_4→
29E	→4)-α–KDO(1→3)-α-D-GalNAc(1→

Recently, the USA Centers for Disease Control and Prevention issued a health advisory to alert for an increase in invasive meningococcal cases, mainly due to MenY. In addition, the neuronal damage happening during bacterial meningitis, could be correlated to the appearance of neurodegenerative diseases and the study of the molecular interactions and pathways linking infections with neuronal damage remain mostly unexplored.[121]

Given the significance of sialic acids in immune evasion, as previously discussed, it is possible that *Neisseria meningitidis*, particularly sialylated serogrops, may exploit similar mechanisms evading immune detection by the interaction with Siglec-7. The presence of sialic acid in serogroups like Men-B, Men-C, Men-Y and Men-W, suggests that this pathogen may engage Siglec receptors on immune cells.

This chapter will focus on several *Neisseria* serogroups having this binding potential to Siglec-7 through the interaction between their sialic acid-rich CPS. We aim to deepen on novel aspects of *Neisseria meningitidis* pathogenesis by means of a detailed analysis on the structural features of Nm CPS and binding to Siglecs, to provide a comprehensive understanding on the role of both receptor and pathogen, with potential implications for both vaccine development and therapeutic strategies.

3.1.2 Biophysical techniques for the depiction of Siglec-7 and sialylated ligands

Deciphering the interactions between Siglecs and sialylated glycans from pathogens at the atomic level is crucial for understanding immune recognition mechanisms. A multitechnique approach involving experimental and computational methods provides a comprehensive view of these molecular interactions.

By combination of NMR, computational and biophysical approaches, the evaluation of the factors that govern the selective binding of Siglec-7 to serogroup Y (MenY) of *Neisseria meningitidis* were studied in detail, unravelling the basis of this interaction. (Figure 3.2)

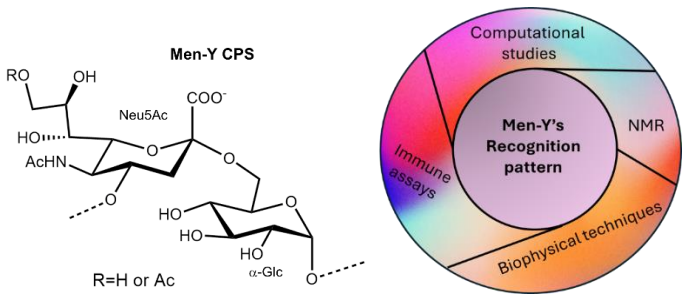

Figure 3.2. Repeating unit of the serogroup Y CPS from *Neisseria meningitidis*. Schematic representation on the techniques employed in this chapter.

3.1.3 Fluorescence titrations, ELISA immunological assays for the binding Serogroup Y and Siglec-7.

Solid-phase enzyme immunoassay (ELISA) between serogroups from *Neisseria meningitidis* and Siglecs. Different recombinant soluble forms of Siglecs (Siglec-5, Siglec-7 and Siglec-9) and human C-type lectins (DC-SIGN, Langerin, MGL and MR) were used with the full extracellular region fused to the Fc region of the human Ig1. As showed in the Figure 3.3, the ability of Siglec-7 to recognize Men-Y and no other serogroups was demonstrated, indicating that Siglec-7 possesses different affinities towards two structurally similar sialylated CPS, Y and W.

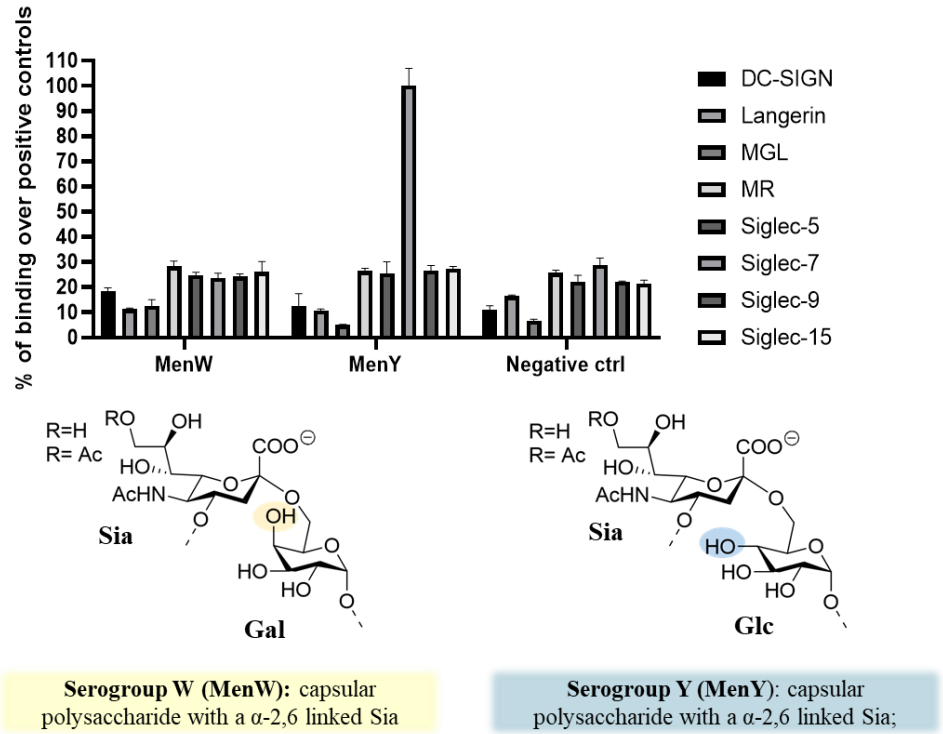

Figure 3.3 ELISA solid-phase assay performed between different C-type lectins and Siglecs-Fc. Men-Y CPS showed a strong and selective binding to Siglec-7 compared to Men-W. Polyacrylamide polymers (PAA) coated with different glycans were used as positive controls. Wells were coated with the reported CPS and the binding to different C-type lectins and Siglecs was evaluated. The experiment has been performed three times in duplicate with similar results.

Data are normalized over binding signal from the positive controls used for each lectin (set as 100% of binding). Error bars indicate standard deviations. OD: Optical density. Below the schematic representation of the MenY and MenW serogroups.

After the ELISA, subsequent steady-state fluorescence analysis was performed to determine the binding affinity between MenY and Siglec-7. Addition of increasing concentrations of Men-Y CPS allowed to build the curve, recording the emission between 310 and 450 nm, whilst through non-linear regression the binding constant K_b was obtained.[122] We obtained a K_D of 62 in the micromolar range (Figure 3.4). This result confirms the ability of Siglec-7 to interact with MenY from *Neisseria meningitidis* with relatively high affinity.[123]

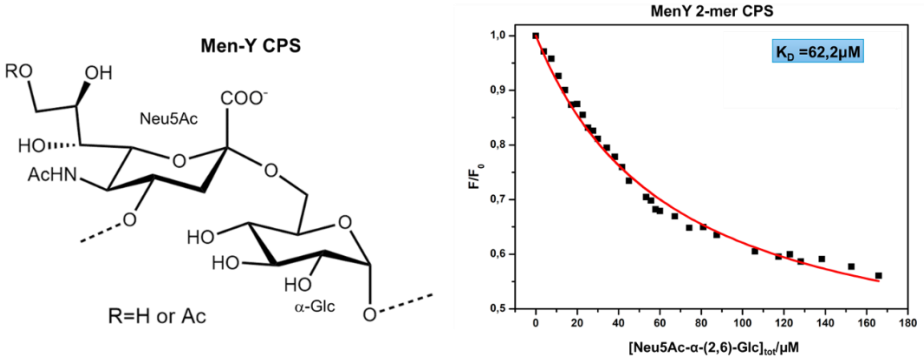

Figure 3.4. Fluorescence titration of Siglec-7 upon the addition of increasing concentrations of Men-Y. The emission spectra were recorded by using an excitation wavelength of 295 nm and a temperature of 25ºC.

3.1.4 Depolymerization and purification of Serogroup Y from *Neisseria meningitids* CPS

Dr. Fabrizio Chiodo provided us the capsular polysaccharide (CPS) serogroup Y, MenY, from *Neisseria meningitidis*. In order to successfully determine its recognition pattern by Siglec-7, we first depolymerized the oligosaccharide,

that was around 35-50 kDa, to perform ligand-based NMR (Figure 3.5).

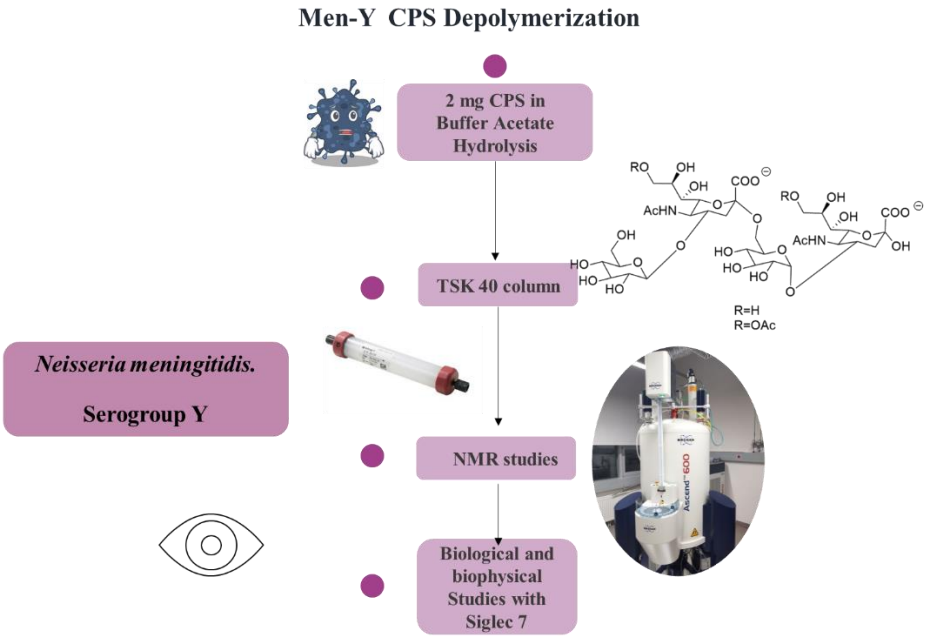

Figure 3.5 Schematic representation on the workflow on the experimental procedures for MenY depolymerization and purification.

Depolymerization was carried out through a mild acid hydrolysis in acetate buffer, pH=4.3, followed by a purification step on gel filtration chromatography, TSK40, as indicated in the Figure 3.6 below.

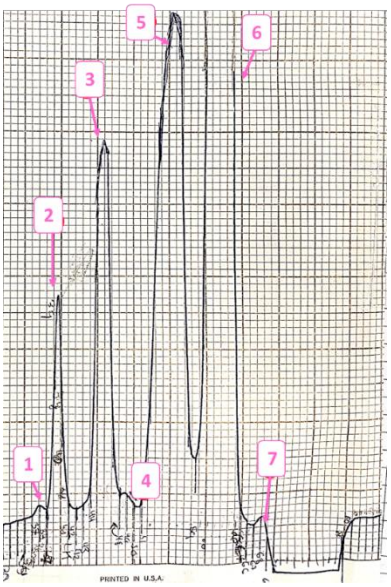

Figure 3.6. TSK40 gel filtration column peaks obtained from the purification of the depolymerized portions of MenY CPS. They represent from left two right lower to higher molecular weight. Peaks 6 and 7 represent the salts purified from the mixture.

The fractions were analyzed through NMR (Figure 3.7), achieving successfully oligomers from 1 to 3 repeating units. (1-Mer to 3-Mer). They were afterwards used for binding studies.

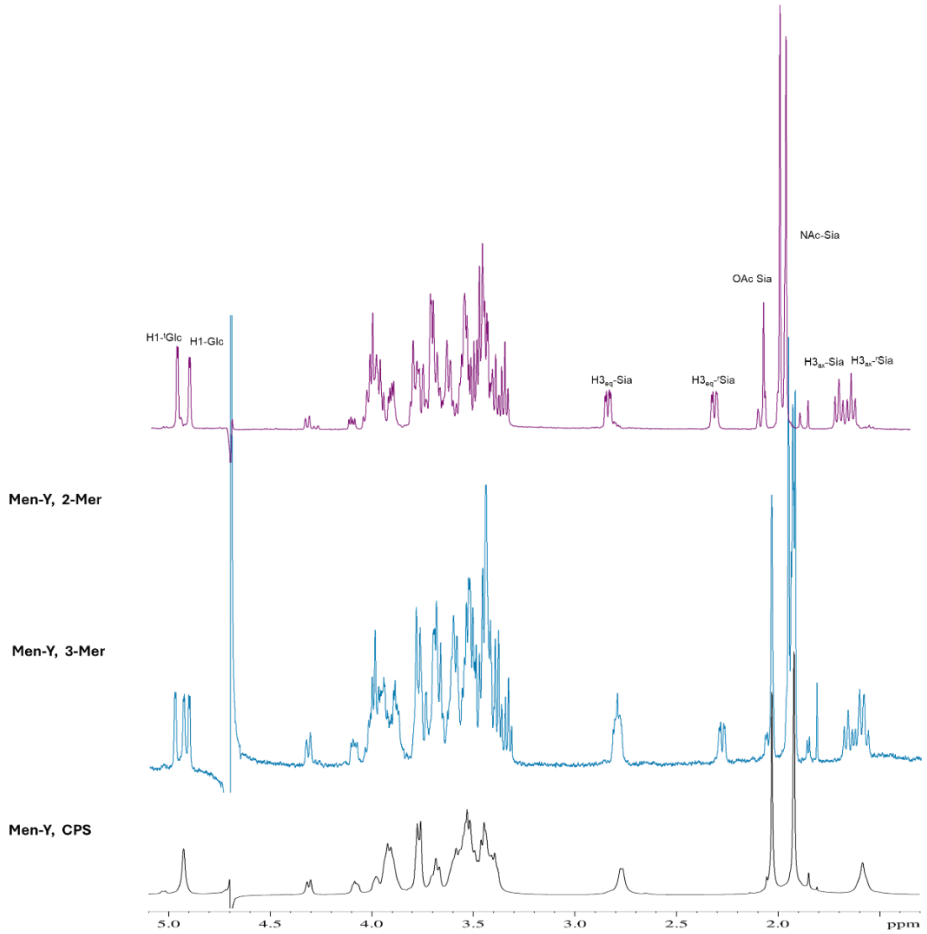

Figure 3.7 ^{1}H-NMR before and after depolymerization of Men-Y CPS from *Neisseria meningitidis*, obtaining different lengths of oligomer: 2-Mer and 3-Mer.

3.1.5 STD NMR and tr-NOESY analysis of the molecular binding between Siglec-7 and Serogroup Y – NMR

Once proved that Siglec-7 can bind MenY, we used NMR to map the recognised epitope with STD NMR and the achieved bioactive conformation thanks to trNOESY experiments.[124, 125] We initially studied the conformational behavior of MenY oligomers in the free state. The flexibility of α-2,6 sialoglycans depends on the torsion angles around the Neu5Ac-Glc glycosidic linkage, namely φ, (C1-C2-O-C6'), ψ (C2-O-C6'-C5') and ω (O6'-C6'-C5'-O5')[126]

where this last torsion is the one providing further rotation freedom to the ligand, as indicated in Figure 3.8. The different conformations existing for α-2,6 sialoglycans come from the different values of the ω torsion: *gg* (-60º), *gt* (+60º) and *tg* (180º).[127]

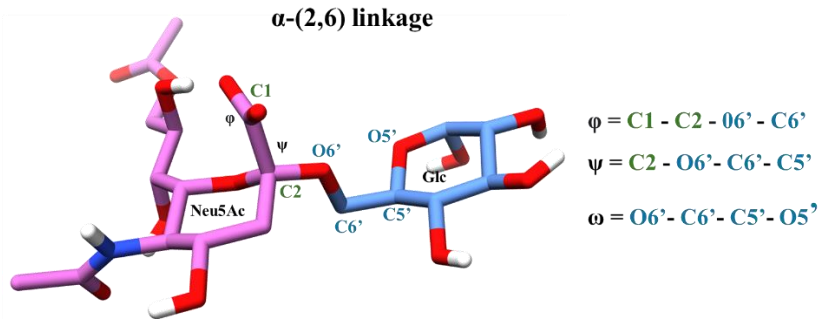

Figure. 3.8 Representation of the three staggered rotamers from the torsional angles described for the linkage α-Neu5Ac-(2,6)-α-Glc: φ, ψ and ω. Where the ω have as main values -60, 60° and 180° namely *gg*, *gt* and *tg* respectively around the C5-C6 bond.

The structure of the 2-mer and 3-mer of Men-Y oligosaccharide portions was determined by analysing NMR data Kuttel et al[128] reported that this serogroup possesses a di-OAc at position 7 and 9 position of sialic acid glycerol chain, whereas we clearly observed only shifts on the chemical environment for 9-OH of sialic acid, and CPS was acetylated on the O9 of the sialic acid on a percentage below the 30% (Figure 3.9).

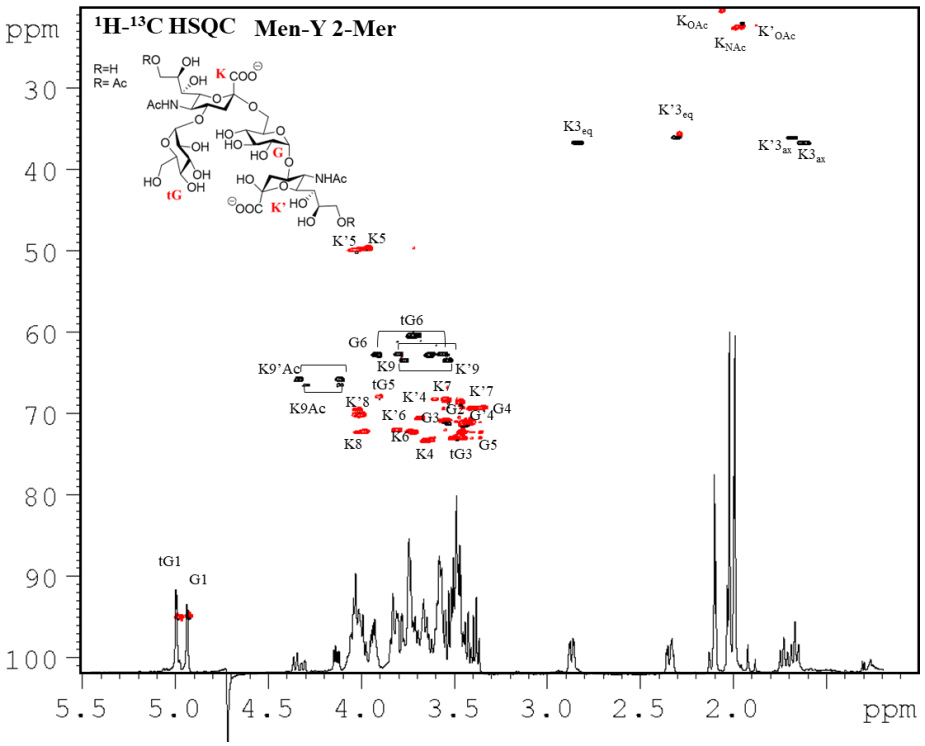

Figure 3.9 ^{1}H-^{13}C- HSQC spectrum of the depolymerized 2-Mer Men-Y CPS from *Neisseria meningitidis*

NOESY experiments were conducted to acknowledge its conformation in the free state, where we could observe a mixture of *gg* and *gt* conformations (Figure 3.10). STD NMR experiments were acquired the Siglec-7 and 2-Mer from MenY CPS, permitting to map the interacting epitope when this ligand is accommodated into the Siglec-7 binding pocket. Therefore, the crucial role of the sialic acid K) in mediating the interaction with Siglec-7 was elucidated; interestingly STD-NMR analysis, the internal sialic acid (Neu5Ac, residue K) was the key residue and showed high relative STD percentages in the binding epitope, as depicted in Figure 3.10. The highest STD enhancements were observed for the acetyl group of the central Neu5Ac and K6, the other STD values were normalized accordingly. Discrete STD effects were observed for the other protons of Neu5Ac moieties (K and K'), together with protons at position 6 of the internal Glc (G). The anomeric proton and H6 of the terminal

Glc (tG) exhibited a lower contribution to the binding. No STD signals were observed for the isolated protons H2, H3 and H4 of the glucose moieties (tG and G), indicating that they were solvent exposed. Analogously, no STD contribution from the O-acetylated diastereotopic protons at position 9 of Sia was observed.

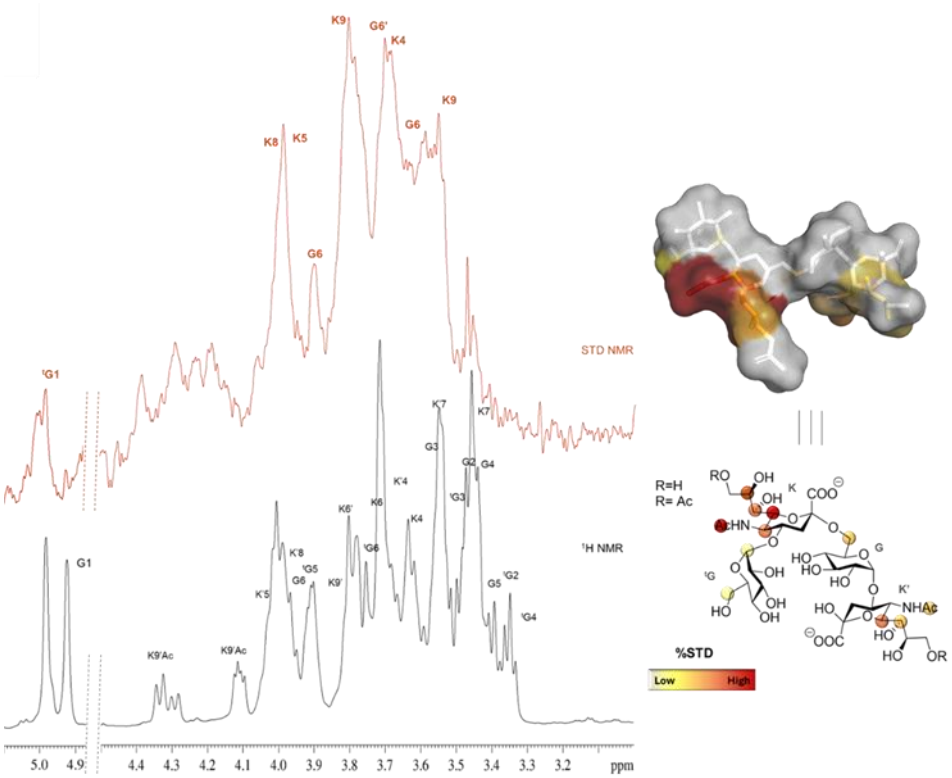

Figure 3.10. STD NMR analysis of Siglec-7 and Men-Y 2-Mer with the ligand epitope mapping calculated by $(I_0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ was the signal intensity in the STD-NMR spectrum (black) and I_0 was the peak intensity of the off-resonance spectrum (black). The highest signal was set to 100% and the other protons were normalized accordingly. On the right: the bioactive conformation of Men-Y 2-Mer as obtained by NMR; the ligand surface was colored according to the STD effects as well as the protons of the ChemDraw structure of Men-Y 2-Mer.

Through the tr-NOESY experiments (Figure 3.11), the bioactive conformation was obtained, giving the most representative conformers displayed by the 2-Mer CPS from *Neisseria meningitidis* when bound to Siglec-7. The multiplicity of the diastereotopic protons at the position 6 of glucose, **G**, together with the vicinal coupling constant between protons H5 and H6, gave a value of $^{3}J_{5,6} < 3$ *Hz*, indicating a preferred conformation, based on the of ω (O6'-C6'-C5'-O5'), corresponding to the *gg* rotamer, where ω=-60º.

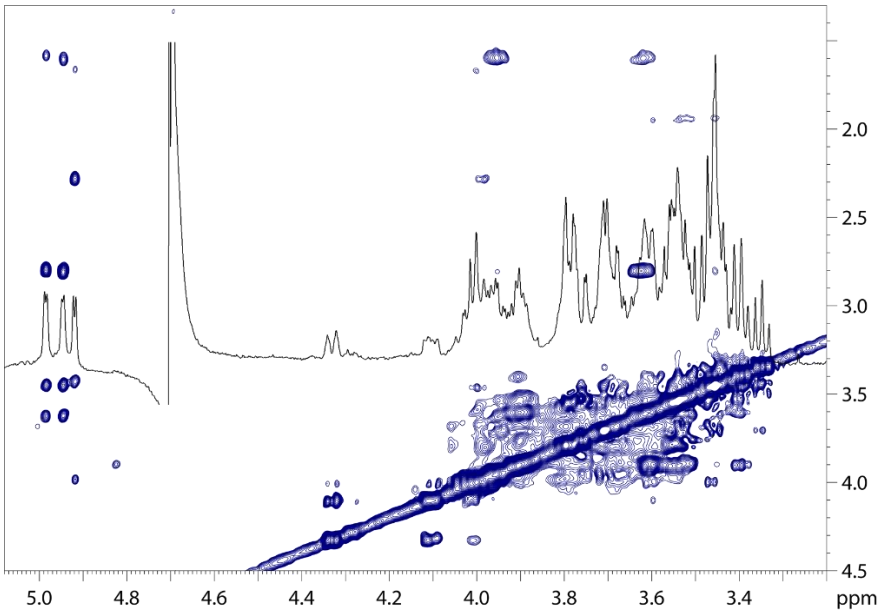

Figure 3.11. tr-NOESY of the 2-Mer in its bound state with Siglec-7 at 400 ms.

3.1.6 Protein-based NMR analysis of the binding between Siglec-7 and Serogroup Y CPS

To unveil the molecular intricacies of the interaction between Siglec-7 and Men-Y, protein-based NMR was also employed. This technique allowed the analysis of the changes on the chemical shifts that suffer the aminoacids of a protein when it forms a complex within a ligand. Aliquots of Men-Y CPS were sequentially added to [^{15}N] labelled Siglec-7 CRD and ^{1}H-^{15}N HSQC NMR

spectra were acquired. As expected, during the NMR titration some of the protein signals experienced chemical shift perturbation (CSP) as explained in Chapter II section 2.5 (Protein-based NMR experiments). We could observe that, as shown in Figure 3.12, some residues, in particular those located in the canonical binding site, such Arg124 and Lys131, together with residues in the BC' and CC' loops, including Asn55 and Ala76, experienced decrease in signal intensity, as depicted in the Figure 3.1.6.1, suggesting their involvement in the interaction with Men-Y CPS.

These changes also involved Ala76 from the CC-loop that seemed to play a pivotal role, disappearing upon binding, resulting in the highest variation when values were normalized.[129] Analysis of the CSP showed similar results, particularly in the signals located near the canonical Arg124, including Tyr26 and Ser27. In fact, it was demonstrated that several signals affected by CSP were located near to and on the GG-loop, including Lys131, Trp132 and Tyr136, Glu137 and Q138, also conserved residues among Siglecs binding sites. These amino acids are indeed generally involved in the establishment of hydrophobic interactions with sialic acids residues, helping the ligands to the stabilization of the complex. Other CSP effects were attributed to Arg92, Phe95 and His96, which suggest the formation of the second binding pocket of Siglec-7, indicating thus an extended binding mode for longer oligosaccharides.

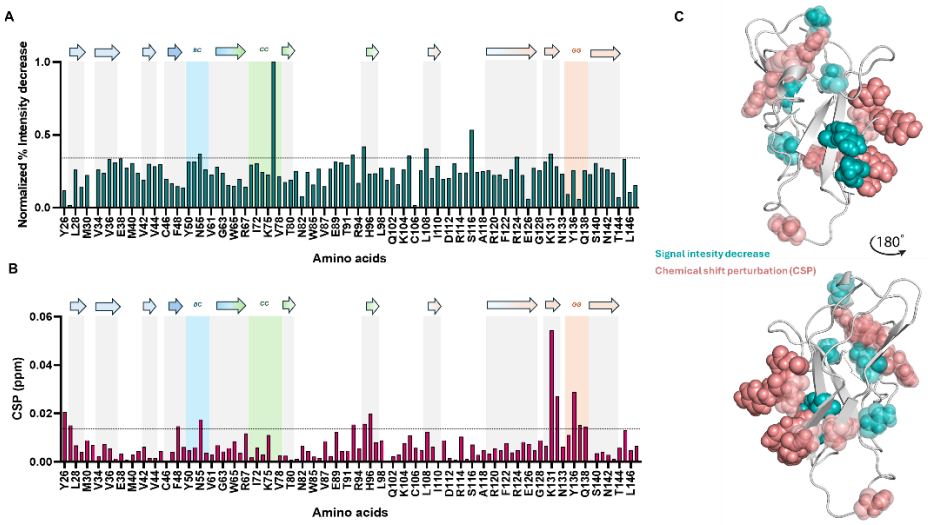

Figure 3.12. A. Diagram of the % intensity decrease of aa of 100 µM Siglec-7 in the presence of 1 mg CPS from MenY. The % intensity decrease effects were evaluated from the of variation of chemical shift heights between the protein in apo form and bound to CPS, then normalized to the maximum value. A threshold (dotted line) was set to 0.35. The residues experiencing the largest signal intensity decrease were Q55, R92, F95, N105, L108, S116, R124, K131. **B.** Diagram of the chemical shift perturbation (CPS) of aa of 100 µM Siglec-7 in the presence of 1 mg CPS from MenY. The CSP effects were evaluated with the formula $\mathrm{CSP} = \frac{1}{2}\sqrt{{\Delta\delta_H}^2 + (\Delta\delta_N/5)^2}$ and a threshold of 0.015 (dotted line) was set. The residues experiencing the largest CSP were Y26, S27, F48, N55, R92, F95, H96, K131, W132, Y136, D137, Q138. **C.** 3D structure of Siglec-7 (PDB 2HRL) evidencing the most perturbed amino acids in signal intensity decrease (teal spheres) and CSP (salmon spheres).

3.1.7 Computational analysis of the molecular binding between Siglec-7 and Serogroup Y

The experimental data obtained through fluorescence, ELISA, ligand-based NMR and protein-based NMR was complemented with computational studies. The approach was to study the 2-Mer and also different lengths of the oligosaccharide in its free and bound state.

The conformational behavior of Men-Y in its free state was evaluated. We began building first the disaccharides to know the most populated conformers of the α2,6 linkage for the Neu5Ac-Glc oligomer. Previous studies and adiabatic energy maps (Figure 3.13) performed with MM3*, showed that φ could explore two predominant torsion values of -60° and 180° while ψ assumes almost stably a value of 180°. On the other hand, the ω torsion is the one displaying the highest flexibility, adopting three different values in the free state corresponding to -60°/60°/180° (*gg/gt/tg*) rotamers, with a preference for the values around -60° *gg* and 60° *gt* on the Neu5Ac-Glc α-2,6 linkage.

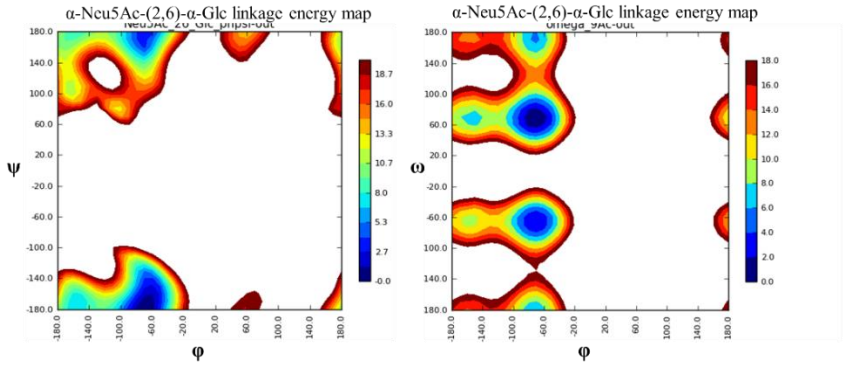

Figure 3.13. Adiabatic energy maps of the two different glycosidic linkages present in the serogroup Men-Y, illustrating the φ, ψ and ω torsions of the Neu5Ac-Glc linkage.

It was also performed the adiabatic energy maps for the α-1,4 Glc-Neu5Ac linkagewith one possible energy minimum allowed, thus possessing only one conformations with φ=-60º and for ψ=-30º, respectively. (Figure 3.14)

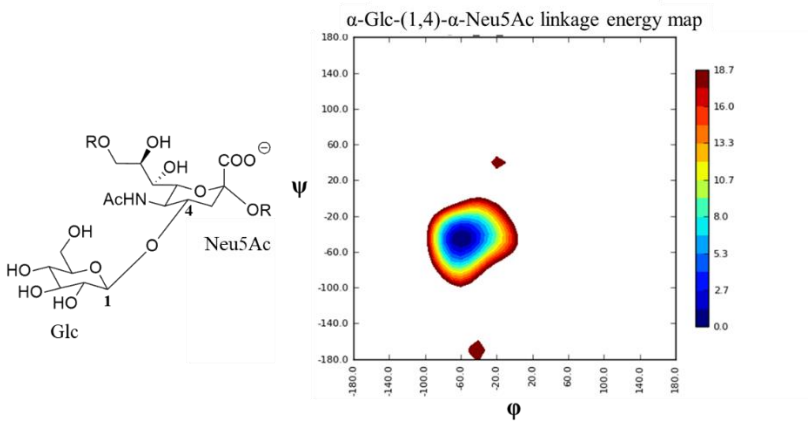

Fig. 3.14. Adiabatic energy map for the φ and ψ torsions or the α-Glc-(1,4)-α-Neu5Ac bond.

Then we could build the 3D structure of the 2-Mer from Men-Y with GLYCAM[130] forcefield. To checked which of the conformers were populated, or which were more representative, several Molecular Dynamic simulations (MD simulations) on the free state were carried out with different timescales, from 100 ns to 500 ns. Through the analysis of the results, we obtained a mixture of the three staggered rotamers *gg/gt/tg* (60%/30%/10%) for α-2,6 linkage on Neu5Ac-Glc. The stability of the structure was also monitored through RMSD (Root Mean Square Deviation), as shown in the figure below. Given the potential biological role of the acetylation at position 9 of Neu5Ac, we computationally explored the binding and conformational features of different acetylation patterns for the oligomers (Figure 3.15). We took in consideration all the possibilities for the 2-Mer: **K** and **K'** both acetylated, only **K'**, only **K**, and none of them. In agreement with the previous NMR results in, we did not observe differences in the population of different conformations based on the position of the acetyl group. This feature will further be studied in complex with Siglec-7. In the next studies we selected **K**, internal Neu5Ac sialic acid residue, as the only residue acetylated in its 9 position, computational studies were performed starting with the crystal structure (2HRL) of the Siglec-

7 and GT1b ganglioside, that acted as a template for the orientation of the 2-Mer from MenY CPS in the binding pocket of Siglec-7.

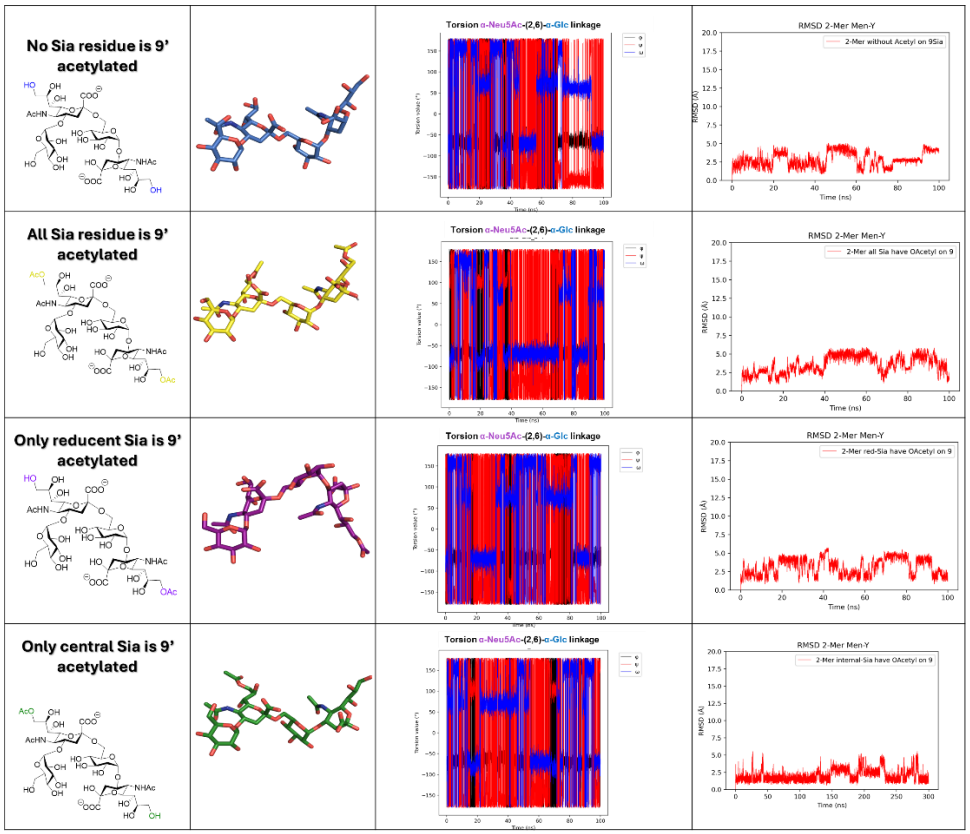

Figure 3.15. Different acetylation patterns for Men-Y MD simulation in the free state. The behaviour of φ, ψ and ω torsion angles during MD simulations of 2-Mer oligomers are reported. Stability of the ligand is represented by the RMSD, plotted in red.

The complex was refined and docking calculations of the tetramer, 2-Mer, inside the Siglec-7 (PDB:2HRL)[131] binding site were performed with AutoDock4.2.The best docked poses were selected, according to their ability to represent STD and NOESY data. The chosen complexes were submitted to MD simulations in explicit solvent and the interactions between sialic acids the binding pocket residues were monitored. In agreement with STD-NMR and tr-

NOESY data, we obtained a reliable 3D complex where the ω torsion angle was restrained at -60º, gg conformation, while other torsion angles were sampled along the simulations. MD analysis have revealed that φ mainly adopted a value around -60º, occasionally sampling 180º, while the ψ torsion remained consistently constant at 180º among all the different simulations, as expected. The stability of the complex formation was demonstrated by RMSD analysis, confirming the stable positioning of the ligand in the Siglec-7 binding pocket. (Figure 3.16)

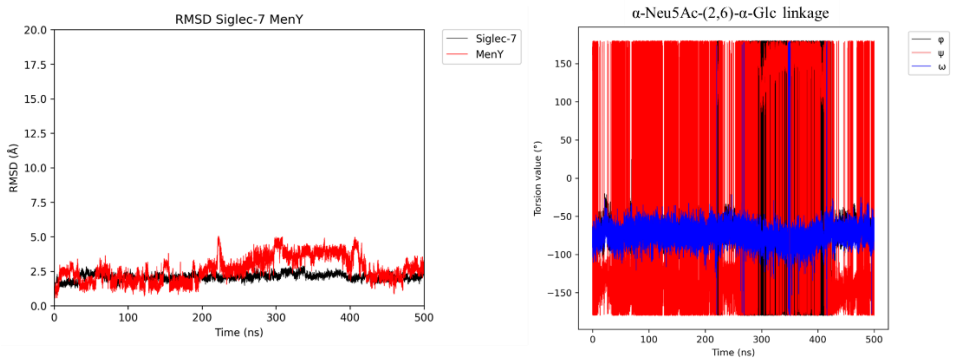

Figure 3.16. RMSD trajectory and dihedral angles around the glycosidic linkage of -Mer monitored along MD simulation in the bound state. A restriction of ω at -60° was applied, according to NMR results.

The interactions were thoroughly mapped between the 2-Mer and Siglec-7, and thanks to this we could describe the molecular intricacies of the binding in a 3D structure. (Figure 3.17) The analysis of the key contacts mainly involved electrostatic and hydrogen bonds interactions, most of them were retained along the MD simulation.

This network of polar interactions was established with highly conserved residues conserved on the B-strand, F, surrounding the canonical Arg124, and on the g-B strand, in which typical amino acids of Siglec-7, like Lys131 and Asn133 are located. In detail, the highly conserved Arg124 established a stable salt bridge with the carboxylate group of the internal Neu5Ac (K) via its guanidinium group. This result is in line with STD-NMR, and in full agreement,

we could also observe that the glycerol moiety of Neu5Ac, **K**, was involved in hydrogen bonds between the 8'-OH position with the Asn133 and Lys131, both located on the G-strand that it is mentioned above. Conversely, the acetyl group at position 9 did not interact with Siglec-7 and remained solvent-exposed through the entire time of the MD simulation. Almost all the protons of both internal (G) and terminal (tG) glucose units were solvent exposed, except those at positions 1 and 6 of tG and G residues, respectively. The hydroxyl group of O6 of tG established a single interaction with the side chain of Asn129, inside the binding pocket of the protein. The model further predicted H-bonds between Lys75 and Asn133 with the carboxyl and acetyl moieties of the reducing sialic acid, K'.

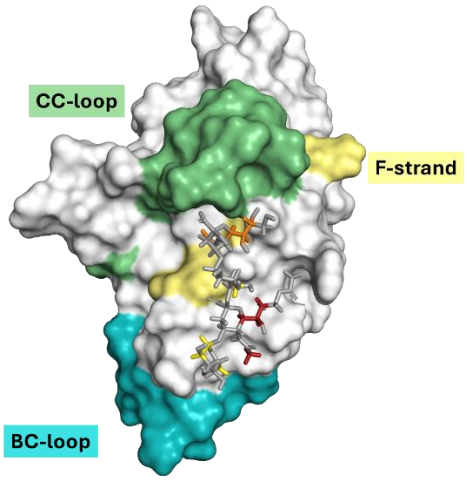

Figure 3.17 3D complex with protein surface colored by strand and loops, as indicated in the figure: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. The ligand surface was colored according to the STD edit code.

Therefore, MD analysis was in line with the findings of STD-NMR and trNOESY, emphasizing the primary involvement of Neu5Ac, K, in the recognition and binding event, and showing the minor role of the reducing K and glucose residues G and tG in the interaction. By the integration of the analysis of the contacts observed in MD simulations, the epitope mapping and

its bioactive conformation derived from NMR data, a comprehensive depiction of Siglec-7 and Men-Y 2-Mer model was attained.

To further investigate the impact of the non-stoichiometric acetylation pattern at the C-9 of Neu5Ac on the conformational behavior of 2-Mer from Men-Y, we computationally investigated the corresponding de-acetylated oligomer, where no significant changes were predicted. Worthy, the absence of the acetyl group allowed to esblish a novel H-bond with Asn133, already involved in the interaction. (Figure 3.18) These findings fully agreed the STD NMR experiments, from which we could observe only STD effects from the non-acetylted protons at position 9 of Neu5Ac.

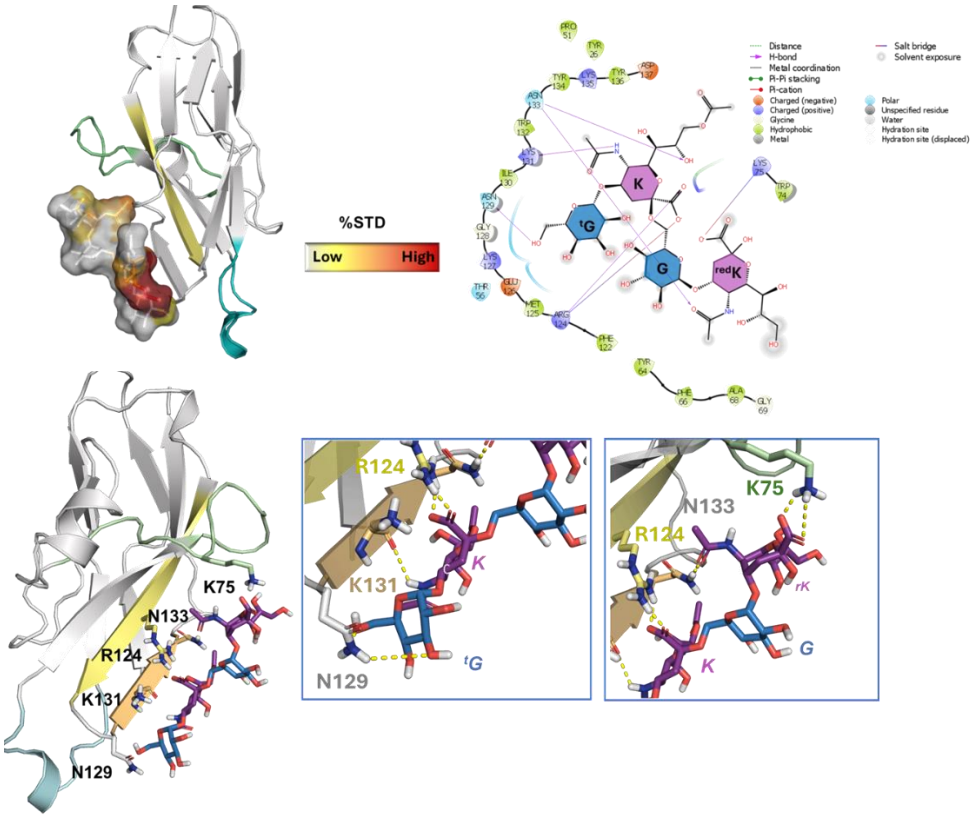

Figure 3.18. A. 3D views of the Siglec-7/ Men-Y 2-Mer complex: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. **B.** 2D plot of the interactions occurring at the Siglec-7 and 2-Mer's from Men-Y interface. **C.** Different views highlighting the H-bonds monitored by MD simulation were shown. Ligand 2-Mer is represented following SNFG color-code

Longer oligomers were built and faced with Siglec-7, obtaining similar results, since the protein-based NMR experiments were performed on the full-length capsular polysaccharide, as shown in Figure 3.19. The data obtained through MD simulations resulted in good agreement with the protein-based experiments. In this case, we decided to use a complete de-acetylated structure, due to the null involvement with the binding pocket. Again, internal sialic acid was the moiety primarily recognized by the protein, interacting with its carboxylate group and the conserved Arg124. The longer oligomer permitted further accommodation of the ligand in the binding pocket of Siglec-7, observing novel interactions, like the presence of Tyr26 and Asp53, in agreement to the previous protein-based NMR studies.

Some other interactions were predicted, including the reducing hydroxyl and Arg120, which helped the stabilization of the ligand in the pocket. Side interactions were also observed like the glucose unit G'' and Ile72, and Gly69. Therefore, we obtained a similar accommodation for 2-Mer and the longer 4-Mer, having the internal sialic acid establishing the key salt bridge with Arg124.

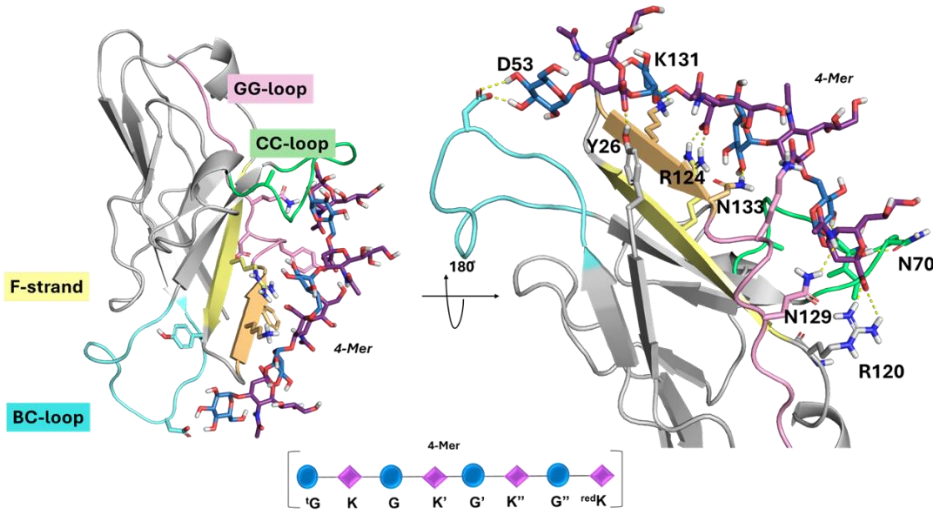

Figure. 3.19. 3D model of the Siglec-7 and Men-Y 4-Mer complex. Different views highlighting the H-bonds monitored by MD simulation were shown. The strands involved in the recognition are colored by type: F-β strand in yellow, BC-loop in cyan, CC'-loop in green and GG-loop in pink. Ligand 4-Mer is represented following SNFG color-code. Also, represented in surface the 3D model of Siglec-7/4-Mer complex.

Preliminary immunological studies were also conducted. A first evaluation of the ability of Men Y CPS to affect the host immune response was explored by investigating the effects of low and high doses of MenY on U937-derived macrophages cells that express Siglec-7 on their surface, and THP-1 derived magrophages that we used as control, having a lower Siglec-7 expression as previously described. The inflammatory response of macrophages is characterized by an increased the secretion of many pro-inflammatory and chemotacticcytokines; we chose to analyze TNF□ and IL-8 as pro-inflammatory cytokines and CXCL10 expression as chemotactic cytokine. The treatment of THP-1-derived macrophages induced a weak but significant pro-inflammatory and chemotactic response to 0.1 μg/ml of MenY treatment as evidenced, respectively by IL-8 and CXCL10 upregulation. This response was even more amplified after treatment with 1 μg/ml of MenY, as displayed by the very strong upregulation of the three genes, especially for CXCL10. On the other hand, U937-derived macrophages appeared indifferent to both doses of

the CPS of MenY, as the expression of TNF and IL-8 cytokines and CXCL10 chemokine was comparable to non-treated cells, indicating that this capsular polysaccharide was not able to activate the immune response, due also thanks to the interaction with Siglec-7, which inhibits the immune response by its ITIM domain. This evidence was further confirmed by treating both U937- and THP-1-derived macrophages with same amounts of LPS. As showed in Figure 6, both cells could efficiently induce a pro-inflammatory response already after 0.1 µg/ml of LPS treatment, and this response was of increased magnitude at 1 µg/ml.

This evidence was confirmed by treating both U937 and THP-1 derived macrophages with the same amounts of LPS. (Figure 3.20)

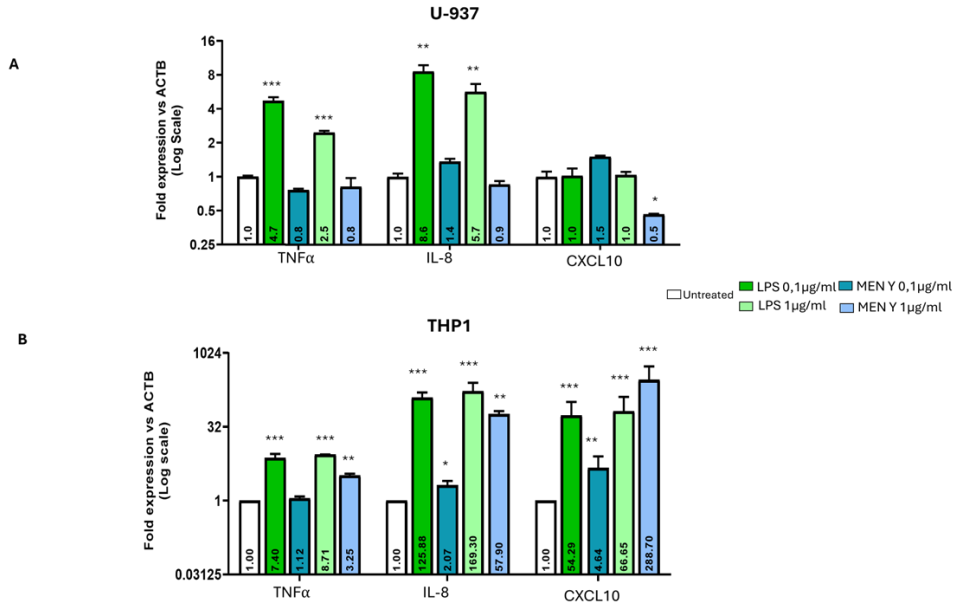

Figure 3.20. Semi-quantitative analysis on the production of the proinflammatory cytokines from U937 derived macrophages cell-line (**A**) and THP-1 derived macrophages cell-line (**B**) after the stimulation with Men-Y CPS. LPS was used as a positive control.

3.1.8 Discussion

The specific recognition of sialylated CPS from *Neisseria meningitidis* serogroup Y (Men-Y) by Siglec-7 has been demonstrated. The preliminary immunological studies found that the CPS selectively binds Siglec-7, respect with other human lectins, and induces a significant pro-inflammatoyry and chcemotactic response, suggesting that Siglec-7 interaction inhibits the immune response by the interaction with MenY. To dissect the binding features and unravel the molecular basis of the recognition process, a multidisciplinary approach was used, combining fluorescence titrations studies, ligand and protein- based NMR, in silico analysis and biological assays. The carbohydrate recognition domain (CRD) of Siglec-7 in *E. coli* and the amino acids resonance were assigned to provide the protein fingerprint and perform the binding studies. The extracellular domain FED of Siglec-7 was produced in HEK293S cells as glycosylated protein for the binding studies. The depolymerization of Men-Y CPS allowed to obtain oligosaccharides with repeating units of increasing length, enabling a detailed evaluation of their binding interactions with Siglec-7. Through NMR spectroscopy we could detail the structure of Men-Y CPS, confirming an acetylation pattern of the Neu5Ac O-9 below the 30%. Fluorescence titrations provided a K_D in the micromolar range. STD-NMR provided the epitope mapping and highlighted the role of the internal sialic acid K in the binding event, particularly with its acetyl group and glycerol chain. Computational studies, including MD simulations allowed to describe the interactions at the Siglec-7/2-Mer interface, occurring primarily hydrogen bonds and electrostatic interactions, showing the crucial role of the salt bridge between the conserved Arg124 and the internal Neu5Ac, K carboxylate. Indeed, this was the key interaction, that together with Asn133 and Lys131 were in complete accordance with the STD NMR experiments.

This chapter provided insights in the importance of capsular polysaccharides from *Neisseria meningitidis*, since they are crucial virulence factors currently used in a plethora of glycoconjugate vaccines. We here demonstrated that

sialylated CPS have gained particular interest for the interaction with inhibitory host receptors, like Siglec-7 to evade immune surveillance. This understanding offered vision into the mechanism used by *Neisseria meningitidis* to manipulate the host immune responses thereby paving the way for the development of new targeted interventions to mitigate its impact on human health.

To this aim, the comprehensive analysis developed in this work provides a full view of the specific recognition *Neisseria* Men-Y by Siglec-7. In particular, preferential binding, representative interactions and role of the conformational flexibility around the α-2,6 sialylated glycans, offering high complexity and specificity on host-pathogen interactions.

.

3.2 Molecular Recognition of Siglec – 7 and other sialylated CPS from *Neisseria meningitidis*

3.2.1 Introduction

In section 3.1, we used a multitechnique approach to study how at a molecular level the recognition mechanism between Siglec-7 and serogroup Y CPS from *Neisseria meningitidis*.

This chapter will instead focus on the investigation of the recognition between Siglec-7 with other sialylated serogroups of the same pathogen, specifically serogroups W and B. In the one hand, serogroup W, an oligomer composed by [4)-α-Neu5Ac-(2→6)-α-Gal-(1→]$_n$ has been associated with an increase in invasive meningococcal disease in the last years. And, on the other hand Serogroup B which is a homopolymer of sialic acid with α-2,8 linkages, offers a unique challenge to vaccine development due to its molecular similarity to host neural tissues since they express polysialic acid, a feature that we are sure that will also impact Siglec-7 recognition.

The investigation on the interaction of Siglec-7 with these pathogenic serogroups can give a deep insight into the high specificity of Siglec-7 among serogroups, not only by linkage type, but also through conformation. This can provide new information for the design of therapeutic strategies targeting Siglec-mediated immune modulation in meningococcal infections.

3.2.2 Molecular details of Serogroup W and Siglec-7

In the previous chapter, it has been described the interaction between Serogroup Y of *Neisseria meningitidis* and Siglec-7, providing a comprehensive understanding, at a molecular level, of the structural and biophysical aspects that contribute to the binding event. Serogroup Y, while structurally similar to Serogroup W, displayed distinct affinity patterns toward Siglec-7, which was

highlighted through ELISA assays. We also focused on Serogroup W[132,133] to understand and elucidate why these two capsular polysaccharides (CPS), despite their structural similarities (Figure 3.21), exhibit very different binding affinities to Siglec-7.

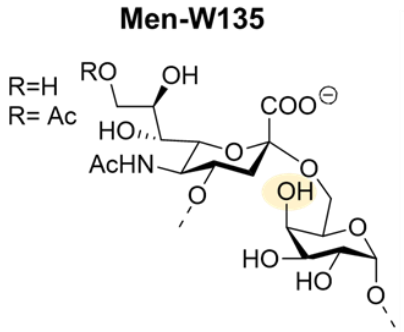

Figure 3.21. Schematic representation of Serogroup W CPS from *Neisseria meningitidis.*

Serogroup W contains Neu5Ac-Gal units, with a α-2,6 sialyl linkage and Gal α-1,4 linked. It is known for its significant pathogenicity and global impact on the meningitis disease and its increase worldwide.

While the structural similarities between Serogroups Y and W suggest that they should bind similarly to Siglec-7, our data revealed differences in their affinity, which suggest that this could be due to subtle structural variances in their 3D disposition. We deepen into the molecular dynamics of Serogroup W and Siglec-7 interaction by a combination of fluorescence techniques and computational analysis that helped to elucidate the binding mode of this particular CPS. The fluorescence titrations helped to explore the affinity with Siglec-7, and computational modelling allowed to visualize the molecular details in the binding pocket of Siglec-7. This assessed how these small structural differences between Serogroups Y and W influenced their receptor-binding behavior.

Fluorescence titrations and ELISA immunological assays for the binding Serogroup W and Siglec-7.

ELISA assays between *Neisseria meningitidis* serogroups and Siglecs different recombinant soluble forms of Siglecs (Siglec-5, Siglec-7 and Siglec-9) and human C-type lectins (DC-SIGN, Langerin, MGL and MR)) were (Figure 3.22) showed that Siglec-7 showed low affinity against Men-W (Neu5Ac-Gal) compared to the very similar serogroup Men-Y (Neu5Ac-Glc), only differing in the 4' hydroxyl position being equatorial for glucose, and axial for galactose. It seems that these differences will induce a difference in the 3D arrangement that affects the recognition events. Steady-state fluorescence analysis was performed to determine the binding affinity between MenW and Siglec-7. We obtained a K_D of 132 μM result that confirms the ability of Siglec-7 to interact with MenW from *Neisseria meningitidis* with a low affinity. The conformational features that govern these different binding preferences between MenW and MenY were thoroughly studied through computational studies in the next paragraphs.

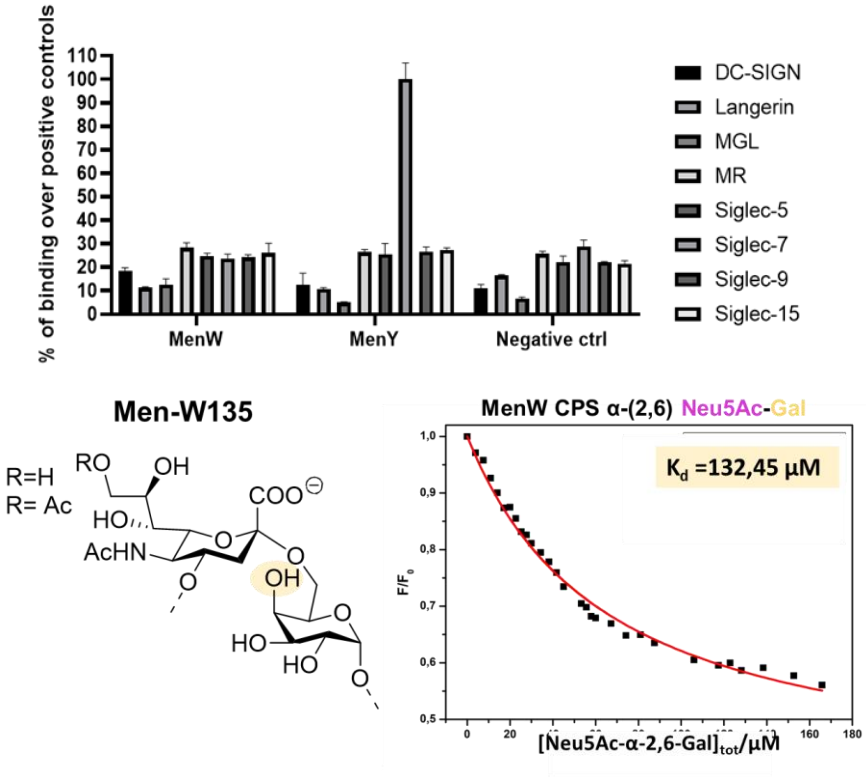

Figure 3.22. A. ELISA solid-phase assay performed between different C-type lectins and Siglecs-Fc. Men-W CPS showed a low binding to Siglec-7 compared to Men-Y. Polyacrylamide polymers (PAA) coated with different glycans were used as positive controls. Wells were coated with the reported CPS and the binding to different C-type lectins and Siglecs was evaluated. The experiment has been performed three times in duplicate with similar results. Data are normalized over binding signal from the positive controls used for each lectin (set as 100% of binding). Error bars indicate standard deviations. OD: Optical density. Fluorescence titration of Siglec-7 upon the addition of increasing concentrations of Men-W The emission spectra were recorded by using an excitation wavelength of 295 nm and a temperature of 25ºC.

Computational analysis of the molecular binding between Siglec-7 and Serogroup W

The three-dimensional disposition of sialoglycans with α-2,6 linkage depends on the ω torsion (O6'-C6'-C5'-O5'), which possess three main values which

confer the conformer names *gg, gt,* and *tg,* respectively. The adiabatic energy maps were calculated by confronting first, φ and ψ linkages. (Figure 3.2.2.3) Since the ψ, as expected, remained constant with one possible value at -180º, the ω torsion was analysed. We obtained, as shown in Figure 3.23 three staggered conformations defined by the ω torsion of α-2,6 linkages: *gt, gg* and *tg*. Several reports have demonstrated that Neu5Ac-Gal linkages rarely populate the *gg* conformation.[134] So, for the Neu5Ac-Gal linkages the main rotamers expected will be *gt* and *tg* for this specific sialoglycan bond. In the case with galactose, it has been shown in previous studies that the *gt* is one of the preferred ones. So this starting 3D conformation was used to build our MenW oligomer, (as 2-Mer, 3-Mer).

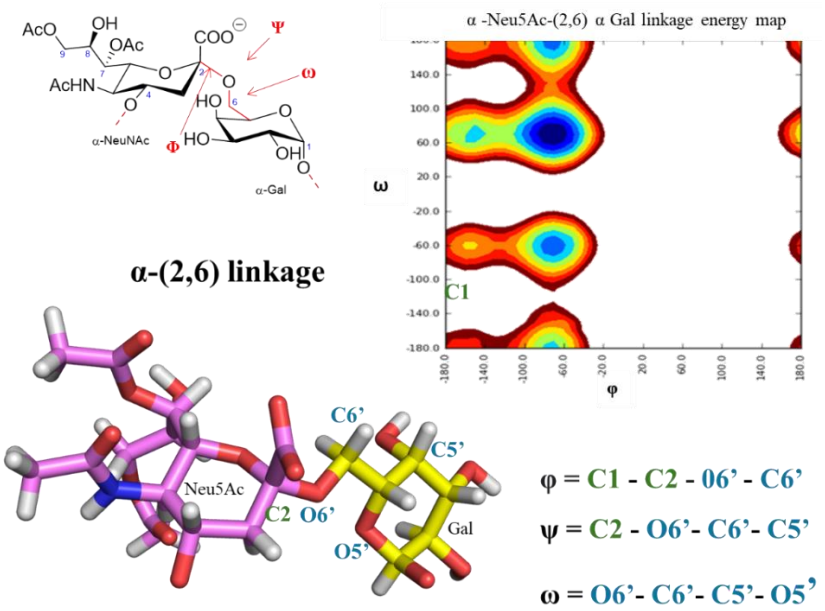

Figure 3.23. Adiabatic energy maps of the two different glycosidic linkages present in the serogroup Men-Y, illustrating the φ, ψ and ω torsions of the Neu5Ac-Glc linkage. Below, the representation of the three staggered rotamers from the torsional angles described for the linkage α-Neu5Ac-(2,6)-α-Glc: φ, ψ and ω. Where the ω have as main values -60, 60º and 180º namely *gg*, *gt* and *tg* respectively around the C5-C6 bond

MD simulations in explicit solvent were run in the free state; we obtained a preferred conformer for the α-2,6 torsion on Neu5Ac-Glc, at ω=+60º, known as *gauche-trans*, or *gt.* (Figure 3.24) The stability of the structure was also monitored through RMSD (Root Mean Square Deviation), as shown in the figure below. MenW was also reported to be acetylated in the 9 position of the Neu5Ac, so given its potential biological importance, we computationally explored the binding and conformational features of different acetylation patterns for the oligomers. We took in consideration two possibilities for the 2-Mer of MenW: both sialic acids acetylated and none of them. After an intense analysis where RMSD was gave high, and torsion values were sampled, no observed differences in the population of different conformations were observed based on the position of the acetyl group.

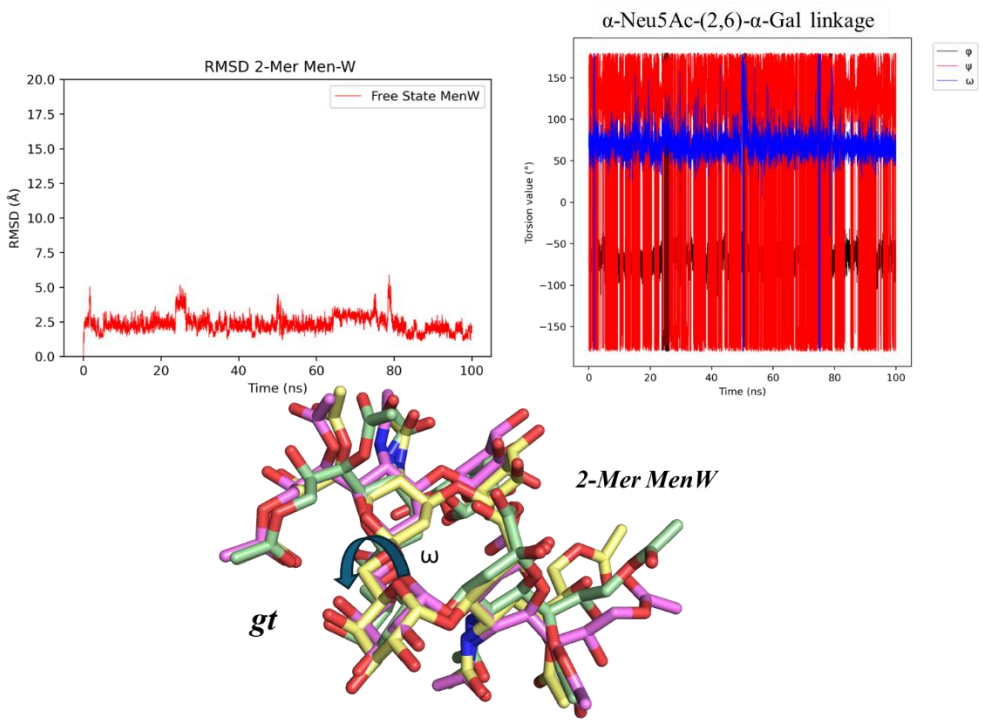

Figure 3.24 The behaviour of φ, ψ and ω torsion angles during MD simulations of the 2-Mer of MenW oligomer are reported. Stability of the ligand is represented by the RMSD, plotted in red and the glycosidic torsion of the α-2,6 Neu5Ac-Gal linkage

sampled. Below, most representative poses for the MD simulation on the free state of the 2-Mer for MenW.

Free state computational analysis permitted to define *gt* as the most representative conformer of this serogroup. We also studied the conformational behavior on longer oligosaccharides, such as the 3-Mer, obtaining analogous results. With this, the 3D model of Siglec-7 and MenW was built. Analogously to the MenY bound state, crystal structure 2HRL of Siglec-7 was used as a template for the 3-Mer of Men-W CPS (*gt* rotamer). We used a longer oligosaccharide to be able to do a fair comparison within the MenY results. So, first, the structure was refined and docking calculations of the hexamer, 3-Mer inside the Siglec-7 binding site were performed with AutoDock4.2. The chosen complexes from Docking were submitted to MD simulations in explicit solvent and the interactions between sialic acids the binding pocket residues were monitored. We obtained a reliable 3D complex where the ω torsion angle adopted a prevalent value of 180º, *gt* conformation, differently from the free state, wher the only rotamer was *gt*. Further analysis has revealed that φ mainly adopted a value around -60º, while the ψ torsion remained consistently constant at 180º among all the different simulations, as expected. The stability of the complex was demonstrated by RMSD analysis, confirming the positioning of the ligand in the Siglec-7 binding pocket. (Figure 3.25)

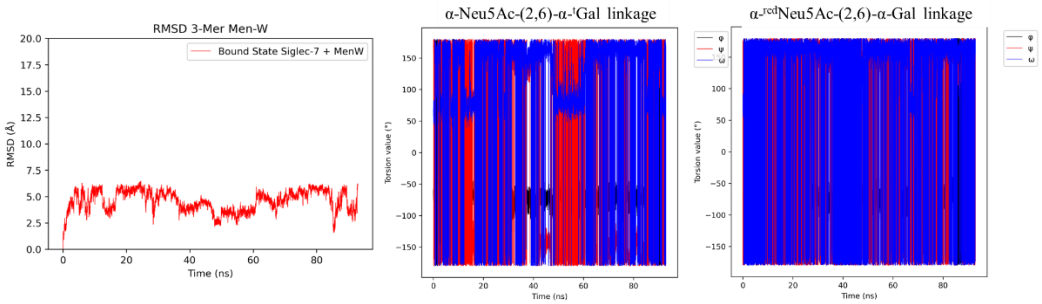

Figure 3.25. RMSD trajectory and dihedral angles around the glycosidic linkage of 2-Mer monitored along MD simulation in the bound state.

The interactions were mapped between the 2-Mer of MenW and Siglec-7, and thanks to this we could describe the molecular details of the binding event with the 3D structure. The analysis of the key contacts mainly involved electrostatic and hydrogen bonds interactions, most of them were retained along the MD simulation. This network of polar interactions was established with conserved residues conserved on the β-strand, F, surrounding the canonical Arg124, and on the G-β strand, in which typical amino acids of Siglec-7, like Lys135 and Asn129 are located. In detail, the highly conserved Arg124 established a stable salt bridge with the carboxylate group of the internal Neu5Ac via its guanidinium group, as shown in the Figure 3.26. This result is the expected, since it is behaving in a similar way like MenY did. We could also observe that the glycerol moiety of Neu5Ac, was involved in hydrogen bonds with the Asn129 and Lys135, both located on the G-strand that it is mentioned above. Analyzing the role of the acetyl at position 9, it did not interact with Siglec-7 and remained solvent-exposed throughout the MD simulation, as expected. Almost all the protons belonging to both internal (Gal) and terminal (tGal) galactose units remained solvent exposed, except those at positions 3 of tGal.

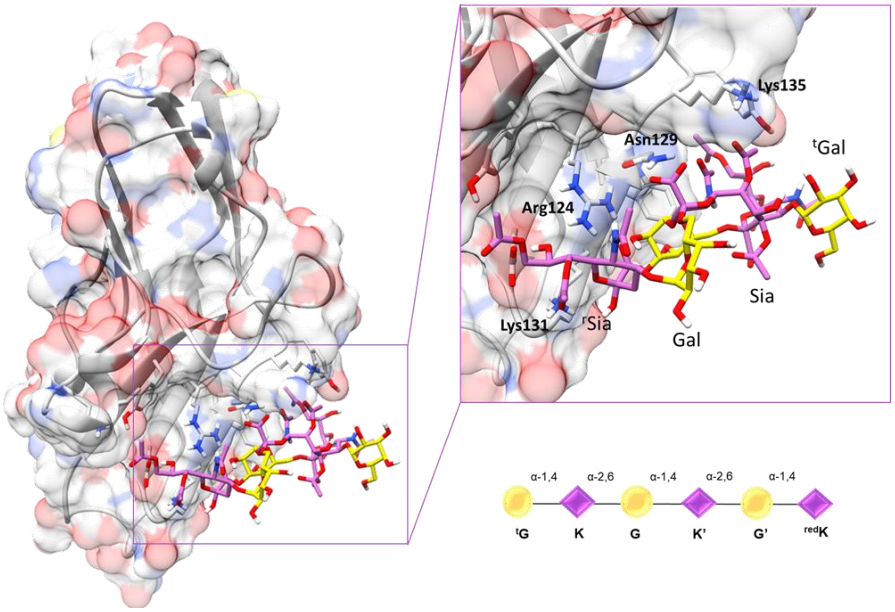

Figure 3.26 Different views highlighting the H-bonds monitored by MD simulation were shown. Ligand 2-Mer is represented following SNFG color-code.

By comparing the 3D model before and after simulation, we could definitively see a shift on the reducing part of the Siglec-7, moving from the CC-loop (before MD) to the BC-loop (after MD), as decpited in the Figure 3.27.

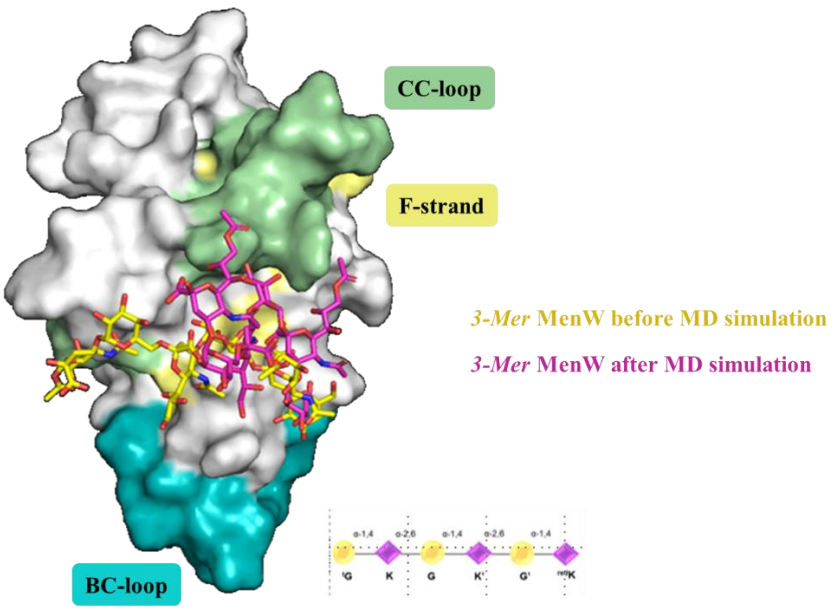

Figure 3.27 3D complex with protein surface colored by strand and loops, as indicated in the figure: the strands involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green. The ligand surface was colored according to: 3-Mer before simulation in yellow and 3-Mer after MD simulation in violet purple.

The spatial rearrangement of MenW seemed to be induced by the value of its ω torsion, since the initial 3D structure was based on the free state results, which showed a unique value of ω=60º, *gt* conformer. The conformational change from *gt* to *tg,* ω=180º provoked a different accommodation of MenW inside the binding pocket of Siglec-7.

Discussion

The recognition of sialylated CPS from *Neisseria meningitidis* serogroup W (Men-W) by Siglec-7 has been demonstrated. Preliminary ELISA studies found that the CPS binds Siglec-7 with a lower affinity respect with the other its structurally similar serogrou colleague, MenY. To dissect the binding features and unravel the molecular basis of the recognition process to understand the difference of these affinities, a multi-disciplinary approach was used, combining fluorescence titrations studies, *in silico* analysis and biological assays. Repeating units of increasing length were evaluated for the conformational behavior and after building of the 3D structure, enabling a final detailed analysis of their binding interactions with Siglec-7. The fluorescence titrations provided the K_D in the micromolar range for the system of Siglec-7 Men-W. Computational studies, including MD simulations allowed to describe the difference between the conformation performed by MenW in the free and bound state. In the free state + *gt* rotamer was the only value for the ω torsion of this serogroup. Conversely, once it interfaced the Siglec-7, a conformational change occurred obtaining as main conformation the *tg*. This shift provoked a 3D rearrangement of the 3-Mer MenW inside the Siglec-7 binding pocket.

The comprehensive analysis developed in this work provides a full view of the specific recognition *Neisseria* Men-W by Siglec-7. Notably, this study exhibits the difference in the binding affinities of MenY (Neu5Ac-Glc) and MenW (Neu5Ac-Gal), due to the different values acquired by the ligands on their ω torsions, confirming the high complexity and specificity on host-pathogen interactions.

3.2.3 Molecular Recognition of Siglec – 7 and Serogroup B CPS from *Neisseria meningitidis*

Neisseria meningitidis holds thirteen CPS surrounding cell envelope Nm and representing their major virulence factors, playing a crucial role in *N.*

meningitidis pathogenicity, making them the main component of meningococcal vaccines as CPS or as protein-conjugate vaccines. Between the thirteen serogroups, six of them are the cause of the invasive meningococcal disease, (A, B, C, X, Y, W). Serogroup B has become the major cause of this infection, being reported by the health authorities from America, Europe, South America and Australia.[135] PolySia is mainly present in the nervous system, particularly in microglia.[136] It can cause neuronal damage during this bacterial infection, and it could be correlated to the appearance of neurodegenerative disease. The study of the molecular interactions and pathways linking infections with neuronal damage remain mostly unexplored.Apart from the α-2,6 linkage, Siglec-7 possesses even higher affinity for α-2,8 endogenous ligands. Unlike other capsular polysaccharides, Serogroup B CPS is the only from N. *meningiti*is holding a α-2,8 polysialic acid (PolySia), likely binding Siglec-7. The molecular intricacies of Serogroup B and Siglec-7 interaction was assessed through a combination of fluorescence techniques, ligand-based NMR and computational analysis, definitively key to elucidate the binding mode of this particular CPS. The fluorescence titrations helped to explore the affinity with Siglec-7, ligand-based NMR to describe epitope mapping and preliminary results on bioactive conformation, and computational modeling allowed to visualize the molecular details that happen in the binding pocket of Siglec-7.

Characterizacion on the Free State of Serogroup B from Neisseria meningitids

MenB was extracted and purified from the dried cells *of Neisseria meningitidis* MC58 strain. In parallel, oligomers containint MenB repeating units were furnished by Prof. Hiromune Ando from the Gifu University in Japan (Figure 3.28).

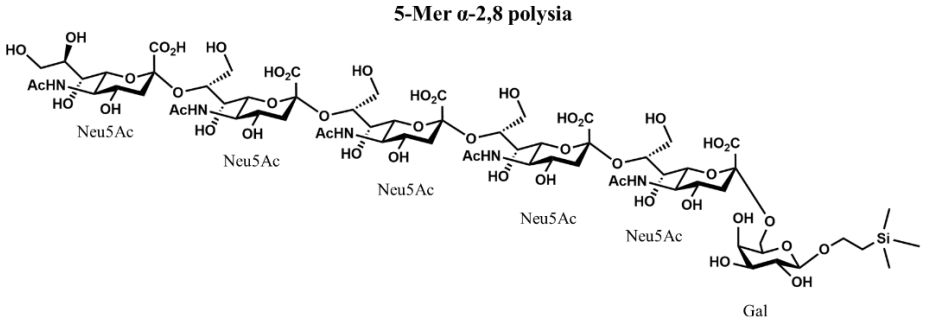

Figure 3.28. Schematic representation of 5-Mer from MenB.

PolySia contains α-2,8 linkages also present in terminal sialic acid from gangliosides, We can observe high conformational flexibility, due to the presence of three ω torsions, known as ω_7, ω_8 and ω_9, as depicted in the Figure 3.29.

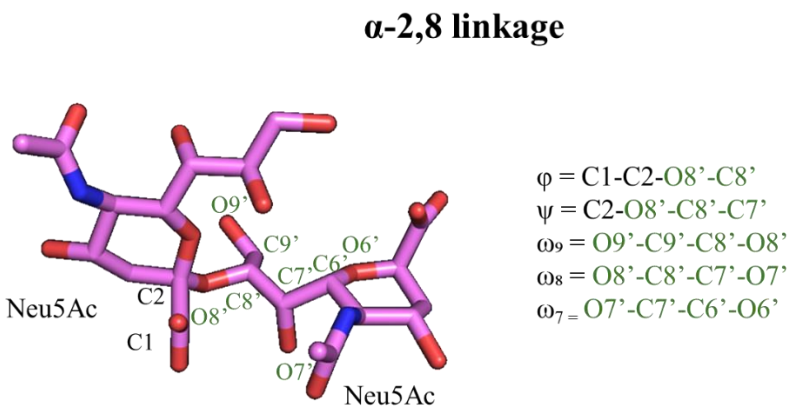

Figure 3.29. Representation of the three staggered rotamers from the torsional angles described for the linkage α-Neu5Ac-(2,8)-α-Neu5Ac φ, ψ , ω_7, ω_8 and ω_9, respectively.

The free state NMR spectra of the 5-Mer were analysed, confirming five sialic acid residues and a protected galactose moiety. So, computational calculations were performed to help to depict the conformational behavior of MenB in its free state (Figure 3.30).

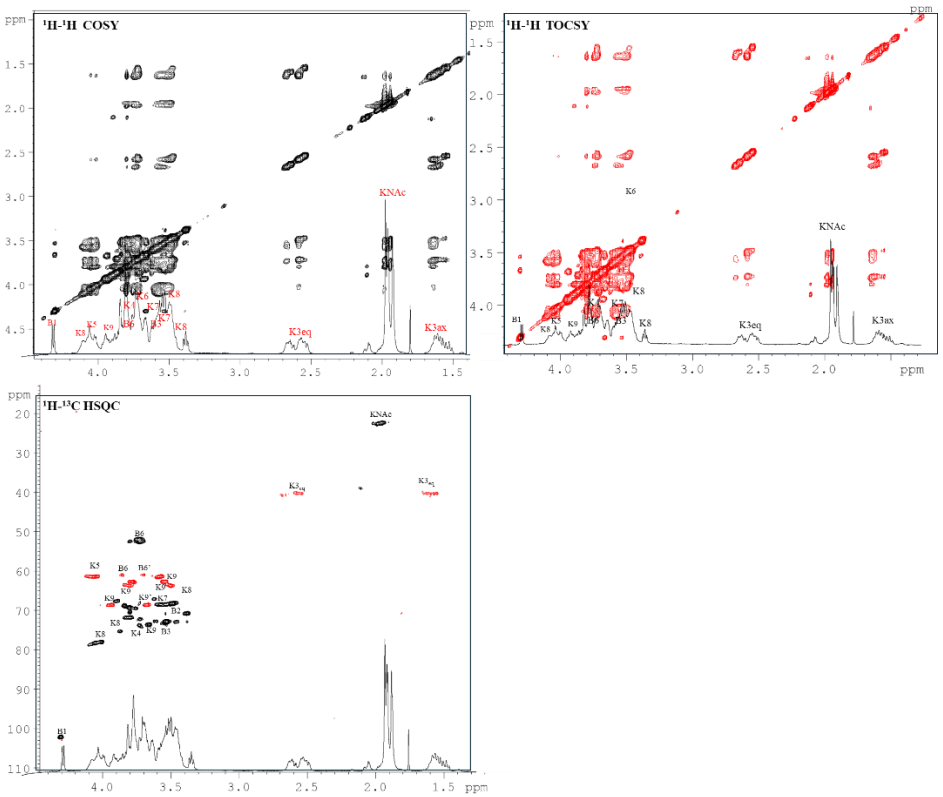

Figure 3.30. NMR experiments for the conformational analysis of 5-Mer of MenB CPS *Neisseria meningitidis* in the free state. A. ^{1}H-^{1}H COSY and ^{1}H-^{1}H TOCSY on the upper part. Below, the ^{1}H-^{13}C HSQC spectrum.

The conformational behavior of 5-Mer from Men-B CPS was evaluated in its free state via computational analysis. The disaccharide of Neu5Ac-(2,8)-Neu5Ac unit was built to describe the most populated conformers for the α-2,8 linkage in MenB oligomer. (Figure 3.31) Through a molecular mechanics approach, employing MM3*, we obtained the adiabatic energy maps that provided the φ, ψ and the different ω_7, ω_8, and ω_9 values. Minima for this linkage were first based on the φ and ψ torsions, which acquired φ= -80º and ψ at 60º, and also secondaty φ= -60º and ψ at 110º. Then we confronted φ with

each ω torsion, ω_7, ω_8, and ω_9, respectively. Even though it is a highly flexible linkage, many studies suggest that the inter-hydrogen bonds that the molecule acquires on the PolySia[137,138]provokes a clear preference on the formation of a helicoidal structure, providing unique value for each conformation. ω_9 torsion mainly adopted a value of 60º as preferred, ω_8, being the only one having more than one rotamer, defined by its values of -60º or 60º, and lastly, a unique conformer for the ω_7 of 60º has been shown.

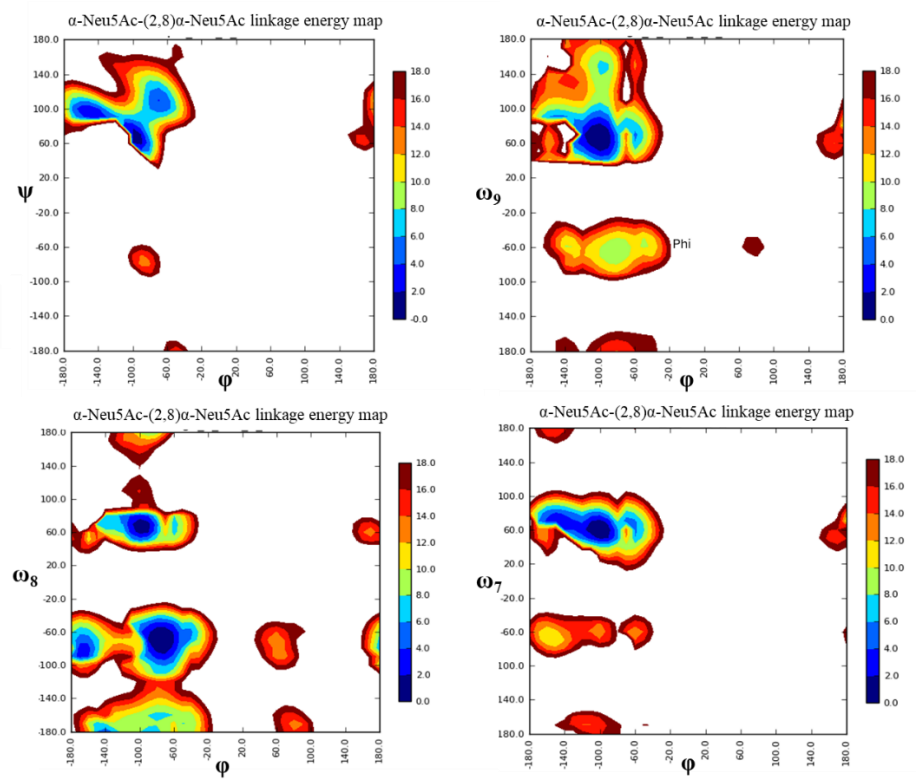

Figure 3.31 Adiabatic energy maps for the α-Neu5Ac-(2,8)-α-Neu5Ac linkage.

Once each α-2,8 linkage for MenB was set to its minimum, the 3D structure was built employing the GLYCAM forcefield and database. In order to comprehend which was the most representative conformer of MenB, Molecular Dynamic simulations (MD simulations) on the free state were carried out on

100 ns timescales. The stability of the structure was also monitored through RMSD (Root Mean Square Deviation), as shown in the Figure 3.32 below.

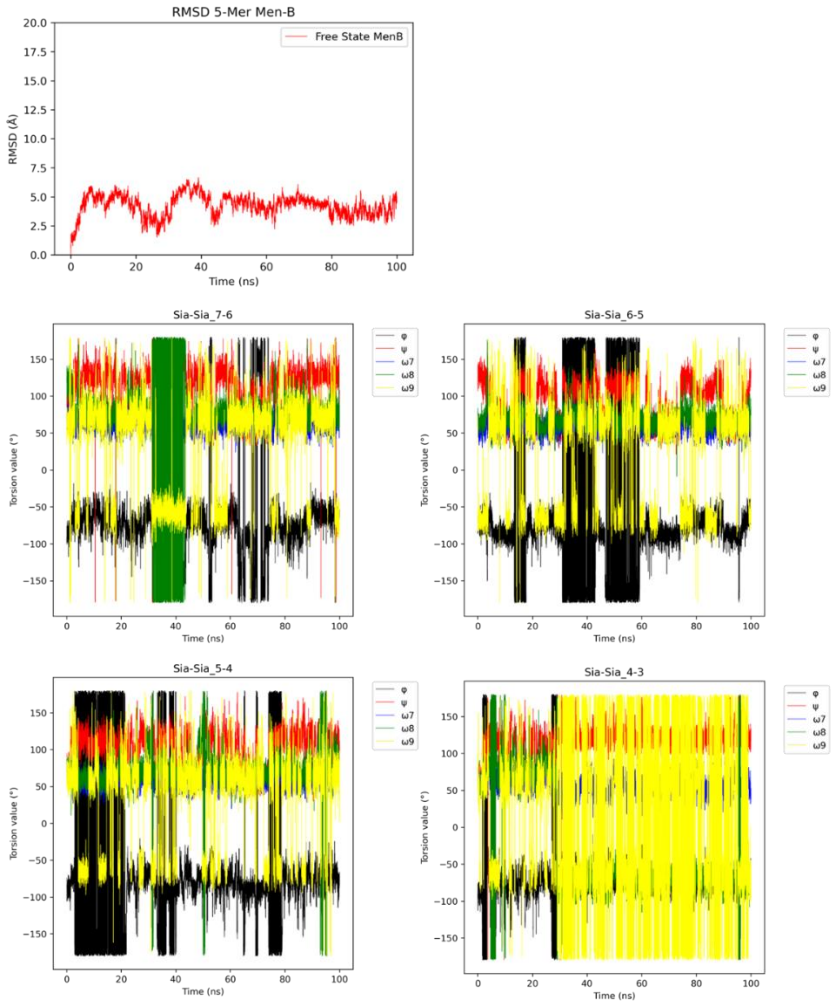

Figure 3.32. The behaviour of φ, ψ and ω_7, ω_8 and ω_9 torsion angles during MD simulations of the 5-Mer of MenB oligomer are reported. Stability of the ligand is represented by the RMSD, plotted in red and the glycosidic torsion of the α-2,8 Neu5Ac-Neu5Ac linkage sampled.

Looking at the 3D structure of the 5-Mer, it was observed an helicoidal-like disposition of the sialic acids, through the formation of internal hydrogen bonds that stabilized this orientation. (Figure 3.33)

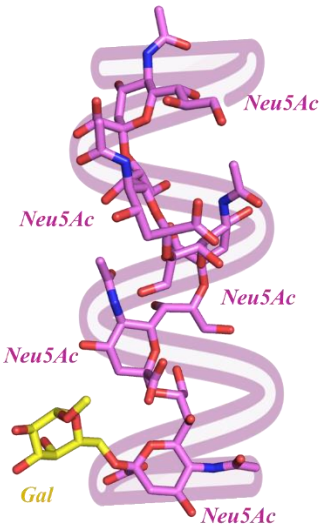

Figure 3.33. 3D representation of the spatial arrangement of the 5-Mer of MenB from *Neisseria meninigitidis* in its free state after a 100 ns of simulation.

Fluorescence titrations for the molecular binding of Serogroup B and Siglec-7

Fluorescence analysis was performed to determine the binding affinity between MenB 5-Mer and Siglec-7.

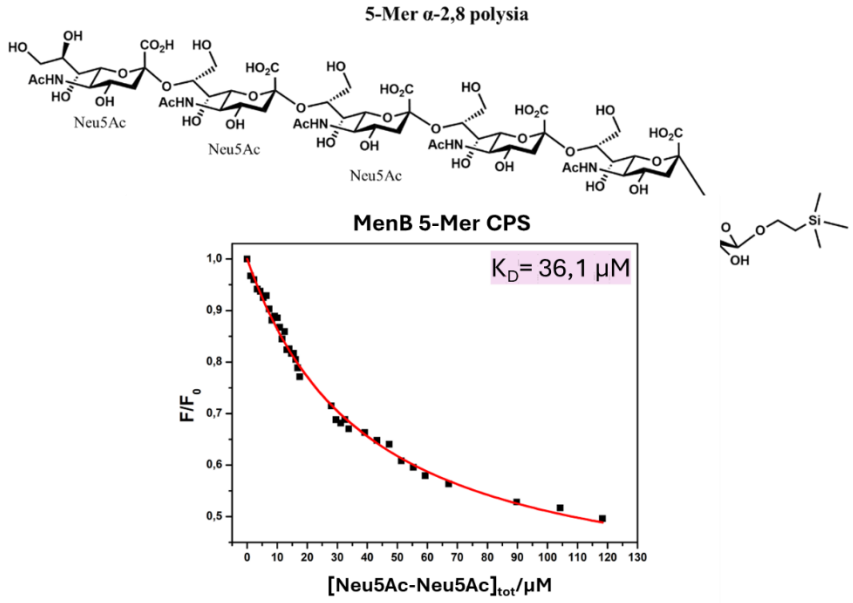

Figure 3.34. Fluorescence titration of Siglec-7 upon the addition of increasing concentrations of Men-B The emission spectra were recorded by using an excitation wavelength of 295 nm and a temperature of 25ºC.

We obtained a K_D of 36,1 in the micromolar range. This result confirms the high affinity of Siglec-7 towards α-2,8 sialylated ligands. (Figure 3.34)

STD NMR and tr-NOESY analysis of the molecular binding between Siglec-7 and Serogroup B – NMR

STD NMR experiments were acquired on Siglec-7 and 5-Mer from MenB CPS from *Neisseria meningitidis* complex, permitting to obtain a preliminary interacting epitope when this ligand is accommodated into the Siglec-7 binding pocket. Due to the high complexity of the oligomer, STD enhancements were detected for sialic acid, denominated as **K**, with the highest STD effects belonging to its acetyl group, AcK. The other STD values were normalized accordingly. (Figure 3.35) Lower STD effects were observed for the other protons on the glycerol chains of **K,** and no STD responses were observed for protons at position 4 of sialic acids or internal protons from galactose residue, indicating the position of these groups towards the solvent.

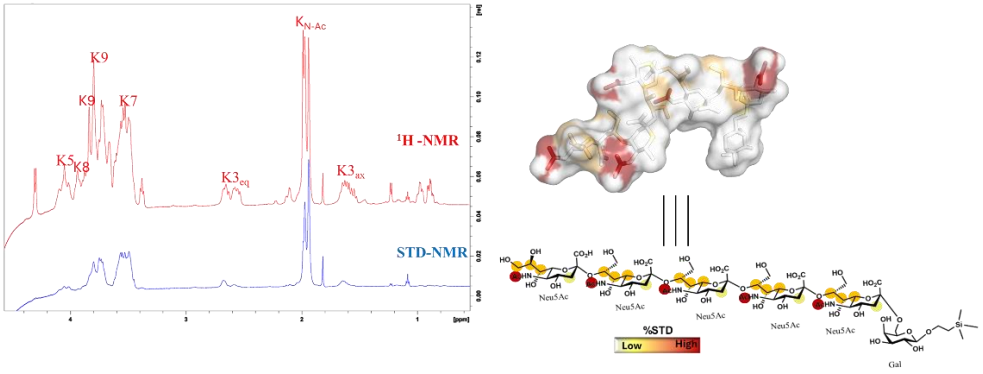

Figure.3.35. STD NMR analysis of Siglec-7 and Men-B 5-Mer with the ligand epitope mapping calculated by $(I_0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ was the signal intensity in the STD-NMR spectrum (blue) and I_0 was the peak intensity of the off-resonance spectrum (red). The highest signal was set to 100% and the other protons were normalized accordingly. On the right: the bioactive conformation of Men-B 5-Mer as obtained by NMR; the ligand surface was colored according to the STD effects as well as the protons of the ChemDraw structure of Men-B 5-Mer.

trNOESY and trROESY experiments were performed to define the bioactive conformation. In this case, the most important peaks that we could define to acknowledge this conformation came out overlapped, as observed in the Figure 3.36

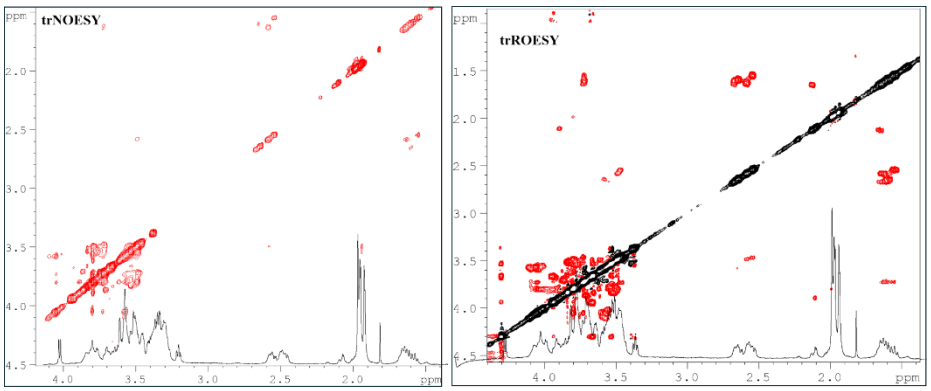

Figure 3.36. tr-NOESY (left) and trROESY (right) of the 5-Mer MenB in its bound state with Siglec-7 at 400 ms and 600 MHz.

Computational analysis of the molecular binding between Siglec-7 and Serogroup B

Docking calculations of the 5-Mer inside the Siglec-7 binding site (PDB:2HRL) were performed with AutoDock4.2. The chosen complexes from Docking were submitted to MD simulations in explicit solvent and the interactions between sialic acids and Sige-7 binding pocket were monitored. We obtained a reliable 3D complex where the values for the different torsion angles adopted on the sialic acids interacting with Siglec-7, as indicated in the Figure 3.37, we observed that the sialic acid attached to the galactose might be too far away to be successfully recognize by the binding pocket of Siglec-7. In fact, it gets positioned towards the solvent, indicating that in Siglec-7, only three sialic acids residues get fully recognized.

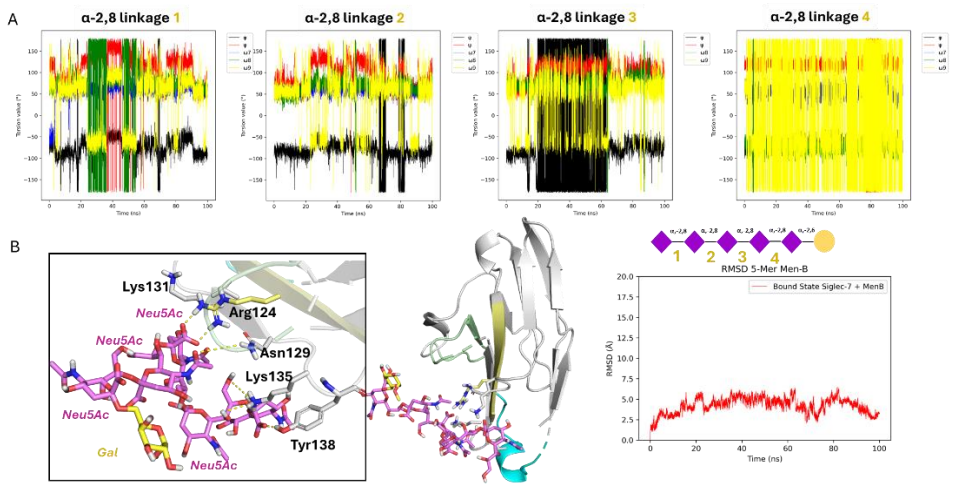

Figure 3.37. RMSD trajectory and dihedral angles around the glycosidic linkage of 5-Mer monitored along MD simulation in the bound state.B**.** Different views highlighting the H-bonds monitored by MD simulation were shown. The strands of of the protein involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green Ligand 5-Mer is represented following SNFG color-code

The analysis of the key contacts mainly involved electrostatic and hydrogen bonds interactions, most of them were retained along the MD simulation. This network interactions were established with conserved residues conserved on the β-strand, F, surrounding the canonical Arg124, and on the G-β strand, in which typical amino acids of Siglec-7, like Lys135 and Asn129 are located. In detail, the highly conserved Arg124 established a stable salt bridge with the carboxylate group of the internal Neu5Ac via its guanidinium group.

It was also observed that the N-Ac moiety of the adjacent internal Neu5Ac (linkage 2) was involved in hydrogen bonds with the Asn129 and Lys135, both located on the G-strand that it is mentioned above.

3.3.6 Discussion

The specific recognition of sialylated CPS from *Neisseria meningitidis* serogroup B (Men-B) by Siglec-7 has been demonstrated.

Fluorescence titrations provided the K_D in the micromolar range for the system of Siglec-7 Men-B. Computational studies, including MD simulations allowed to describe the difference between the conformation acquired by MenB in the free and bound state. We showed the difference in the binding affinities of different serogroups MenB > MenY > MenW. It was also revelead the importance of the flexible omega torsions on the 3D disposition of these ligands towards a certain receptor, since the achievement of a bioactive conformation is key and is due to the different values acquired by the ligands on their omega torsions, confirming the high complexity and specificity on host-pathogen interactions.

3.4 Molecular Recognition of Siglec – 7 and OPS from *Fusobacterium nucleatum* ATCC 51191

3.4.1 Introduction

Fusobacterium nucleatum is an anaerobic Gram-negative bacterium, one of the most significant components of the oral microbiota.[139,140]

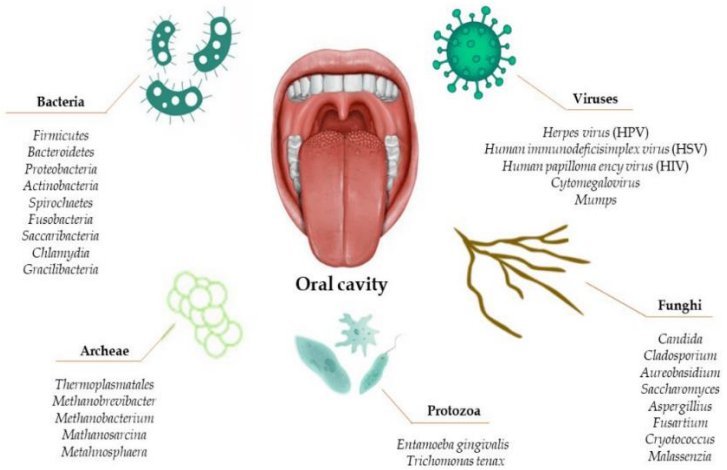

Figure 3.38. Microbioma division in the oral cavity.

However, it has been demonstrated that several disorders are directly related to the presence of *F. nucleatum* inside or outside the human oral cavity, positioning it as a potential pathosymbiont. Firstly, it is involved in the progression of periodontitis, specifically in the oral epithelium[141]. Due to this duality, and the fact that through the oral cavity can arrive to the digestive system and bloodstream, it highly contributes to the etiology of other infections and diseases (Figure 3.38). For example, the case of the intrauterine infections, provocking preterm birth miscarriages and neonatal sepsis. *Fusobacterium* strains contribute to severe throat infections, to Lemierre's[142] syndrome, as well as to gastrointestinal diseases, like inflammatory intestine disease, appendicitis and colorectal cancer. [143,144]

Bacterial Lipopolysaccharide (LPS) are main component of the bacterial outer membrane (Figure 3.39) and are potent activator of the host innate immune response primarly through interaction with TLR4/MD2 receptorial complex.

The LPS from *F. nucleatum* ATCC 51191 O-antigen was characterized, leading to the identification of several glucuronic acids as well as a fucose residue as components of the repeating unit.[145] The structure is reported below.

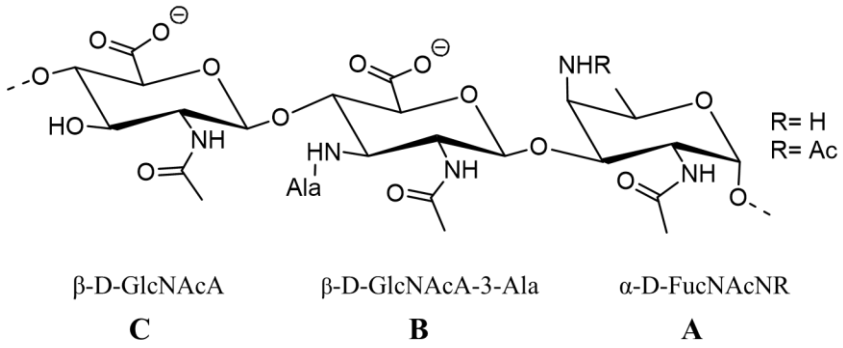

Figure 3.39. Schematic representation of the structure of the OPS from *Fusobacterim nucleatum* ATCC 51191.

Recently, Siglec-7 has been shown to mediate the immumodulation by interacting with lipopolysaccharides (LPS) and OMVs of colorectal cancer associated *Fusobacterium nucleatum* strains. (Figure 3.4.1.3) through the induction of a pro-inflammatory profile in human monocyte-derived dendritic cells (moDCs) and the tumour associated profile in human monocyte-derived macrophages (moMfs), that was followed by the *F. nucleatum* Siglec-7 interaction. [146]

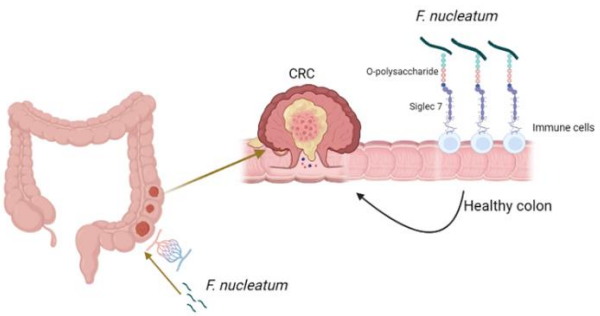

Figure 3.40 Schematic representation of the interaction between F. nucleatum and immune cells in the context of colorectal cancer.

For this reason, we dissected the molecular basis of the recognition of the LPS from *F. nucleatum* 51191 by Siglec-7, through ligand-based NMR techniques, and *in silico* methodologies, to propose a structural model of the interaction. Globally, the acquired results allowed to determine preliminarily the elements of *F. nucleatum* LPS binding by Siglec-7, thus improving the knowledge on the mechanism of action of this opportunistic oncopathogen.

3.4.2 *F. nucleatum* ATCC 51191 OPS and Siglec-7 binding studies through STD NMR

First, the LPS *of F. nucleatum* was extracted from 51191 strain by hot aqueous phenol method and purified as described in the experimental section. Then, the O-polysaccharide was isolated from the lipid A portion upon mild acid hydrolysis The chemical structure of the OPS, showed the presence of [→4)-β-d-GlcpNAcA-(1→4)-β-d-GlcpNAc3NAlaA-(1→3)-α-d-FucpNAc4NR-(1. A partial acetylation (around 30%) of the α-N-Acetyl-Fucose-NHR was observed. Once the OPS was isolated and purified, the recognition binding events between the *F. nucleatum* 51191 and Siglec-7 was explored by STD NMR experiments. From STD NMR, it was possible to qualitatively assess the recognition of the OPS. Particularly, the involvement of the three residues comprising the OPS: B, C and A, respectively, was detected. It seemed to have a higher involvement on residues B and A, but it is challenging task to distinguish between some proton signals of B and C.

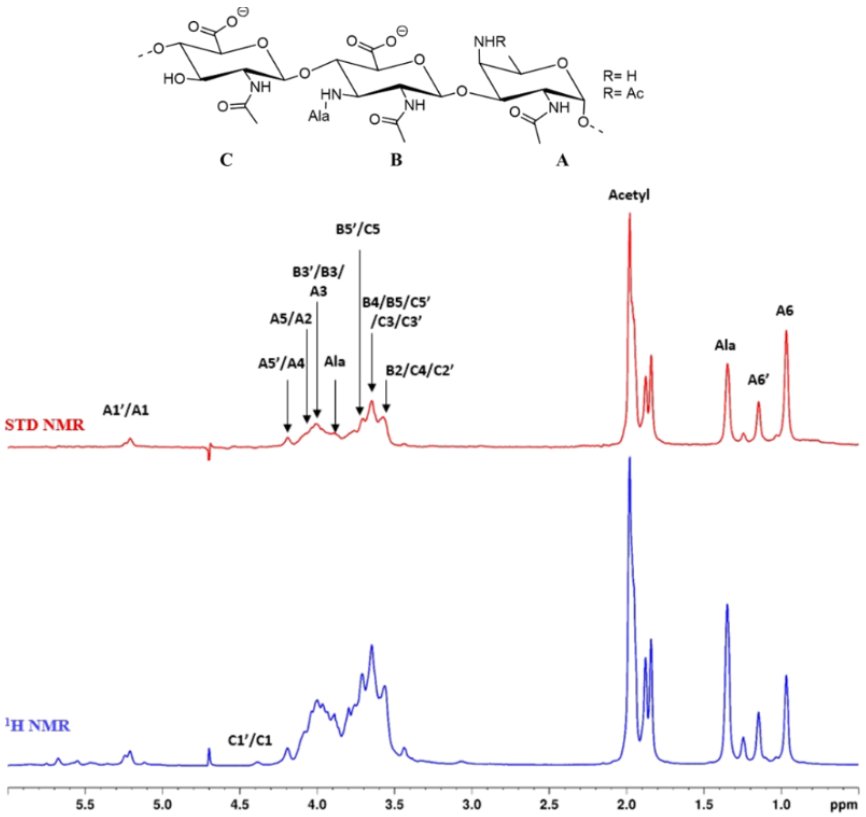

Figure 3.41. STD NMR analysis of Siglec-7 and OPS from *F. nucleatum* ATCC 51191. STD-NMR spectrum (red) the off-resonance spectrum (blue).

Starting from the protons of C and B, the pyranose ring was embedded in the binding pocket of Siglec-7, since many of their protons gave STD enhancements.

[1]Characterization performed by Pilar García-Vello.
[2]STD-NMR experiment performed by Ferran Nieto Fabregat.

3.4.3 Computational analysis: free and bound state of OPS *F. nucleatum* 51191 and Siglec-7

The investigation of the binding features between OPS from *Fusobacterium nucleatum* ATCC 51191 and Siglec-7 was also studied through computational analyses, so new insights could be added to the recognition process. No N-

acetylated, or substituted glucuronic acids were available within this database. So, for this reason, the aim has become the parametrization of the building blocks that are present in this OPS (Figure 3.42). To do so, old protocols and new libraries of forcefields were taken into account for the development of a new protocol that allowed us not only to parametrize new ligands, to use with GLYCAM forcefields.

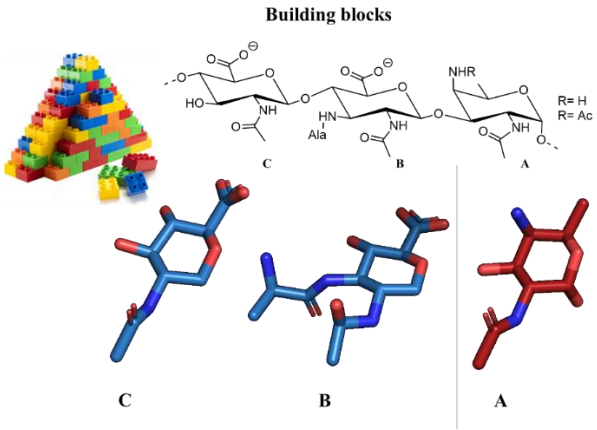

Figure 3.42. Building blocks conforming the OPS from *F. nucleatum* ATCC 51191 for the construction of its 3D structure. Each residue is represented in sticks colored with SNFG color coding (Fucose, A: red, Glucuronic acids, B and C: blue)

Once parametrised, molecular mechanics were applied to calculate the linkage properties of this oligosaccharide, to calculate the glycosidic torsions that govern their 3D disposition, divided in pairs. The adiabatic energy maps were calculated as it follows: B-C, B-A and A-C linkages, respectively as shown in Figure 3.43

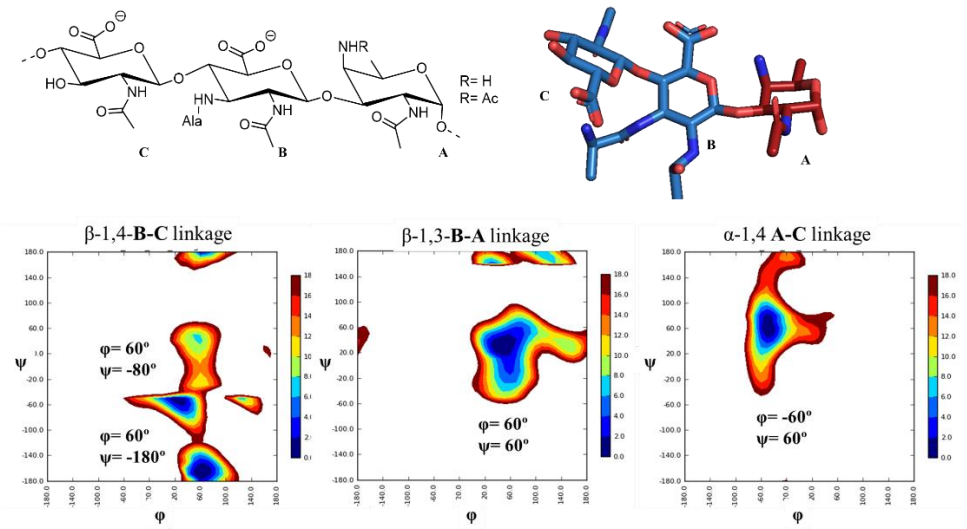

Figure 3.43. Adiabatic energy maps for the: B-C GlcNAcA -(1,4)- β-GlcNAcA-3Ala β-; B-A β-GlcNAcA-3Ala-(1,3)-α-FucNAc-NHR; A-C α-FucNAc-NHR-(1,4)- β-GlcNAcA, respectively.

Minima for these linkages are based exclusively on the φ and ψ torsions, which acquired the following values: for the B-C linkage, a minimum of φ= 60º and a value of ψ at 180º, and another possible minimum at φ= 60º and ψ at -80º were found. Then, B-A linkage gave as possible minimum a unique value for each torsion of φ= 60º and a value of ψ at 60º. Lastly, the only α- configured linkage, A-C, also had one unique value for each torsion, being the value at φ= -60º and ψ at 60º its minima. With this information, the 3D structure of one repeating unit of OPS from *F. nucleatum* 51191 ATCC was built, and its conformational behavior was furtherly studied through MD simulations on its free state. The stability was also monitored through RMSD (Root Mean Square Deviation), and torsions (φ, ψ) for each linkage were sampled along the simulation, as shown in the Figure 3.44 No conformational changes were observed during the MD simulation.

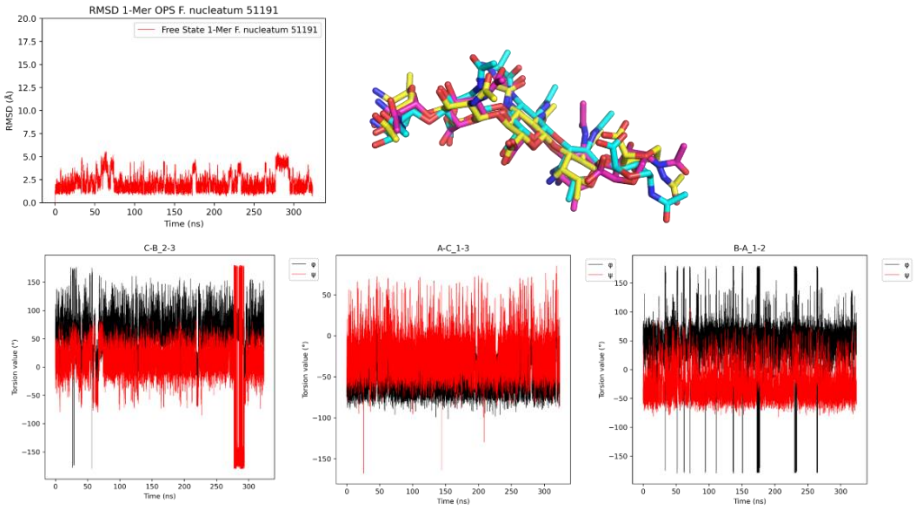

Figure 3.44. The behaviour of φ, ψ during MD simulations of the different linkages of 1-Mer of OPS from *F. nucleatum* 51191 are reported. Stability of the ligand is represented by the RMSD, plotted in red and the glycosidic torsions are sampled and divided into C-B, A-C and B-A, as it corresponds

It is noteworthy that this oncopathogen is able to produce an OPS that gets recognized by a Siglec without possessing a sialic acid in its core. So, the molecular details to take into account are based on the "molecular mimicking" of the glucuronic acid of the sialic acid, in order to fit into the binding pocket of Siglec-7. We decided to use first one repeating unit and also longer oligosaccharide lengths. So, first, the structure was refined and docking calculations of the 1-Mer (with caps, as a pentasaccharide), inside the Siglec-7 binding site (PDB:2HRL) were performed with AutoDock4.2. In this case we can have B interacting with Arg124, or residue C interacting. Due to the ambiguous STD-NMR results, since both of them establish interactions with some residues of the binding site, we decided to use both and simulate the best result among all. The chosen complexes from Docking were submitted to MD simulations in explicit solvent and the interactions between sialic acids the binding pocket residues were monitored. Residue B, the GlcNAcA-3Ala, was the first docked and establishing the key salt bridge with conserved residue Arg124 and submitted to MD simulations in explicit solvent. The stability of

the complex was monitored through the RMSD, and as depicted in Figure 3.45, it goes up to 7.5 A, demonstrating this was not a reliable complex. Nevertheless, the interactions that were governing the complex were mainly polar contacts with other conserved residues of the binding pocket, apart from the salt-bridge between B residue and Arg124. Indeed, A and C established hydrogen bonds with amino acids from the F such Asn129, and from the G-strand like Lys131 and Lys135. Hydrophobic interaction from the Trp132 located on the CC-loop interaction and B was also observed.

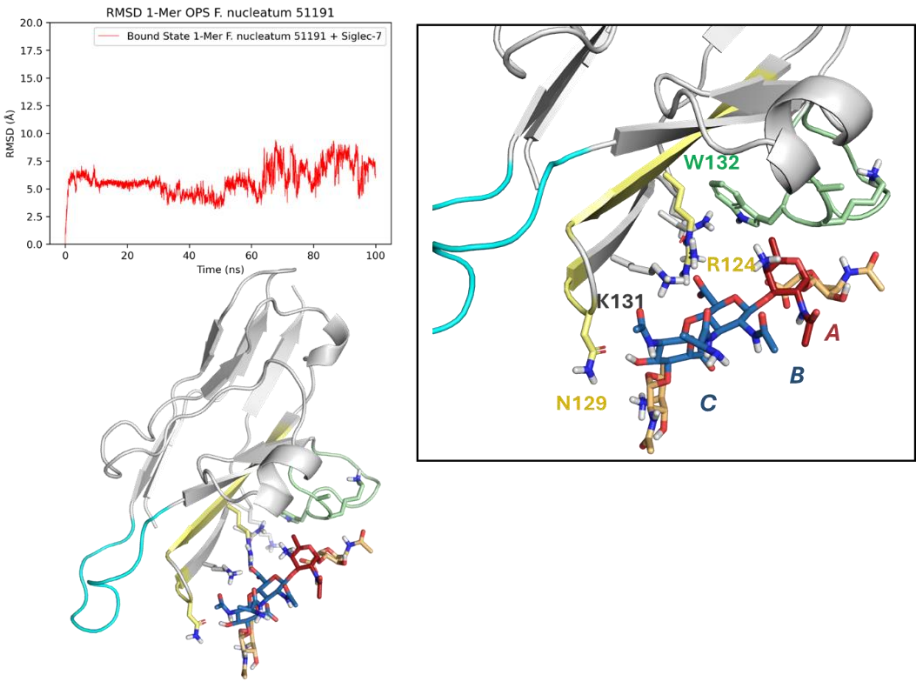

Figure 3.45. RMSD trajectory of 1-Mer monitored along MD simulation in the bound state**. B** residue is the one stablishing a salt-bridge with R124. Different views highlighting the H-bonds monitored by MD simulation were shown. The strands of the protein involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green Ligand 1-Mer OPS is represented following SNFG color-code

Therefore, Docking calculations with the C residue establishing a salt bridge with the key conserved residue Arg124 were performed. The most

representative poses were afterwards submitted to MD simulations in explicit solvent and the interactions between OPS from *F. nucleatum* 51191 and binding pocket residues from Siglec-7 were monitored, as depicted in Figure 3.4.6

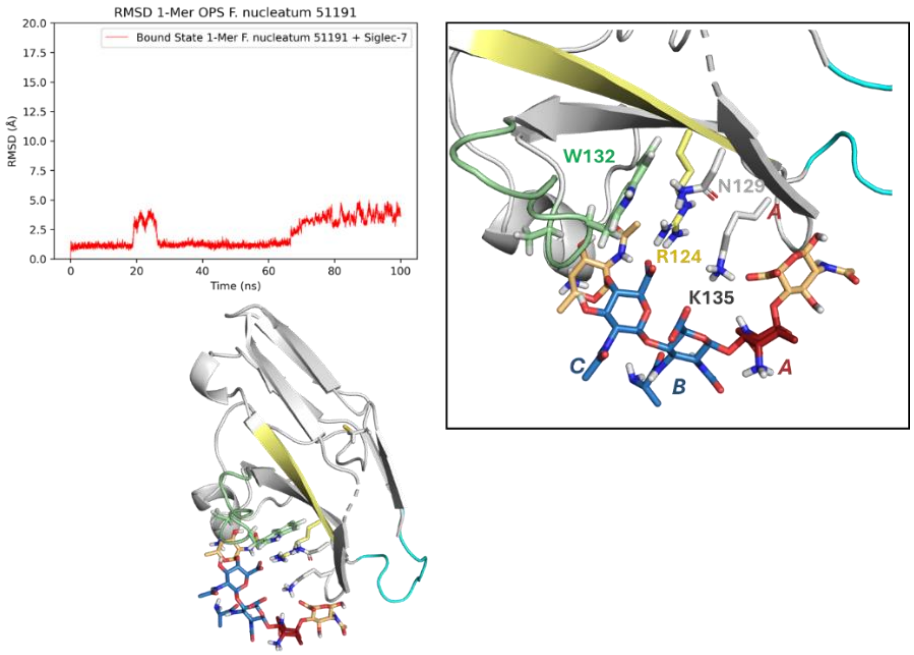

Figure 3.46. RMSD trajectory of 1-Mer monitored along MD simulation in the bound state**. C** residue is stablishing a salt-bridge with R124.Different views highlighting the H-bonds monitored by MD simulation were shown. The strands of the protein involved in the recognition are colored by type: F-β strand in yellow, BC'-loop in cyan and CC'-loop in green Ligand 1-Mer OPS is represented following SNFG color-code.

Analogous interactions were found, but is noteworthy that the RMSD gave lower values with an average of 3A, demonstrating that **C** residue (β-GlcNAcA) from the OPS of *F. nucleatum* 51191 seems to be the main character on the recognition event with Siglec-7, by establishing a key salt-bridge with Arg124, and other polar contacts with conserved residues such Lys135, Trp132 and Asn129, help stabilizing the complex along the simulation.

3.4.4 Discussion

The specific recognition of the OPS from *Fusobacterium nucleatum* ATCC 51191 by Siglec-7 has been described. The finding that Siglec-7 can bind ligands that mimic sialoglycans has been demonstrated by a combination of NMR techniques and computational analysis. Repeating units of *F. nucleatum* 51191 OPS were evaluated for their conformational behavior after parametrization of their specific force fields in order to build the 3D structure. This enabled a final and detailed analysis of their binding interactions with Siglec-7. The computational studies, including MD simulations allowed to describe the selectivity by Siglec-7 between β-GlcNAc-3Ala (**B**) and β-GlcNAcA (**C**) in the complex formation by its RMSD. It demonstrated that Siglec-7 shows selectivity for terminal residues, like **C**, to allow the recognition event happens.

4. Unveiling the recognition pattern of Siglec – 9 and Endogenous Ligands

Siglecs interaction with endogenous sialoglycan ligands—those present on host cells—remains a crucial, yet underexplored aspect in this dissertation. Those are essential for maintenance of immune homeostasis, as they prevent unnecessary immune activation by recognizing "self" sialylation patterns. Through this, Siglecs contribute to the self-tolerance, reducing the risk of autoimmune reactions and ensuring an equilibrated immune response. The inhibitory signaling pathways activated by endogenous ligands, such as those mediated by the immunoreceptor tyrosine-based inhibitory motifs (ITIMs) in some Siglecs, are key to this regulatory function. Within this framework, we aim to describe in the following sections the intricacies and specificities between endogenous sialylalted ligands and Siglec-9.

4.1 The role of Siglec – 9 in binding the Ganglioside GDa-α and its sulphated cognate

4.1.1 Introduction

Sialic acid immunoglobulin-type 9 -Siglec-9, a CD-33 related Siglec, is an immune checkpoint receptor broadly expressed on leukocytes, specifically in monocytes, neutrophils, dendritic cells, and in lower levels, macrophages and natural killers.[147,148] CD-33 related Siglec-9 shows a preference for binding α-2,3 and α-2,6 sialoglycans, specially gangliosides with multiple terminal α-2,3 linkages. Several glycan arrays showed a prevalence for the binding with sialyl Lewis X (SLex, α-Neu5Ac-(2,3)- β-Gal-(1,4)-α-Fuc-(1,3)-β-GlcNAc) and 6-sulphated sialyl Lewis X glycans. Differently from other Siglecs, it also showed a high affinity for non-sialylated ligands, since it was demonstrated the binding between hyaluronic acid and this receptorv (β-GlcNAc- (1-4)-β-GlcA-(1,3))$_n$.[149] The concentration of Siglec-9 is higher in the neonatal period of host

and its ligands expression is upregulated in chronic airway inflammation demonstrated by its binding to airway mucin MUC5B.[150] It has also been found to recognize several sialoglycans expressed by some human cancers due to the hyper-sialylation and hyper-glycosilations on mucins. Within this framework, Siglec-9 is becoming an important immune checkpoint and tumor immunotherapy development in the last years. [151,152]

Highlighting that Siglec-9 shares 84% identity in the coding sequence with Siglec-7, and both share 8 potential N-glycosylation sites – sialic acid binding sites – and their V-set domain shares over 98% similarity in their proteins, it is important to study this receptor together with its "sister" Siglec-7. These two immune checkpoints are potential immune checkpoints to cancer and immune therapy. The ability of Siglec-9 to bind sulphated ligands creates an interesting research field to decipher at a molecular level how this interaction take place, and which are the amino acids on its V-set domain to be in charge of this recognition event, and we aim to offer some insights into this novel receptor.[153]

4.1.2 Computational analysis: free and bound state of GD1a-α and Siglec-9

The crystal structure of Siglec-9 has not been solved yet. For this reason, the AlphaFold prediction for this receptor has been the one used during this section (AlphaFold ID: Q9Y336) with Human Siglec-9. The structure that was obtained by AlphaFold holds a 90% confidence in the model prediction on the V-set domain, so we chose to refine this structure dynamically through MD simulations. (Figure 4.1) This approach allowed the protein structure to arrive to one of their conformations minima by employing simulations on the protein by itself, and then with a well-known ligand, such siallyl-lactosamine (α-Neu5Ac-(2,3)-β-Gal-(1,4)-β-GlcNAc)[154] to check binding site, and possible conformational changes that may occur in the CC-loop and F-β strand of the receptor.

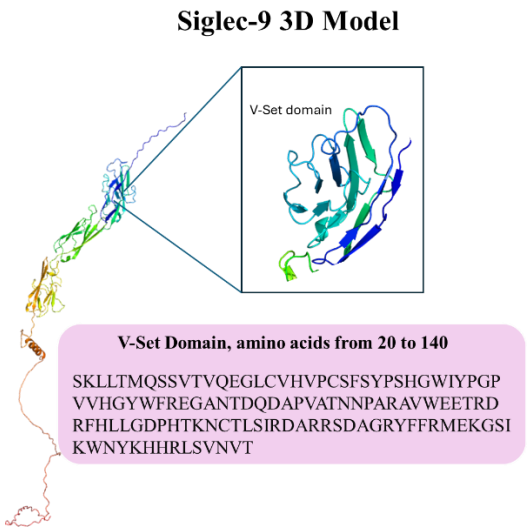

Figure 4.1 3D representation of Siglec-9 and the amino acids contained on the V-set domain from AlphaFold prediction.

The stability of this protein was monitored through RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation). The RMSF allowed to depict which were the amino acids implicated in the conformational changes and the fluctuation in Angstroms that have received during the simulation. Together with RMSD, helped to select the most stable structure of Siglec-9 during the calculation timescale. Surprisingly, no big conformational changes were observed during the calculations, except for the BC and CC-loop, which was expected due to its high flexibility, as shown in Figure 4.2.

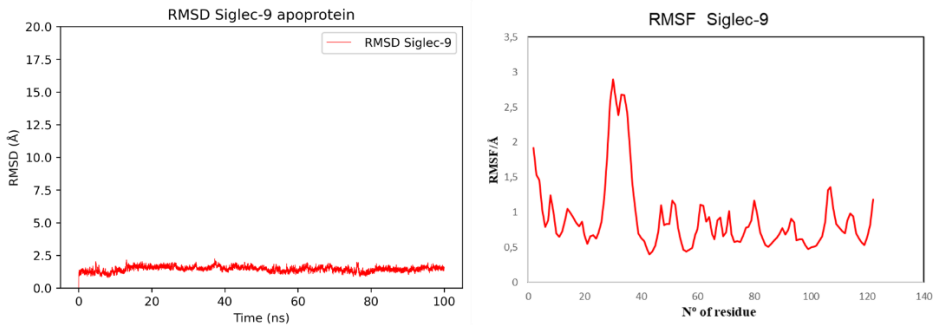

Figure 4.2. RMSD and RMSF graphics from the MD simulation of Siglec-9 protein

After obtaining a reliable result on the stability of Siglec-9 protein, its binding site was checked with a known ligand: siallyllactosamine (3'SLN), α-Neu5Ac-(2,3)-β-Gal-(1,4)-β-GlcNAc, by using it with a template of Siglec-7 and Siglec-10, to build the Siglec-9-3'SLN 3D model. First, the complex was refined and docking calculations of the 3'SLN inside the Siglec-9's binding site were performed with AutoDock4.2. The chosen complexes from Docking were submitted to MD simulations in explicit solvent and the interactions between sialic acids the binding pocket residues were monitored.

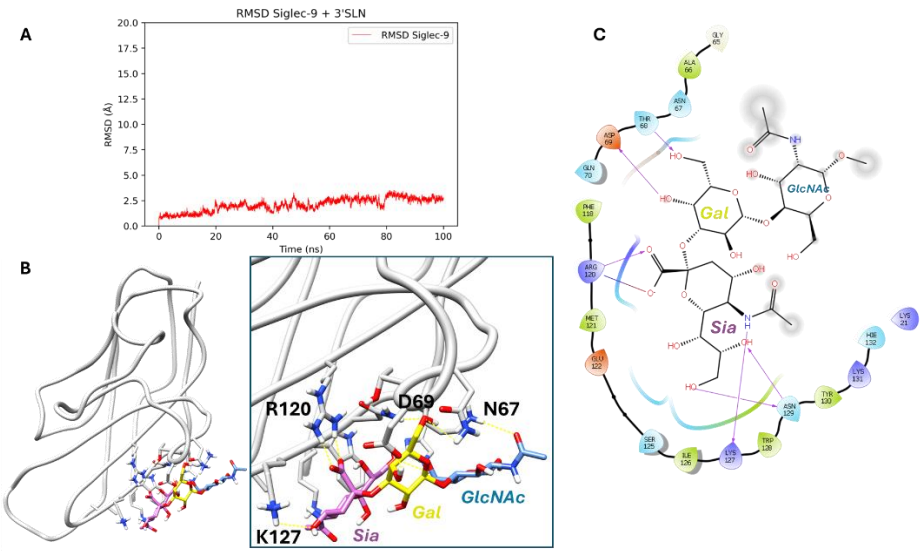

Figure 4.3. RMSD trajectory of 3'SLN monitored along MD simulation in the bound state**.** Sialic acid residue is stablishing a salt-bridge with R124. Different views highlighting the H-bonds monitored by MD simulation were shown. The protein involved in the recognition are colored grey cartoon. Ligand 3'SLN is represented following SNFG color-code.

By comparing the results of the amino acids implicated in the binding pocket with those published, we concluded we obtained a reliable model of Siglec-9. Located on the F-β-strand, conserved key residue Arg120 (Figure 4.3) is the main character on the recognition event, by salt-bridging the carboxylate of the sialic acid on the 3'SLN. Other polar contacts, such the ones implicating Asp69, Asn67, Lys127 and Asn129 helped to accommodate the α-2,3 sialylated glycan

inside the binding pocket. Within this information, we now have a reliable structure for Siglec-9 to be studied among other ligands. GD1a ganglioside family, in particular GD1a-α, is a complex glycosphingolipid that is found on the outer leaflet of cell membranes, specifically in neurons and other cell types in the nervous system. Gangliosides are characterized by a ceramide backbone that can be linked to one or even more sialic acid residues. GD1a-α contains two branched sialic acid residues contained in a carbohydrate chain.

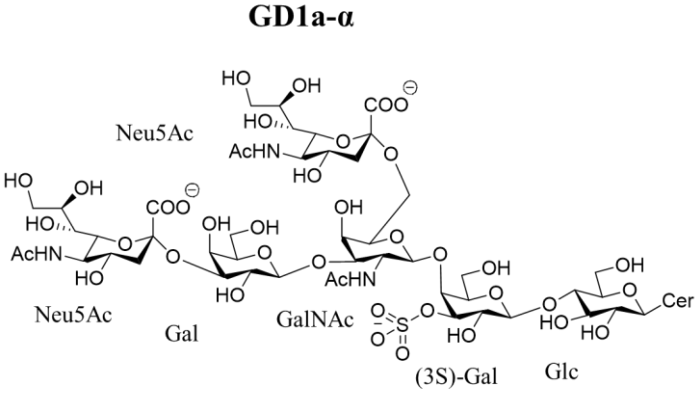

Figure 4.4 Schematic representation of the chemical structure of sulphated-GD1a-a.

It is involved in cell recognition, modulating synaptic transmission. The α-2,3 and α-2,6 linkages, and the abundance of sulphated, as shown in Figure 4.4 Gd1a-α, makes this ganglioside a potential ligand for binding Siglec-9. The complex formation between this receptor and GD1a-α can trigger immune activation, relating this with certain cancers and inflammation by the upregulation with Siglec-9. To decipher the possible engagement between this ligand and Siglec-9, a computational approach has been performed. 3D structure of GD1a-α, in its free and bound state was first performed. Through molecular mechanics simulation, the adiabatic energy maps were calculated. (Figure 4.5) Since ψ, as expected, remained constant with one possible value at -180º Fig 4.1.2.5, the three staggered conformations defined by the ω torsion of α-2,6 linkage: *gt*, *gg* and *tg* were evaluated Several reports have demonstrated

that Neu5Ac-Gal linkages rarely populate the *gg* conformation. So, for the Neu5Ac-Gal linkages the main rotamers expected will be *gt* and *tg*. In the case with galactose, it has been shown in previous studies that the *gt* is the preferred one. In α-2,3 linkages, a minor flexibility was present, with only two main conformers: -*g* and *t*, φ at -60º or 100º and a ψ at 0º, respectively.

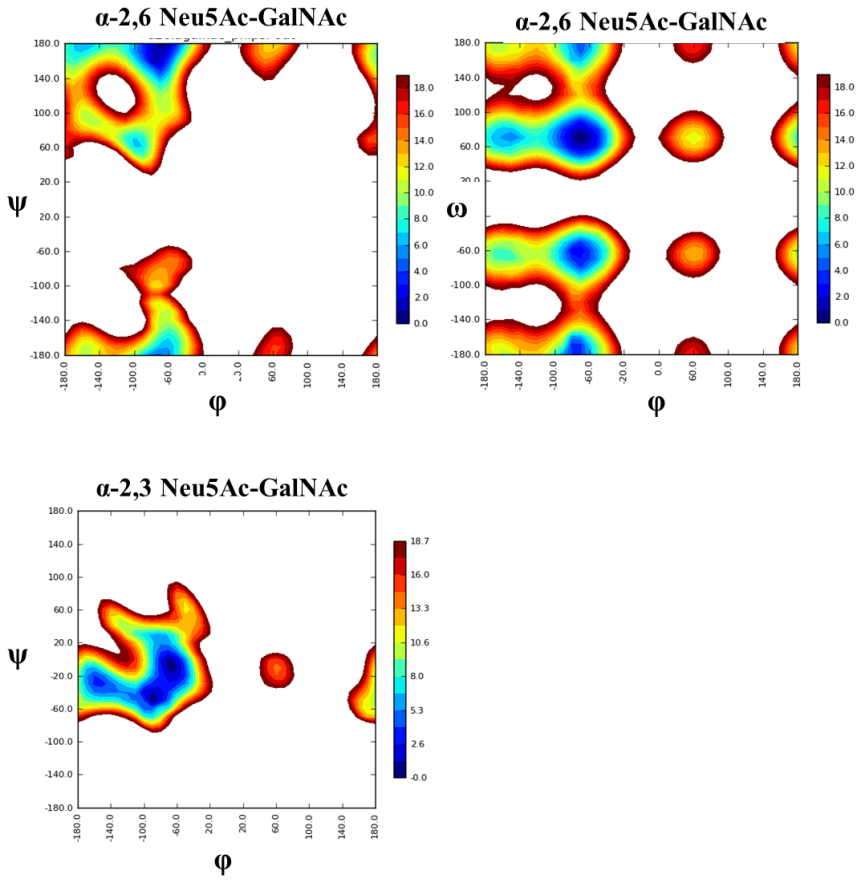

Figure 4.5. Adiabatic energy maps of the two different glycosidic linkages present in the GD1a-α sialyl linkages illustrating the φ, ψ and ω torsions of the Neu5Ac-GalNAc linkage. Below, the representation of the three staggered rotamers from the torsional angles described for the linkage α-Neu5Ac-(2,6)-α-GalNAc: φ, ψ and ω. Where the ω have as main values -60, 60º and 180º namely *gg, gt* and *tg* respectively around the C5-C6 bond; and α-Neu5Ac-(2,3)-α-GalNAc, φ, with main values of -60º or -100º, and ψ with a 0º torsion value, respectively.

Docking calculations with GD1a-α and Siglec-9 were performed, and the most representative complexes were submitted to MD simulations. Due to the presence of two sialic acids, both were considered for binding to Siglec-9.

First, the case of α-2,3 linkage facing Siglec-9 was submitted to MD simulations in explicit solvent. Through the analysis of the results, we examined the stability of GD1a-α by sampling the torsion values for each linkage. We obtained a preferred conformer for the α-2,3 on Neu5Ac-Gal, which was the one that held φ =-60º, *-g* and a value of ψ of 0º (Figure 4.6).[155] Instead, the α-2,6 gave φ value of -60º and ψ of 180º, and different rotamers populated the ω torsion, being the most representative the *tg* and the *gt*, with values of 180º and +60º, respectively. The stability of the complex was also monitored through RMSD (Root Mean Square Deviation), as shown in the figure below, showing high stability of the protein with no conformational changes observed during the timescale of the simulation, corroborating a reliable 3D complex of GD1a-α and Siglec-9.

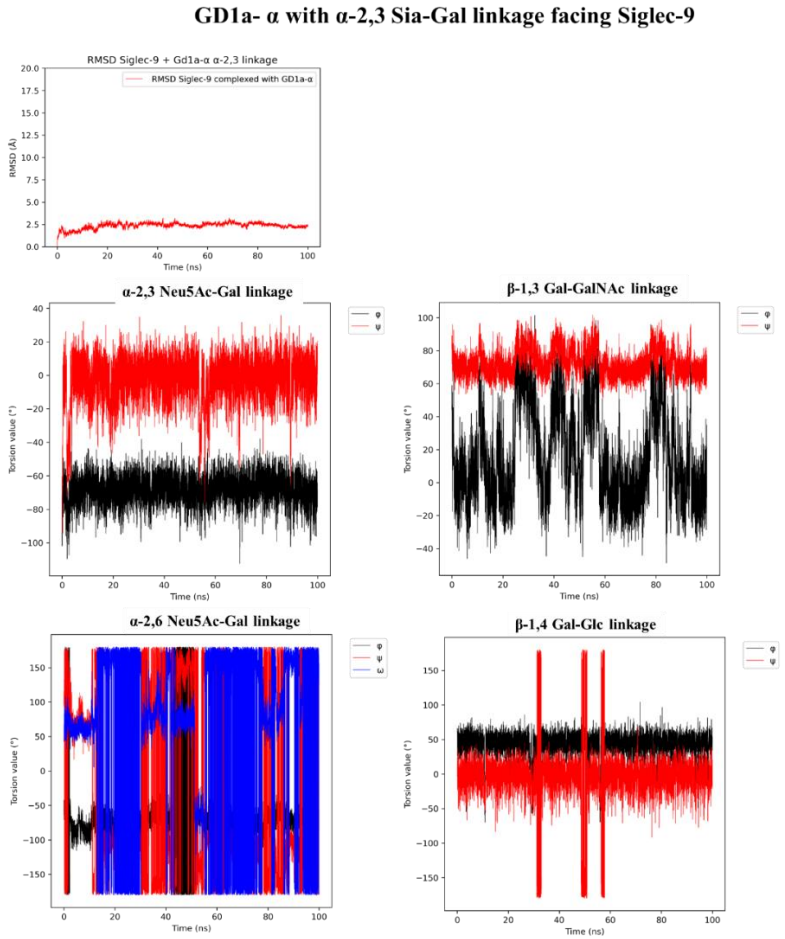

Figure 4.6 The behaviour of φ, ψ and even ω during MD simulations of the different linkages of GD1a-α are reported. Stability of the ligand is represented by the RMSD, plotted in red and the glycosidic torsions are sampled and divided into the different linkages, as it corresponds.

The interactions taking place in the binding pocket of Siglec-9 were also monitored, as depicted in Figure 4.7. Through this computational analysis we observed that the sialic acid attached to the galactose with the α-2,3 linkage is successfully recognized. The analysis of the key contacts mainly involved electrostatic and hydrogen bonds interactions, most of them retained along the MD simulation. In detail, the highly conserved Arg120 established a stable salt bridge with the carboxylate group of the internal Neu5Ac via its guanidinium

group. The network interactions were established with conserved residues conserved on the β-strand, F, surrounding, like the canonical Arg120, the G-βstrand, in which typical amino acids of Siglec-9, like Lys135 and Asn129 are located, and also involved interactions with the CC-loop such as hydrogen bonds with Asp69 and Ala67.

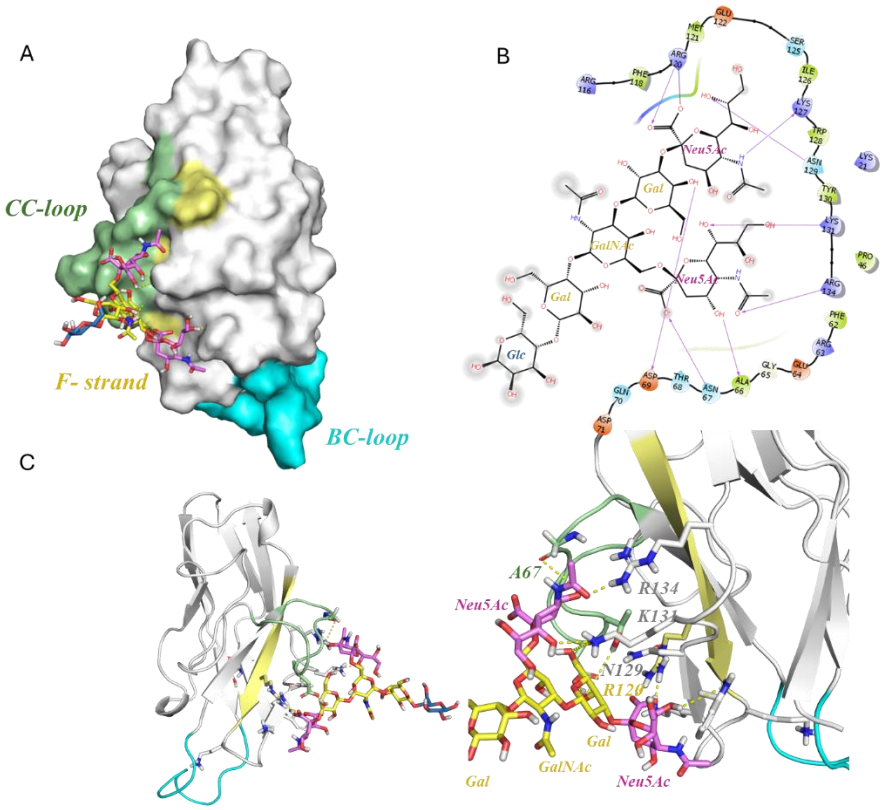

Figure 4.7 A. 3D model of the Siglec-9 and GD1a-α in complex with the protein in surfaced colored by loop and strand. CC-loop in green, F-β strand in paleyellow and BC loop in cyan. **B**. 2D plot of the interactions occurring at the Siglec-9 and GD1a-α from lgand interface. **C.** Different views highlighting the H-bonds monitored by MD simulation were shown. Ligand Gd1a-α is represented following SNFG color-code.

The description of the ability to bind sulphated ligands by Siglec-9 remains crucial in the underlying of the molecular mechanisms taking place to improve the binding. GD1a-α has been found also with a sulphate moiety in the GalNAc

in the 3'OH position. MD simulations on the 3D model of the complex between Siglec-9 and sulphated GD1a-a were submitted in explicit solvent. Through the analysis of key contacts, no big differences were observed as shown in Figure 4.8, only the sulphate group is anchored through the Lys135, helping to accommodate the ligand in the binding site of the protein. Within this, more studies will be conducted within this project through experimental techniques that can complement these data.

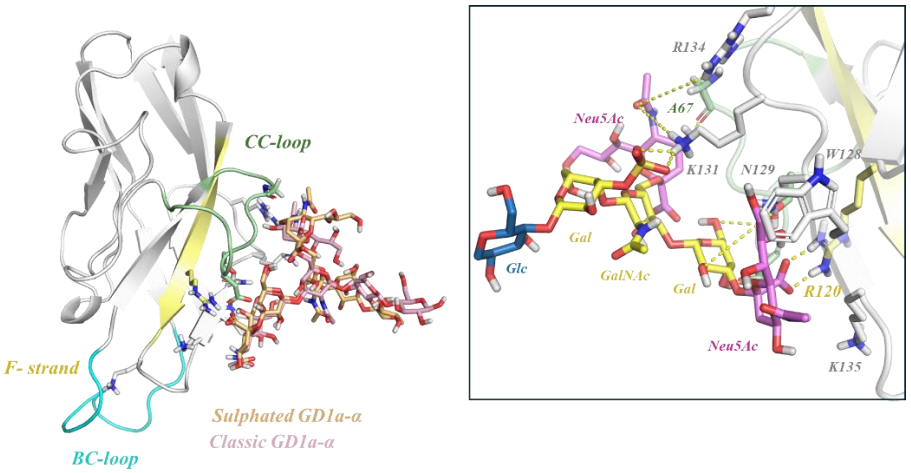

Figure 4.8. 3D model comparison of the Siglec-9 - sulphated GD1a-α and the classic GD1a-α in complex with the protein in surfaced colored by loop and strand. CC-loop in green, F-β strand in paleyellow and BC loop in cyan. Classic GD1a-α is colored in lightpink sticks, and uslphated GD1a-α in light orange sticks. On the right different views highlighting the H-bonds monitored by MD simulation were shown. Ligand sulphated Gd1a-α is represented following SNFG color-code.

4.1.3 Discussion

The 3D model of Siglec-9 was defined with AlphaFold. The refining with well-known recognized ligands like 3'SLN has been also shown. Within this, the specific recognition of the GD1a-α classic and sulphated ganglioside by Siglec-9 has been described. Siglec-9 can bind sulphated ligands has been demonstrated by computational analysis. The computational studies, including

MD simulations allowed to describe the selectivity by Siglec-9 with α-2,3 and α-2,6 linkages by the monitorization of its RMSD and its torsion values within 3'SLN and GD1a-α ganglioside host ligands.

5. Siglec – 10 and Synthetic Glycomimetics –a cancer immunotherapy target

5.1 Introduction

Siglec-10 belongs a to inhibitory Siglecs'family. Is mainly expressed on dendritic cells, macrophages and B cells.[156] Structurally, it is constituted by five extracellular domains and three tyrosine-based motifs in its cytoplasmic tails, being also a CD-33 related Siglecs. It is able to dampen autoimmune responses upon recognition of α-2,6 and α-2,3 sialoglycans, and only modulates these responses to DAMPS, suggesting its exclusive role in mediating sensing of infection vs tissue injuries in the immune system through dendritic cells.[157] Interaction between CD-24, a DAMPs, and Siglec-10 can dampen the activation of Toll Like Receptor 4 (TLR4), leading to a downregulation on the inflammatory response.[158,159] Siglec-10 is associated to several patho-physiological processeslike ovarian and breast cancers, suggesting is the key mechanism through which cells promote tumor evasion. Siglec-10 has also been associated with autoimmune diseases like Systemic Lupus Erythematosus (SLE); neurodegenerative diseases like Alzheimer's, where Siglec-10 is highly expressed on microglia and inflammatory processe like sepsis.

As part of the GLYTUNES project, a the crystal structure of Siglec-10 has been solved in CIC-Biogune in Jesús Jiménez Barbero group, where they experimented problems due to the difficulty on the crystallization, as the CC-loop is lacking on the solved structure. Conversely, ICENI diagnostics have synthesized some glycomimetics that could work as inhibitors for Siglec-10. Within this framework, collaborative research was established and possible ligands against this receptor were evaluated with molecular docking.

5.2 Modelling Siglec-10 *CC*-loop

Provided by the CIC-BIOGUNE, the 3D model of the V-set domain of Siglec-10 was incomplete on the CC-loop. Due to this, the set of aminoacids lacking on the structure were taking from UNIPROT (ID: Q96LC7) where these

residues were taken and submitted to structural refinement. Using as template Siglec-7, and a model built in our group by R. E Forgione et al,[67] the 3D model of the V-set of Siglec-10 was constructed. Checking for conformational changes, this PDB was submitted to several MD simulations on the protein uncomplexed and complexed with well-known ligands, such the 3' siallyllactosamine, 3'SLN. Stability of the apoprotein was monitored through RMSD and comparison with the complexed V-set domain was also performed, giving as result a RMSD between structures of 1,104 Å, and an average on the timescale of 1,5Å as shown in Figure 5.1. Thus, a reliable 3D model to use for the building of 3D complex for Docking calculations.

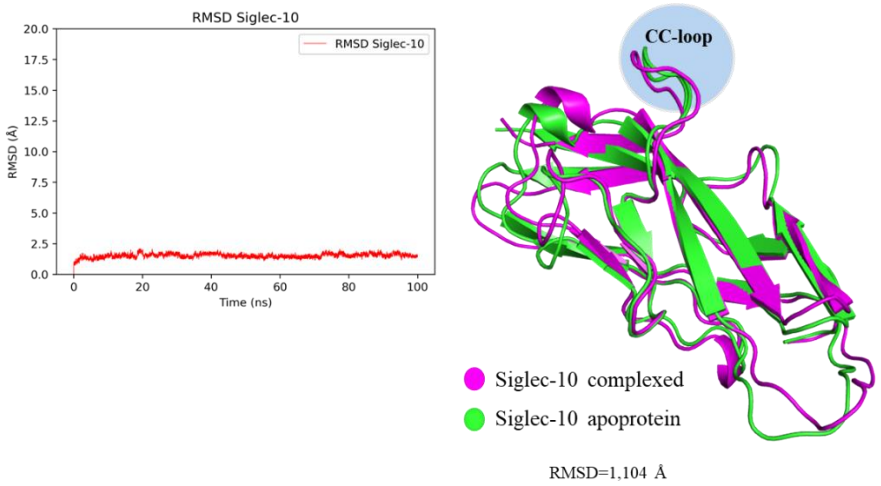

Figure 5.1 Stability of the V-set domain of Siglec-10 is plotted in red through its RMSD. Comparison between the complexed and un-complexed protein shows great stability and no particular conformational changes observed.

5.3 Molecular Docking on a glycomimetics library targeting Siglec-10

Glycomimetics, being fully synthetic or semi-synthetic molecules, have emerged as promising therapeutic agents in the field of immunology, targeting infectious diseases, autoimmune disorders and cancer. It has been shown a particular interest on Siglecs. The specific binding between these immune

receptors and glycomimetics aims to modulate the receptor's activity by restoring in this way the proper immune function.[160]

Some semi-synthetic glycomimetics have been suggested through the collaboration with ICENI Diagnostics, as part of the project of the ITN. (Figure 5.2) Three glycomimetics with a biphenyl sulphonate substitute on the 9' OH position of the glycerol moiety of a sialic acid residue were proposed. This biphenyl moiety is planar and inside the binding pocket may potentially interact with hydrophobic residues on the F and G- strand of Siglec-10. Conversely, glycomimetics with a carbamate linked to a triazol and an adamantane counteraprts, also connected on the 9'th OH position of the glycerol moiety of the terminal sialic acid were proposed. .

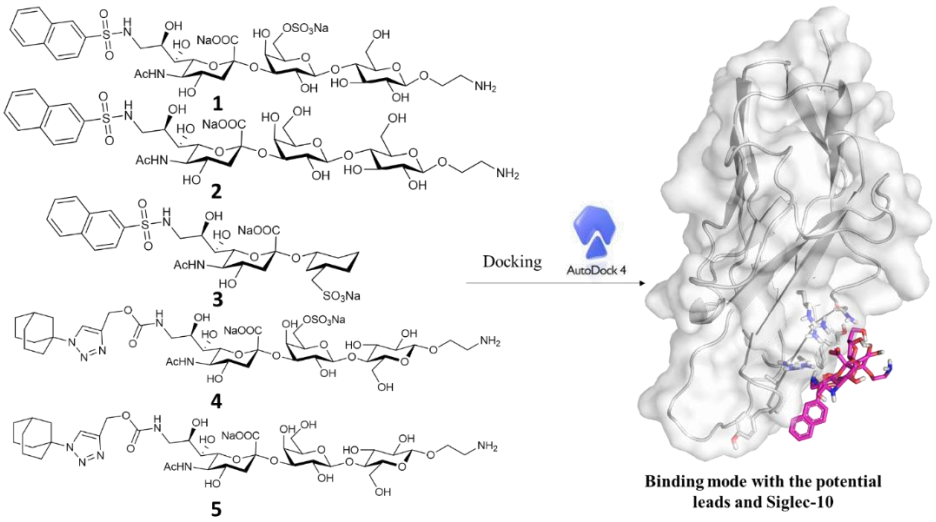

Figure 5.2. Structures of the ligands docked into the binding pocket of Siglec-10. Ligands 1, 2, and 3 feature a biphenyl-sulfonate group at the 9' position of the glycerol chain of sialic acid, while ligands 4 and 5 have an amide linkage at the same position, connecting a carbamate and triazole modified with an adamantane moiety. All docking studies were conducted using AutoDock 4.2, with α-2,3' sialyllactosamine serving as the template.

After the minimization and refinement of the 3D structure of Siglec-10, with Siglec-7 (PDB:2HRL) as template for the positioning of the different ligands,

several molecular Dockings were performed. Analysing the most representative binding poses obtained with Autodock, we observed that the carboxylate group of sialomimetics 1 and 4, established stable salt-bridges with the guanidinium moiety of Arg119, the conserved key residue for the recognition process with Siglec-10. The sulphate group present on the 6'position of the lactose (Gal-GlcNAc) core in 1 and 4, played a key role in establishing of novel hydrogen bonds contacts with other conserved residues from Siglec 10, such as Lys54 and Tyr128. Also, the sulphonate moiety directly linked to the biphenyl functional group present on ligands 1, 2 (Figure 5.3) and 3 (Figure 5.4), established novel contacts with conserved Asn129 residue. Interestingly, ligands 2 and 5 were accommodated in the binding pocket but the salt-bridge was lacking, excluding them primarily as possible candidates for Siglec-10.

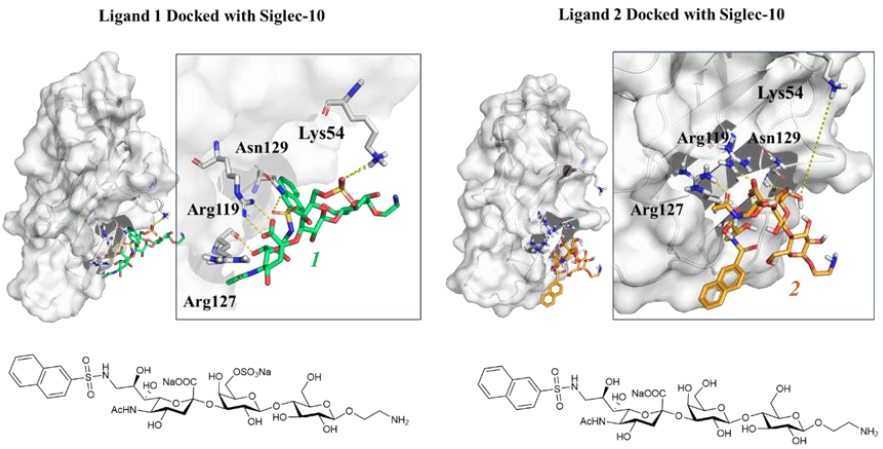

Figure 5.3. Binding Interactions Between Siglec-10 and Ligands 1 and 2: A Demonstration of Pattern Recognition. Ligand 1, shown on the left as green sticks, forms a salt bridge with the canonical Arg119 and and hydrogen bonds with Asn129 through the sulfonate moiety of the biphenyl from Sialic acid. It also establishes another hydrogen bond between the 6'-sulfate group of the galactose moiety and the Lys54 present in the CC-loop. In contrast, ligand 2, represented on the right as orange sticks, lacks the 6'-sulfate group on the galactose residue, resulting in poorer accommodation

of the ligand in the binding site. Consequently, Lys54 and Asn129 are not involved in the recognition of ligand 2.

Additionally, the sulphonate moiety directly linked to the biphenyl functional group present on ligand 1 and 5 seems to be a necessary group, since it can establish novel contacts with conserved Asn129 and additional Arg80. Ligand 3 is an interesting case, as it demonstrated the highest movement inside the binding pocket of siglec-10, flipping the molecule completely where its biphenyl faces towards the CC-loop, losing completely the salt bridge between the carboxylic moiety of the sialic acid and the guanidinium group of conserved Arg119.

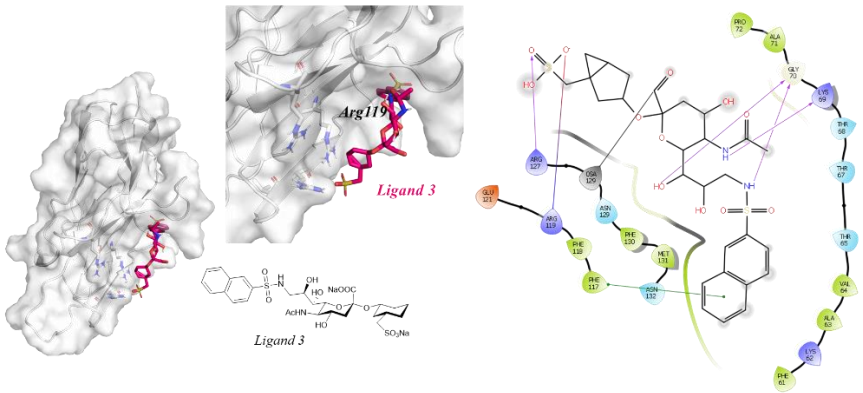

Figure 5.4. Binding interactions between lead 3 and Siglec-10 after Docking calculations. The ligand has given a 180º flip inside the binding pocket losing the salt-bridge between the carboxylate of sialic acid and Arg119, positioning its byphenyl towards the CC-loop and the sulphate to the F-β and G-β strands where conserved Arginines are present.

The lack of the sulphate group on the lactose residue 2 and 5 leads to the destabilization in the recognition event with Siglec-10. Ligand 4 gave interesting results, with a bifurcated salt-bridge with two arginine residues. On the one hand, one oxygen established a hydrogen bond with Arg127, and on the other with canonical Arg119, anchoring in this way the ligand in the binding site. Adamantane established a slight hydrophobic interaction with Val127 and

the sulphate on the 6'position of the galactose established a novel contact with Arg80, as shown in the Figure 5.5. Moreover, even though ligands 4 and 5 are recognized, the adamantane and a triazol moieties do not accommodate as well as the biphenyl moiety present in ligands 1,2,3 in the binding site of Siglec-10.

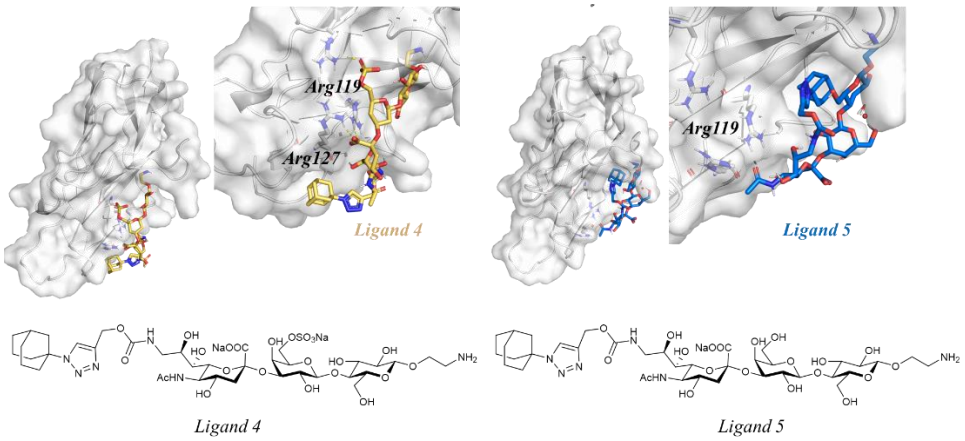

Figure 5.5. Binding interactions between lead 4 and 5 with Siglec-10 after Docking calculations. Ligand 4 established bifurcated salt-bridge between the carboxylate of sialic acid, Arg119 and Arg127 establishing nice anchorage on the binding site of siglec-10. Conversely on the right, Ligand 5, lacking a sulphate on the galactose receives a flip of 180º that pushes the adamantane towards CC-loop, losing the salt-bridge and leading a bad recognition.

5.4 Discussion

With the previous 3D Siglec-10 structure, Molecular Docking with several glycomimetics have been performed. Within this, the specific recognition of these semi-synthetic leads has been described, as demonstrated by computational analysis. The results showed that the preferred leads are those with a sulphate on the galactose residue, that seems to be key in the development of glycomimetics for Siglec-10. Among these, those with a biphenyl group established a better recognition in the binding site due to less steric hinderance with respect to the adamantane moiety.

General conclusions:

Glycans are essential regulators of living organisms due to their biological roles, including structural and modulatory functionalities, molecular mimicry and specific recognition by glycan-binding proteins. In particular, sialic acid, located on mammalian cell surfaces, considered as an effective marker of "self" ligand by the host, acting as key regulator of the immune system by its interaction with some particular GBPs, Siglecs, as extensively described during this PhD dissertation gained importance during the last decades. From only being an endogenous, "self" ligand, sialic acid turned into a target for molecular mimicry performed by bacteria and some cancer cells. This information constituted a step on the deep comprehension of Siglec relevance in health and diseases. These "non-self", exogenous interactions cause serious diseases between allergies, autoimmuninity, pathogenic diseases and cancer. For this reason, Siglecs are nowadays considered as checkpoints in human diseases, becoming a target for the design of therapeutic agents for the treatment of these relevant diseases.

With this context this thesis work aimed to offer new insights into understanding the Siglec-sialoglycan molecular binding processes within different biological events that have been provided setting the basis for the design of novel therapeutic and diagnostic treatments. In this framework, a crucial step is the detailed description of the molecular interaction process, here achieved by an ensemble of advanced NMR techniques, biophysical methods, and more extensively in this dissertation, through computational studies. This approach allowed to delve different Siglecs in different contexts. Siglec-7 in the context of the recognition events with pathogenic exogenous sialylated ligands such as those CPS produced by *Neisseria meningitidis* or the sialic acid mimicry of the LPS from *Fusobacterium nucleatum*. Then, the description of Siglec-9 aiming to set the basis for the understanding of the downregulation of several diseases such as Guillen-barre syndrome, the recognition process

between this receptor and host endogenous ligand was described. Lastly for inhibitory immune checkpoints, we deepened the context of Siglec-10 and some glycomimetics that are potential leads for the drug design was deepened through computational techniques.

We here demonstrated that nowadays computational tools are essential not only comeplementing experimental data, but also for laying a basis for research by providing novel insights. By this approach, applying computational analysis in the beginning of a multidisciplinary investigation, we showed that many experimental procedures can be simplified going to direct experiments from the acquired data. This results in time and resource saving without compromising the depth and accuracy of new experimental findings and underscores the growing and importance of computational approaches in the modern molecular biology research field.

Chapter IV:

Side Projects

IV: Side Projects:

6. Dimerization of the Toll-Like Receptor 2/1 and Toll-Like Receptor 1/6

6.1 Introduction

Toll-like receptors (TLRs) constitute a family of pattern recognition receptors (PRRs) that have a critical role in the innate immune system. TLRs are constituted by 12 members, identified to date in mammals, and they can recognize Gram-positive and Gram-negative bacteria. Through the interaction with the respective ligands, they dampen signalling molecules to the intracellular domain, dimerizing with the complement factor MyD58, activating the secretion of pro-inflammatory cytokines. [161] Each TLR seems to recognize a different type of pathogen; TLR 4[162] is in charge of the recognition with LPS from Gram-negative bacteria, TLR 5 the bacterial flagellin and so on. Conversely, TLR 2[163] seems to be more heterogeneous in its binding with different pathogens, such as peptidoglycan from bacteria, lipoteichoic acid, lipoporoteins, among others.[164] It has been demonstrated that this ability to bind diverse ligands is lead from the unique ability to heterodimerize with TLRs 1[165] and TLR s 6[166]. This heteromer has mainly being studied with bacterial lipoproteins. Dimer TLR 2/1 can recognize diactylated lipoproteins, and TLR 2/6 triacylated ones. The heterodimerization and the 3D behavior of the full-length model is not yet known. As part of my PhD period abroad, in the Prof. Sonsoles Martín Santamaría group, at the CSIC – Madrid, I enrolled to depict the heterodimerization of the TLR2/1 and TLR2/6 on its extracellular domain and transmembrane domain embebbed in a realistic membrane by the employing of novel computational techniques.

Within this framework, it is aimed to describe at a molecular level the extracellular domain of both dimers TLR2/1 and TLR2/6 that may influence the arrangement of the transmembrane domain, which also dimerizes. As a last

part of the project is the intracellular domain within all its possibilities, focusing on the development of the full-length 3D model of these heterodimers. (Figure 6.1)

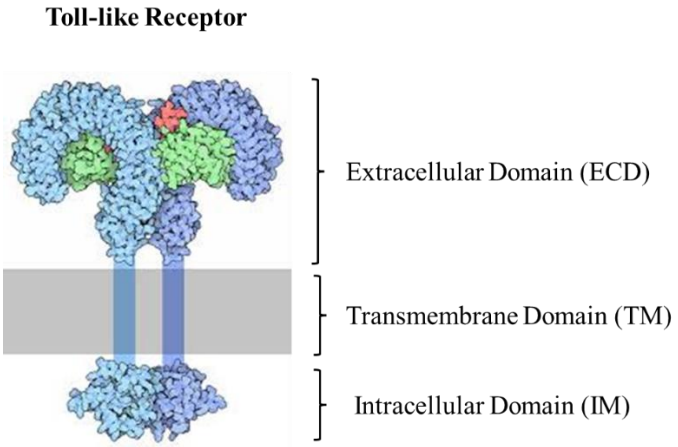

Figure 6.1 Schematic 3D representation of the three domains componing a TLR. On the outer portion of the membrane, the extracellular domain (ECD), embebbed in a liquid-ordered or liquid-disorered membrane the transmembrane domain (TM), and lastly the intracellular domain, responsible for the pro-inflammatory cascade activator (IM).

6.2 Computational studies on the Extracellular Domains of TLR2/1 and TLR2/6 heterodimers.

ECD of the heterodimer TLR2/TLR1 apoprotein and complexed with PAM3CSK4 lipopeptide.

From a previous protein-protein Docking calculation, the apoprotein dimer TLR2/1 and complexed with PAM3CSK4 counterparts were submitted to different MD simulations. The 3D structure complex has been taken from the published X-ray structure 2Z7X, missing terminal region LCRRT has been modelled to complete the amino acidic sequence. The stability of the proteins has been monitored through RMSD and RMSF, as decpited in Figure 6.2., and

interactions at the interface of the dimerization and upon ligand binding have been analysed.

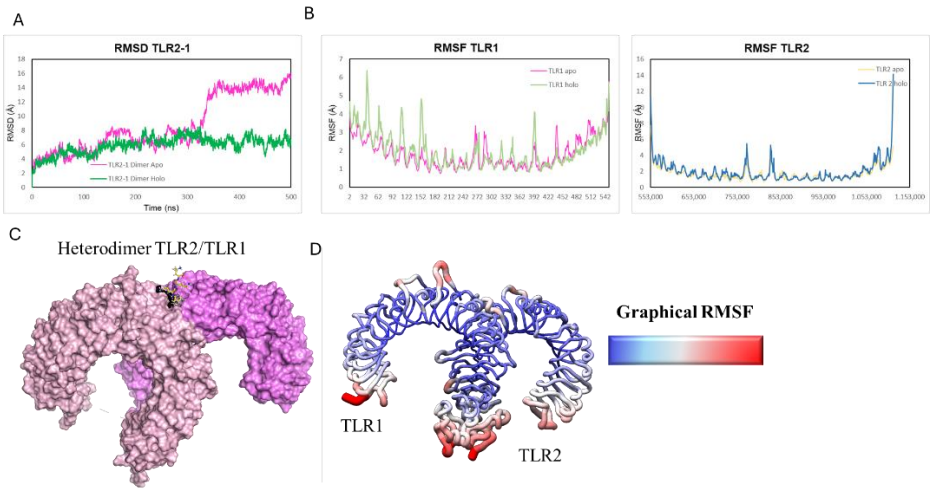

Figure 6.2.A. RMSD monitored through the MD simulation, as plotted in green for the heterodimer TLR2/1 in green complexed with PAM_3CSK_4 lipopeptide and depicted in pink for the TLR2/1 ECD uncomplexed. B. RMSF along the timescale of the simulation of both TLR1 and TLR2 complexed and uncomplexed. C. Surface representation of the TLR2/1 heterodimer after 500 ns of simulation. D. Graphical RMSF performed through Chimera. Red indicates high flexibility, white 1,5 A and blue stable non-flexible counterparts.

A reliable ECD for the TLR-2/1 heterodimer was obtained, only showing some conformational changes on the TLR1. This is due to the dimerization, that provokes an accommodation of the amino acids from 90 to 150 to interface the of TLR2 dimerization site. TLR2 contains a binding site able to hold a diacylated ligand, conversely to the only acyl chain binding site from TLR1, meaning that it can interact with tri-acylated lipopeptides like, PAM_3CSK_4, the natural ligand for this heterodimer. After the submission of this complex to MD simulations, the contacts at the interface of both proteins and among the ligand were monitored. In this section the only interactions that will be analysed are those at the ligand-protein interface. From the TLR1 point of view, hydrophobic contacts are expected, involving residues like Phe306, Val332, Phe330 or

Gln293 holding the hydrophobic counterparts of the ligand, and Asp308, Tyr304 interacting with carbonyl and NH from PAM_3CSK_4 (Figure 6.3). Instead, from the TLR2 point of view, the diacyl chain interacted with hydrophobic residues like Thr862, Phe854, Phe881 and Leu882, as reported by Kang et al.[170] This ligand showed an important role in maintaining the interfaces between both dimers interacting between each-other, provocking a conformational change where the Leucine rich-domain LCRRT from both TLRs interacts, keeping the complex in place.

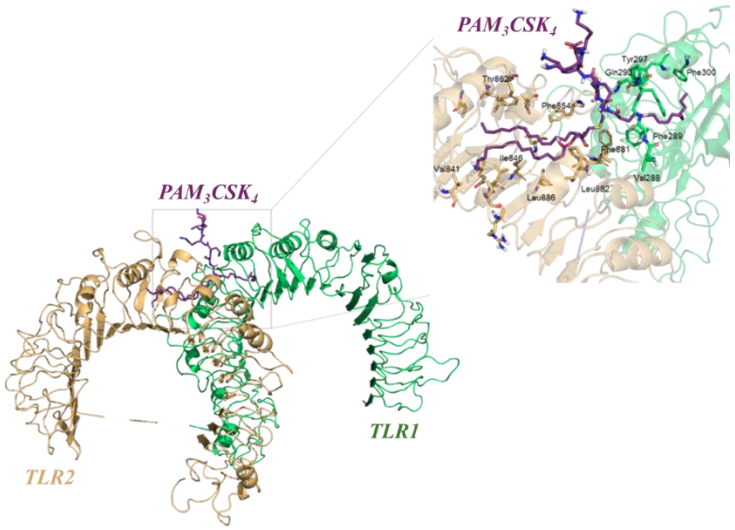

Figure 6.3 Details of the average interactions between ligand PAM_3CSK_4 (purple) and TLR2/TLR1 dimer. (PDB 2Z7X). Represented in cartoon, TLR2 is represented in brightorange and TLR1 in green. Ligand is represented as purple sticks.

ECD of the heterodimer TLR2/TLR6 apoprotein and complexed with PAM3CSK4 lipopeptide

Analogously as the previous heterodimer, TLR-2/6 was submitted to several MD simulations and the stability of the heterodimerization between the TLR2 and TLR6 was monitored through RMSD and RMSF along the timescale of the simulation, as shown in Figure 6.4

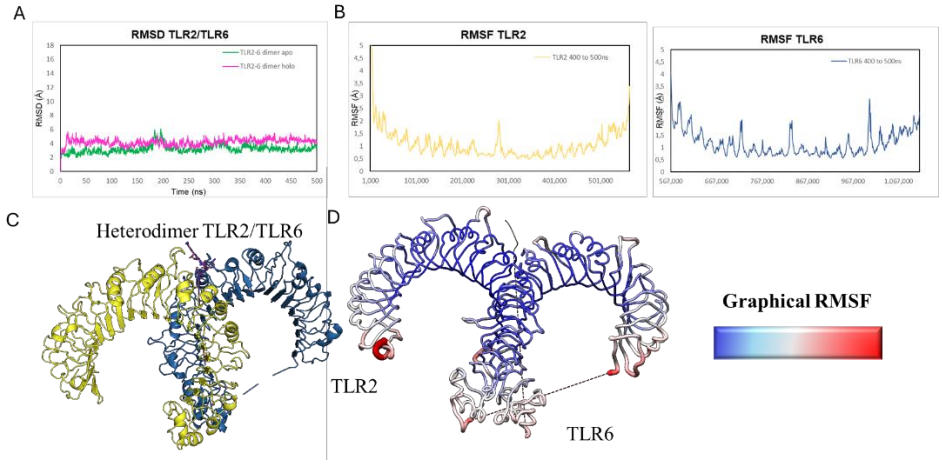

Figure 6.4 A. RMSD monitored through the MD simulation, as plotted in green for the heterodimer TLR2/2 in green uncomplexed and in pink complexed with PAM_2CSK_4 lipopeptide. B. RMSF along the timescale of the simulation of both TLR2 and TLR6 complexed. C. Cartoon representation of the TLR2/6 heterodimer after 500 ns of simulation, TLR2 is represented in yellow and TLR6 in blue. D. Graphical RMSF performed through Chimera. Red indicates high flexibility, white 1,5 A and blue stable non-flexible counterparts.

Stability of this TLR-2/6 heterodimerization was even higher than the TLR-2/1, as almost non-conformational changes are present. Observing how it behaves on the timescale of the simulation, it is noteworthy to say that it remains in the same position as "a rock". As for the interactions, in this case the TLR6 does not contain a hydrophobic pocket for one of the acyl-chains of the PAM_2CSK_4, indicating that almost all the contacts will be held on the TLR2. However, in contrast of the the ligand interacts with TLR2 and less with TLR6, the stability of the dimer depends almost exclusively on the interaction happening at the interfaces of both proteins; specifically, on the LCRRT domain.

With this information, the extracellular domains have been built and submitted to many computational analyses to check stability of both ligand-protein and protein-protein interactions, obtaining a realiable 3D complex.

6.3 Computational studies on the Transmembrane Domains of TLR2/1 heterodimer.

As far as it goes, many of the transmembrane proteins that are published on the PDB lack the TM domain, since they coexist with the lipidic bilayers. These lipidic bilayers are a challenging task for the obtaining of a pseudocrystal, due to its flexibility and low solubitility, making it difficult to achieve experimentally. This section will assess the construction of a membrane bilayer with computational techniques, by the employment of CHARMM-GUI since it covers most lipid types. It is important to develop a membrane with the proper dimensions and characteristics for the protein that is going to be studied embebbed on it. This can determine the different dimerizations on the transmembrane domain and the consequent disposition of the intracellular counterparts, which is the key biological point on the whole TLR. (Figure 6.5)

With this, the aim is to achieve a reliable 3D model on the TM of TLR-2/1 dimer. Apart from the TM, TLRs possess a juxtamembrane domain (JM) that address flexibility to that α-helix, creating a kick-elbow on the dimers that allow the best disposition of the intracellular domain. The different variables existing for this JM + TM, allowed to create a chart with all the possibilities on dimerization between these domains from TLR-2/1. The average 3D structure of TM+JM is published and performed through NMR with PDB code: 8AR0. Stability was monitored on the formed dimer, on the TM by itself and on the JM to select the best possible model.

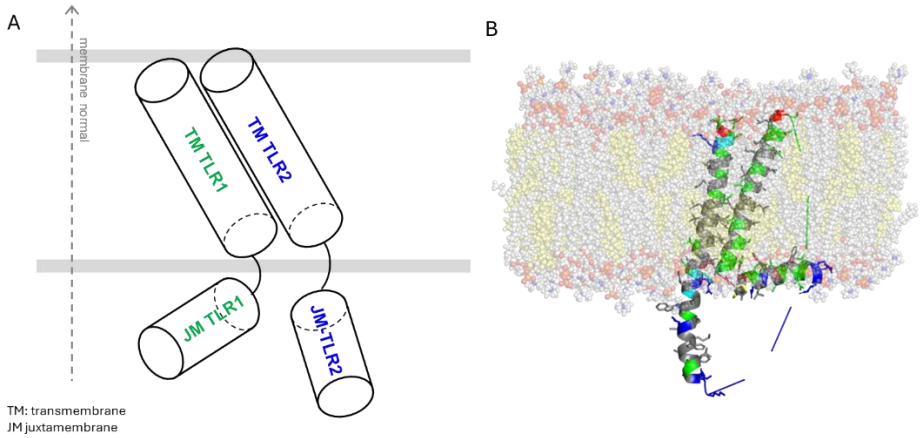

Figure 6.5 A. Schematic representation on the possible dimerization of the transmembrane domains from TLR2 and TLR1. B. Representation of one of the most plausible solutions obtained from different MD simulations on the TLR-2/1 TM+JM interface.

This study continues with the construction, docking and possible dimerizations on the intracellular domains. Once all is achieved, next step is to study the full-length of both dimers.

Chapter V:

Materials and Methods

V. Materials and Methods

8.1 Partial *Neisseria meningitidis* Serogroup-Y CPS depolymerization (Related to Chapter III: Section 3.1)

Dr. Fabrizio Chiodo provided the Serogroup Y CPS, and it was depolymerized in acetate buffer (pH of 4,5) for 8h at 90º, the solution was then neutralized with NaOH 1M and lyophilized. The depolymerized sample was then separated on size exclusion chromatography column TSK-40, obtaining different lengths of oligosaccharide.

8.2 Ligand-based NMR experiments (Related to Chapter III Section 3 and 4)

NMR spectra were acquired on a Bruker 600-MHz Avance Neo instrument equipped with a cryo probe. NMR samples were dissolved in 50 mM deuterate water and [D4] (trimethylsilyl)propionic acid, sodium salt (TSP, 10 μM) was used as internal reference to calibrate all the spectra. Data acquisition and processing were analyzed using TOPSPIN 4.4 software. Homonuclear 2D 1H-1H NOESY experiments were carried out by using data sets of 4096x900 points and mixing times of 100-300 ms. 2D homonuclear spectra were recorded with data sets of 4096x900 (t1 x t2) points and the data matrix processed with zero-filling in the F1 dimension up to 4096x2048 points. To improve the resolution, a cosine-bell function was used before Fourier transformation in both dimensions. Heteronuclear single quantum coherence (HSQC) experiments were carried out with setting data points of 2048x600.

8.2.1 STD-NMR

A protein- ligand ratio of 1:40-100 and a saturation time of 2 s were used with the on-resonance pulse at 7.5 ppm and the off-resonance at 40 ppm. By using

these conditions, no STD signals were observed in the control STD NMR spectrum of the ligand alone. A train of 50 ms (field strength of 21 Hz) Gaussian shaped pulse with an attenuation of 60db has been used to saturate the protein. The epitope mapping of the partially depolymerized CPS Men-Y was achieved by the calculation of the ratio (I0 – Isat)/I0, where (I_0 – Isat) is the intensity of the signal in the STD NMR spectrum and I0 is the peak intensity referred to the unsaturated reference spectrum (off-resonance).

8.2.2 Tr-NOESY/ROESY analysis (Related to Chapter III: Section 3.1)

Protein/ligand molar ratios varied from 1:10 to 1:20. Homonuclear 2D ^{1}H-^{1}H NOESY/ROESY experiments were carried out by using data sets of 2048x512 points and mixing times of 600-700 ms for the free states and of 300 ms and 400 ms for the bound states.

Proton – proton cross relaxation rates were measured integrating the NOE/ROE cross peaks of interest. The raw data of each cross peak were normalized against the corresponding diagonal peak and the ^{1}H-^{1}H distances were calculated using the following equation (Equation 2.5 from Chapter II).

$$r_{ij} = r_{ref} \sqrt[6]{\frac{\sigma_{ref}}{\sigma_{ij}}}$$

8.4 NMR protein assignment (Related to Chapter III: Section 3.1)

Triple resonance experiments for protein NMR assignment HNCA, HNCACB, HNCO and CBCAcoNH were recorded at 298 K on a Bruker's AvanceTM NEO 900 MHz spectrometer, equipped with TCI cryo-probe. 3D HNcaCO was recorded at 298 K on a Bruker's AvanceTM 500 MHz spectrometer. 93 % of the amino acid sequence from Y26 to T147 was assigned, excluding the 5 proline residues. Data acquisition and processing was performed on TOPSPIN

4.1.1 software and spectra were analyzed by using CARA (Computer Aided Resonance Assignment) software.

8.4.1 Protein-based NMR experiments (related to Chapter III: section 3.1)

For ligand binding studies, 2D ^{1}H-^{15}N HSQC NMR experiments were recorded on samples of 200 µM [U-^{15}N] Siglec-7 CRD in 200 µL of aqueous buffered solution (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM of protease inhibitors, 10% D2O in a 3 mm NMR tube. Experiments were acquired on a Bruker's Avance NEO 600 MHz spectrometer, equipped with triple resonance cryo-probe. The spectra were acquired using 32 scans with 2048 data point in t2, 128 increments in the indirect dimension (t1), a recycle delay of 1.2 seconds, and the temperature was kept at 298 K. The interaction of Siglec-7 CRD with the ligands was investigated by adding increasing amounts ligands 1, 2 or 3, to three different tubes containing Siglec-7 CRD, to reach ligand concentrations of 12.5, 25, 50, 100, 200, 400, 800, 1600 and 3200 µM. 2D ^{1}H-^{15}N HSQC spectra were added upon the addition of each ligand aliquot. Data acquisition and processing were performed with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA.

8.5 Immunoarrays: ELISA (Related to Chapter III: Section 3.1 and 3.2)

C-type lectins ELISA. A solution of 50 uL of from Men-Y and Men-W CPS from *Neisseria meningitidis* isolates (Neu5Ac-Glc) and Neu5Ac-Gal 50 ug/mL in PBS (10mM, pH= 7.4) provided by Finlay Institute Cuba were used to coat the Nunc MaxiSorp plate 2h at room temperature. After discarding and washing (2x150µL) with calcium and magnesium-containing buffer TSM [20 mM tris(hydroxymethyl)aminomethane (Tris)-HCl, pH 8.0; 150 mM NaCl; 1 mM CaCl2; 2 mM MgCl2], the wells were blocked with 100 µL of 1% BSA (Sigma-Aldrich, lyophilized powder, ≥96%, agarose gel electrophoresis) in TSM at

37°C for 30 min. The blocking solution was discarded and 50µL of the different C-type-lectins human-Fc at 1 µg/mL in assay buffer (TSM, 0.5% BSA) were added to the wells. After 1h at room temperature, the wells were washed with TSM (2x150µL) and then 100 µL of anti-human horseradish peroxidase (0.3 µg/mL, Goat anti-Human IgG-HRP from JacksonImmuno) were added. After 30 min at room temperature, the wells were washed with TSM (2x150µL). Finally, 100 µL of substrate solution (3,3′,5,5′-Tetramethylbenzidine, TMB, in citric/acetate buffer, pH=4, and H2O2) were added and after max 15 min incubation at room temperature the reaction was stopped with 50 µL of H2SO4 (0.8 M) and the optical density (OD) was measured at 450 nm in an ELISA reader. Polyacrylamide polymers (PAA), at functionalized with different glycans were purchased from Lectinity, MW approx. 20 KDa, Carbohydrate content around 20% mol. Galβ1-4(Fucα1-3)GlcNAcβ-OCH2CH2CH2NH2 0044-PA (PAA-LeX, positive control for DC-SIGN), and GalNAcα-OCH2CH2CH2NH2 0030-PA (PAA-Tn, positive control for MGL), we used a concentration of 40µg/mL PAA-glyco-conjugates to coat the ELISA wells. The experiment has been performed three times in duplicate with similar results. Data were normalized over binding signal from the positive controls used for each C-type lectin (set as 100% of binding). Siglecs-ELISA: A solution of 50 uL of from Men-Y and Men-W CPS from Neisseria meningitidis isolates (Neu5Ac-Glc) and Neu5Ac-Gal 50 ug/mL in PBS (10mM, pH= 7.4) provided by Finlay Institute Cuba were used to coat the Nunc MaxiSorp plate 2h at room temperature. After discarding and washing (2x150µL) with Hanks' Balanced Salt solution (Gibco™ HBSS) the wells were blocked with 200 µL of carbo-free blocking solution (Vector Laboratories, catalog No.NC9977573) at 37°C for 30 min. The blocking solution was discarded and 50µL of the different human Siglecs-Fc constructs (recombinant human chimera purchased from R&D Systems Netherlands) at 5 µg/mL (for Siglec-9 100ng/mL) in assay buffer (carbo-free solution) were added to the wells. After 2h at room temperature, under gentle shaking (150 rot/min), the wells were washed with HBSS (2x150µL) and then 100 µL of anti-human horseradish peroxidase (0.3 µg/mL,

Goat anti-Human IgG-HRP from JacksonImmuno) were added. After 30 min at room temperature, the wells were washed with HBSS (2x150µL). Finally, 100 µL of substrate solution (3,3′,5,5′-Tetramethylbenzidine, TMB, in citric/acetate buffer, pH=4, and H2O2) were added and after max 15 min incubation at room temperature the reaction was stopped with 50 µL of H2SO4 (0.8 M) and the optical density (OD) was measured at 450 nm in an ELISA reader. Polyacrylamide polymers (PAA) functionalized with different glycans were purchased from Lectinity, MW approx. 20 KDa, Carbohydrate content around 20% mol. Neu5Acα6`Lac-C2-PAA 0063a-PA (PAA-Sia 2,6), Neu5Acα3`Lac-Gly-PAA 0060-PA (PAA-Sia 2,3), we used a concentration of 40µg/mL PAA-glyco-conjugates to coat the ELISA wells. The experiment has been performed three times in duplicate with similar results. Data were normalized over binding signal from the positive controls used for each Siglec (set as 100% of binding).

8.6 Fluorescence titrations (Related to Chapter III)

Stationary-state fluorescence spectra were collected on a Fluoromax-4 spectrophotometer (Horiba, Edison, NJ, USA). Emission spectra were recorded in the range of 310 to 450 nm with excitation at 295 nm. The slit widths were set to 5 nm for excitation and to 5 nm for emission wavelength. A quartz cuvette with a pathlength of 1 cm and a volume of 0.2 mL was used. The starting solution of Siglec-7 CRD protein, prepared at a fixed concentration of 4 µM in PBS buffer (pH 7.4), was titrated by adding small aliquots of a stock solution of 100 µM of the ligands Men-Y and Men-W capsular polysaccharides from Neisseria meningitidis, already depolymerized and purified. The experiments were carried out at 25°C. For all ligands analyzed, the fluorescence intensity was quenched. The binding curve was obtained by fitting the data using nonlinear regression with a 1:1 binding model using the function described by Ribeiro et al. (plotting the ratio between the fluorescence intensity at each addition of ligand (F) and the fluorescence intensity of the protein in absence of ligand (F0) against the to the total ligand concentration [L] in µM.

$$\frac{\Delta I_f}{I_0} = \frac{\Delta I_{max}}{I_0} X_{FY} \qquad [8.1]$$

Where If is the fluorescence intensity change upon the ligand, Imax the maximal fluorescence, F the fluorophore Y the interference specie and XFY = $-b \pm \sqrt{b^2} - 4ac/2a$ with a=[F]tKb, b=1+[Y]tKb, c=[Y]tKb, with Kb= association constant. For evalutating the binding interactions, specifically the dissociation constant Kd, the data was analyzed through a non-linear regression equation considering a unique site-specific binding model:

$$Y = \frac{B_{max} * X}{K_D + X} \qquad [8.2]$$

8.7 *In silico* methods (Related to Chapters III-V)

8.7.1 Molecular mechanics

MM studies were conducted using the Maestro software. With it, the adiabatic energy maps were calculated for each disaccharide that was linked by a glycosidic linkage, defined by the torsion angles phi (H1-C1-O-CX'), psi (C1-O-CX'-HX') and in some cases, omega (OX'-CX'-CY'-O5'). The MM3* forcefield that is already included in MacroModel from Maestro, was set at 80 for dielectric constant for calculations. For each disaccharide, the dihedral angles were fvaried incrementally using a grid step of 20 degrees. The corresponding flexible maps were draun as 2D countour plots using the graphical interface of MacroModel.

8.7.2 Ligand preparation for computational analysis.

The 3D coordinates for multiple ligands were generated using the GLYCAM database. Only those whose parameters were not found in GLYCAM were parametrized by a home-made protocol produced in the Prof. Marchetti's lab. Subsequently, the ligand geometries underwent optimization through MD simulations.

8.7.3 Docking calculations.

Docking calculations were performed using AutoDock 4.2 ADT). First, a 3D grid was chosen with a volume necessary to cover all the ligand and the key binding aminoacids of the target protein. When the bioactive conformation of a certain sialoglycan was known, the torsion angles were blocked before the calculation. A distance-dependent dielectric constant and the original Lennard-Jonnes and hydrogen bonding potentials provided by AutoDock were employed. A total 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100. After docking, the 100 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å. The clusters were ranked to the lowest energy representative of each one. The analysis of the docking poses was performed with AutoDock 4.2 Tools and the conformations obtained were ordered by energy from lowest to highest and chategorized by population of a certain conformer protein-ligand.

8.7.4 MD simulation

Molecular dynamics calculations were performed within the the AMBER 18 software package using explicit water with the following forcefields: Glycam06j-1 for the glycans, FF14SB for the protein and Gaff for organic moieties. For the protein preparation, missing hydrogen atoms were added, protonation state of ionizable groups were added (Na+ and Cl-) and cap termini were computed by using Pymol. The input files were generated using the tleap modules of the AMBER packages, the minimization step was performed by using Sander module and the productions of the Molecular Dynamic calculations were performed by using the PMEMD module. The corresponding molecules were positioned within a truncated octahedral box of TIP3P water of the proper size and the remote interactions were calculated using a cut off of 15 Å and counterions (Na+ and Cl-) were added to neutralize the whole system. After the preparation of the input files, an energy minimization process was performed to refine the initial structure. The calculations that employees SHAKE for the C-H bond and 1 fs of integration step. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to

represent the electrostatic interactions, with a grid space of 1 Å. The system was minimized, firstly, by holding the solute fixed, while a second minimization performs the entire system. Afterwards, the whole system got slowly heated using constant pressure and constant volume from 0 to 300 K using a weak restrain on the solute, and then, the system was equilibrated at 300 K using the constant pressure and removing the weak restrains that were put in the solute. The system coordinates were saved and used for 100 ns or 500 ns simulation using the PMEMD module implemented in AMBER. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding and assembling 10000 structures for each complex, which were after analyzed. The stability of energy, pressure, temperature and other thermodynamic parameters were monitored along the trajectory and them, RMSD torsion values, cluster distances and hydrogen bonds were extracted by using the Cpptraj module. This module was useful for the analysis process of the trajectories and the coordinate files that have been created during the MD simulation. VMD software, Pymol software and Chimera software were used to visualize the trajectory. The generated data were also automatized with a personal plotting script.

8.8 Immunological assays (related to Chapter III)

U937 (ATCC® CRL-1593.2TM; Manasas, VA, USA) and THP1 monocytes (ATCC® TIB-202; Manasas, VA, USA) were cultured in RPMI-1640 with L-glutamine, supplemented with 10% inactivated fetal bovine serum (FBS; Gibco, Carlsbad, CA, USA), 100 U/mL penicillin, and 100 g/mL streptomycin (Gibco, Carlsbad, CA, USA), 1X Normocin (Invivogen, San Diego, CA, USA) and 0.05 mM 2-mercaptoethanol (Gibco, Carlsbad, CA, USA) in a 5%-CO2 humidified incubator at 37°C. To induce cell attachment and naïve macrophage differentiation we applied a specific protocol for U937 cells and THP-1 cells. Respectively, 2x105 U937 cells were seeded in a 12-well plate and then treated with phorbol 12-myristate 13-acetate (PMA) at 100 ng/ml concentration for 48 hours, followed by a recovery phase of additional 48 hours (doi:10.1016/j.jprot.2009.08.001). At the same time, 2x105 THP-1 cells were

treated with 20nM PMA for 24 hours, followed by a 72 hours recovery phase (https://doi.org/10.1016/j.jim.2019.112721). Subsequently, both U937-PMA cells and THP1-PMA cells were incubated with varying concentrations (0.1 μg/ml and 1 μg/ml) of capsular polysaccharides (CPS) derived from MenY, or Lipopolysaccharide (LPS) and then maintained at 37°C for 16 hours. After incubation, cells were washed with PBS before being directly lysed on plate using Norgen Biotek RNA extraction kit (Thorold, Ontario, Canada), following the manufacturer's protocol. The extracted RNA (1μg) was reverse transcribed using the Quantitect Reverser transcription Kit (Qiagen, Hilden, Germany). For real-time quantitative PCR (qPCR), 10 ng of cDNA was utilized in a 10 μl PowerUP SYBR Green Master mix PCR kit (Thermo Fisher, Waltham, Massachusetts, USA) reaction, employing primers listed in Table X, and analyzed using the AriaMX system (Agilent, Santa Clara, California, USA). Gene expression was determined by □□Ct method, normalizing TNFα and IL-8 gene expression to β-Actin, before comparing to untreated control (10.3390/ijms22179379).

Human IL-8: FW – ATGACTTCCAAGCTGGCCGT – RV - TACAATAATTTCTGTGTTGGC

Human TNFα: FW – GTTGTAGCAAACCCTCAAGCTG – RV - GAGGTACAGGCCCTCTGAT

Human β-Actin: FW – CCGACAGGATGCAGAAGGAG – RV - GCCTAGAAGCATTTGCGGTG

Chapter VI:

References

VI. References

[1]. Varki, A., & Gagneux, P.. Biological Functions of Glycans. In A. Varki (Eds.) et. al., *Essentials of Glycobiology*. **2017**, *3*,77–88.

[2]. Oswald, D. M.; Cobb, B. A. Emerging Glycobiology Tools: A Renaissance in Accessibility. *Cell. Immunol***. 2018**, *333*, 2–8.

[3]. Zhou, J. Y.; Cobb, B. A. Glycans in Immunologic Health and Disease. *Annu. Rev. Immunol*. **2021**, *39*, 511–536.

[4]. Gimeno, A.; Valverde, P.; Ardá, A.; Jiménez-Barbero, J. Glycan Structures and Their Interactions with Proteins. A NMR View. *Curr. Opin. Struct. Biol*. **2020**, *62*, 22–30.

[5]. C.J. Schultz, Glycoproteins, Plant, Encyclopedia of Biological Chemistry (Second Edition), *Academic Press*, **2013**, 465-468.

[6]. Valverde, A. Ardà, N. C. Reichardt, J. Jiménez-Barbero, A. Gimeno. "Glycans in drug discovery" *Med Chem Commun*, **2019**, *10*, 1678.

[7]. Pinho, S. S., & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, **2015**, 15(*9*), 540-555.

[8]. T. H. Mogensen "Pathogen recognition and inflammatory signaling in innate immune defenses" *Clin. Microbiol. Rev*., **2009**, *22*, 240–273

[9]. H. H. Khan, B. Shi, Y. Tian, T. Wang, S. Hussain, F. U. Khan, Z. Khan, B. Ashfaq, H. A. T. Ahmad. "Glycan regulation in cancer, nervous and immune system: A narrative review" *Biomedical Research and Therapy*, **2019**, *4* 3113-3120.

[10]. R. K. Yu, Y. T. Tsai, T. Ariga and M. Yanagisawa. "Structures, biosynthesis, and functions of gangliosides—An overview" *J. Oleo Sci***., 2011**, 60, *10*, 537–544.

[11]. Gaudette, S., Hughes, D., & Boller, M. (2020). The endothelial glycocalyx: Structure and function in health and critical illness. *Journal of veterinary emergency and critical car*e, **2001**, 30(*2*), 117–134.

[12]. Ryan 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**

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

[14]. T. Schjoldager, Y. Narimatsu, H. J. Joshi, and H. Clausen, "Global view of human protein glycosylation pathways and functions," *Nat. Rev. Mol. Cell Biol.*, **2020,** 21 (*12*), 729–749.

[15]. D. Dutta, C. Mandal, and C. Mandal, "Unusual glycosylation of proteins: Beyond the universal sequon and other amino acids," *Biochim. Biophys. Acta - Gen. Subj.,* **2017**, 1861 (12), 3096–3108.

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

[17]. Chen, M. M., Bartlett, A. I., Nerenberg, P. S., Friel, C. T., Hackenberger, C. P., Stultz, C. M., Radford, S. E., & Imperiali, B. Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, **2010,** 107(*52*), 22528–22533.

[18]. Varki and T. Angata. "Siglecs- the major subfamily of I-type lectins." *Glycobiology,* **2006,** 16, 1R–27R.

[19]. Talabnin, C. Talabnin, et al. "Increased expression of the high-mannose M6N2 and NeuAc3H3N3M3N2F tri-antennary N-glycans in cholangiocarcinoma" *Oncol Lett.* **2018,** 15(*1*), 1030–1036.

[20]. M. E. Taylor and K. Drickamer. Introduction to glycobiology. *Oxford university press Inc.* **2011**, 3rd edition.

[21]. Adamczyk, T. Tharmalingam, P. M. Rudd. "Glycans as cancer biomarker" *Biochim. Biophys. Acta,* **2012**, *1820*, 347–1353

[22]. Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature*, **2007**, 446(*7139*), 1023-1029.

[23]. Schauer, R. Sialic acids as regulators of molecular and cellular interactions. *Current Opinion in Structural Biology*, **2009**, 19(*5*), 507-514.

[24]. Varki, A., & Schauer, R. Sialic acids. In Varki, A., Cummings, R. D., Esko, J. D., et al. (Eds.), Essentials of Glycobiology (2nd ed.). **2009**. Cold Spring Harbor Laboratory Press. Chapter 14: Sialic Acids

[25]. Angata, T., & Varki, A.. Chemical diversity in the sialic acids and related α-keto acids: An evolutionary perspective. *Chemical Reviews*, **2002**, 102(*2*), 439-469.

[26]. Büll, C., & Boltje, T. J. Balancing the roles of sialic acids in health and disease. *Chemical Science*, **2020**, 11(*25*), 6635-6646.

[27]. Suzuki, M., Angata, K., Nakayama, J., & Fukuda, M. Polysialic acid and mucin type o-glycans on the neural cell adhesion molecule differentially regulate myoblast fusion. *The Journal of biological chemistry*, **2003**, 278(*49*), 49459–49468.

[28]. Pomin, V. H., & Mulloy, B. Glycosaminoglycans and Proteoglycans. *Pharmaceuticals,* **2018**, 11(*1*), 27.

[29]. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., & Tabak, L. A. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, **2012**, 22(*6*), 736–756.

[30]. Trinchieri, G., & Sher, A. Cooperation of Toll-like receptor signals in innate immune defence. *Nature reviews. Immunology*, **2007**, 7(*3*), 179–190.

[31]. A. Molinaro, O. Holst, F. Di Lorenzo, et al. "Chemistry of lipid A: at the heart of innate immunity" *Chem Eur J.,* **2015**, 21, 500–19.

[32]. Marchetti, R., Forgione, R. E., Fabregat, F. N., Di Carluccio, C., Molinaro, A., & Silipo, A. Solving the structural puzzle of bacterial glycome. *Current opinion in structural biology*, **2021**, 68, 74–83

[33]. C. Soliman, G.B. Pier, P.A. Ramsland. "Antibody Recognition of Bacterial Surfaces and Extracellular Polysaccharides" *Current Opinion in Structural Biology,* **2020,** 62, 48–55.

[34]. Cress BF, Englaender JA, He W, Kasper D, Linhardt RJ, Koffas MA. Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol Rev*. **2014**, 38(*4*):660-97.

[35]. Paola Cescutti, Bacterial capsular polysaccharides and exopolysaccharides, *Microbial Glycobiology*, **2010**, 93-108.

[36]. Laurence, M. Aringer. "Effector Mechanisms in Autoimmunity" The Autoimmune Disease Sixth Edition, **2020**.

[37]. Bojar, D., Powers, R. K., Camacho, D. M., & Collins, J. J.. Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions. *Cell host & microbe*, **2021**, 29(*1*), 132–144.e3.

[38]. Angata, K., Long, J. M., Bukalo, O., Lee, W., Dityatev, A., Wynshaw-Boris, A., Schachner, M., Fukuda, M., & Marth, J. D. Sialyltransferase ST8Sia-II assembles a subset of polysialic acid that directs hippocampal axonal targeting and promotes fear behavior. *The Journal of biological chemistry*, **2004**, 279(*31*), 32603–32613.

[39]. Mogensen T. H. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clinical microbiology reviews*, **2009**, 22(*2*), 240–27.3

[40]. G. A. Rabinovich, Y. van Kooyk, B.A. Cobb. "Glycobiology of immune responses: Glycobiology of immune responses" Ann. N. Y. Acad. Sci. 1253 (2012) 1–15

[41]. A. Varki, R.D. Cummings, J.D. Esko. et al., editors. "Glycan-Binding Proteins" Essentials of Glycobiology 3rd edition. C*old*

Spring Harbor (NY): Cold Spring Harbor Laboratory Press **2015-2017**

[42]. M. S. Macauley, P. R. Crocker, J. C. Paulson. "Siglec-mediated regulation of immune cell function in disease." *Nat. Rev. Immunol*., **2014**, 14 (*10*), 653– 666.

[43]. P. R. Crocker, A. Varki. "Siglecs in the immune system" *Immunology*, **2001**, 103, 137- 145.

[44]. Kogelberg, H., & Feizi, T. New structural insights into lectin-type proteins of the immune system. *Current opinion in structural biology*, **2001**, 11(*5*), 635–643.

[45]. J.E. Hudak, S.M. Canham, C.R. Bertozzi, "Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion" *Nat. Chem. Biol.* **2014,** 10, 69-75.

[46]. S. Pillai, I.A. Netravali, A. Cariappa, H. Mattoo. "Siglecs and immune regulation" *Annu. Rev. Immunol*. **2012**, 30, 357e392.

[47]. K. L. Hudson, G. J. Bartlett, R. C. Diehl, J. Agirre, T. Gallagher, L. L. Kiessling, D. N. Woolfson. "Carbohydrate-aromatic interactions in proteins" *J. Am. Chem. Soc*., **2015**, 137, 15152–15160.

[48]. J. Lübbers, E. Rodríguez, Y. van Kooyk. "Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions" *Front Immunol*., **2018**, 9, 2807.

[49]. Di Carluccio, et al. "Characterization of the dynamic interactions between complex N-glycans and human CD22" *Chembiochem.,* **2019**, 21, 129.

[50]. M.S. Macauley, P.R. Crocker, J.C. Paulson, "Siglec regulation of immune cell function in disease" *Nature reviews. Immunology* **2014,** 14, 653-666.

[51]. J.C. Paulson, M.S. Macauley, N. Kawasaki. "Siglecs as sensors of self in innate and adaptive immune responses" *Ann N Y Acad Sci*., **2012,** 1253(*1*), 37–48.

[52]. Varki A. Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. *Glycobiology,* **2011**, 21(*9*), 1121–1124.

[53]. Varchetta, et al. "Lack of siglec-7 expression identifies a dysfunctional natural killer cell subset associated with liver inflammation and fibrosis in chronic HCV infection" *Gut*., **2016**, 65, 1998–2006.

[54]. H. Attrill, et al. "Siglec-7 undergoes a major conformational change when complexed with the α2,8-disialylganglioside GT1b" *J. Biol. Chem*,. **2006**, 281, 32774–32783.

[55]. N. Yamakawa, Y. Yasuda, A. Yoshimura, A. Goshima, P. R. Crocker, G. Vergoten, Y. Nishiura, T. Takahashi, S. Hanashima, K. Matsumoto, Y. Yamaguchi, H. Tanaka, K. Kitajima, C. Sato. "Discovery of a new sialic acid binding region that regulates Siglec-7" *Scientific Reports*, **2020**, *10*, 1.

[56]. T. Avril, E.R. Wagner, H.J. Willison, P.R. Crocker. "Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on Campylobacter jejuni lipooligosaccharides" *Infect Immun*., **2006**, 74, 4133–4141.

[57]. D. Lamprinaki, P. Garcia-Vello, R. Marchetti, C. Hellmich, K.A. McCord, K.M. Bowles, M.S. Macauley, A. Silipo, C. De Castro, P.R. Crocker PR, N. Juge. "Siglec-7 Mediates Immunomodulation by Colorectal Cancer-Associated Fusobacterium nucleatum ssp. animalis" *Front. Immunol*. **2021**, *12*, 744184.

[58]. Zheng, Y.; Ma, X.; Su, D.; Zhang, Y.; Yu, L.; Jiang, F.; Zhou, X.; Feng, Y.; Ma, F. The Roles of Siglec7 and Siglec9 on Natural Killer Cells in Virus Infection and Tumour Progression. *J. Immunol. Res*. **2020**, 6, 2020:6243819.

[59]. Atxabal, U.; Nycholat, C.; Pröpster, J. M.; Fernández, A.; Oyenarte, I.; Lenza, M. P.; Franconetti, A.; Soares, C. O.; Coelho, H.; Marcelo,

F.; et al. Unraveling Molecular Recognition of Glycan Ligands by Siglec-9 via NMR Spectroscopy and Molecular Dynamics Modeling. *ACS Chem. Biol.* **2024**, 19 (*2*), 483–496.

[60]. Cheong, K. A.; Chang, Y. S.; Roh, J. Y.; Kim, B. J.; Kim, M. N.; Park, Y. M.; Park, H. J.; Kim, N. D.; Lee, C. H.; Lee, A. Y. A Novel Function of Siglec-9 A391C Polymorphism on T Cell Receptor Signaling. *Int. Arch. Allergy Immunol*. **2011**, 154 (*2*), 111–118

[61]. Munz, S. von Gunten. "Interactions between Siglec-7/9 receptors and ligands influence NK cell-dependent tumor immunosurveillance" *J. Clin. Invest*. **2014**, 124, 1810-1820.

[62]. Choi, H.; Ho, M.; Adeniji, O. S.; Giron, L.; Bordoloi, D.; Kulkarni, A. J.; Puchalt, A. P.; Abdel-Mohsen, M.; Muthumani, K. Development of Siglec-9 Blocking Antibody to Enhance Anti-Tumor Immunity. *Front. Oncol*. **2021**, *11* 1–10.

[63]. Boelaars, K.; van Kooyk, Y. Targeting Myeloid Cells for Cancer Immunotherapy: Siglec-7/9/10/15 and Their Ligands. *Trends in Cancer,* **2024**, 10 (*3*), 230–241.

[64]. G. Y. Chen, N. K. Brown, P. Zheng and Y. Liu. "Siglec-G/10 in self-nonself discrimination of innate and adaptive immunity." *Glycobiology*, **2014**, 24, 800–806.

[65]. Whitney, S. Wang, H. Chang, K. Y. Cheng, P. Lu, X. D. Zhou, W. P. Yang, M. McKinnon and M. Longphre. "A new siglec family member, siglec-10, is expressed in cells of the immune system and has signaling properties similar to CD33" *Eur. J. Biochem*. ,**2001**, 268,6083–6096.

[66]. R.E. Forgione, C. Di Carluccio, J. Guzman-Caldentey, R. Gaglione, F. Battista, F. Chiodo, Y. Manabe, A. Arciello, P. Del Vecchio, K. Fukase, A. Molinaro, S. Martín-Santamaría, P.R. Crocker, R. Marchetti, A. Silipo. "Unveiling molecular recognition of sialoglycans by human siglec-10" *iScience*, **2020**, 23(*6*), 101231.

[67]. G. Y. Chen, J. Tang, P. Zheng and Y. Liu. "CD24 and Siglec-10 selectively repress tissue damage-induced immune responses" *Science*, **2009**, 323, 1722–1725

[68]. E. Bandala-Sanchez, N. G. Bediaga, G. Naselli, A. M. Neale and L. C. Harrison. "Siglec-10 expression is up-regulated in activated human CD4 + T cells" *Hum. Immunol*., **2020**, 81, 101–104.

[69]. A. Barkal, R. E. Brewer, M. Markovic, M. Kowarsky, S. A. Barkal, B. W. Zaro, V. Krishnan, J. Hatakeyama, O. Dorigo, L. J. Barkal and I. L. Weissman "CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy" *Nature*Ya, **2019**, 572, 392–396.

[70]. M. Dobson. "Biophysical Techniques in Structural Biology", *Annu. Rev. Biochem*. ,**2019**, 88, 25–33.

[71]. M. A. Johnson, B. M. Pinto "NMR spectroscopic and molecular modeling studies of protein–carbohydrate and protein–peptide interactions" *Carbohydrate Research*, **2004**, 339, 907–928.

[72]. R. Marchetti, S. Perez, A. Arda, A. Imberty, J. Jimenez-Barbero, A. Silipo, A. Molinaro, "Rules of Engagement" of Protein–Glycoconjugate Interactions: A Molecular View Achievable by using NMR Spectroscopy and Molecular Modeling" *ChemistryOpen,* **2016**, 5, 274 – 296.

[73]. W. Becker, K. C. Bhattiprolu, N. Gubensäk, K. Zangger. "Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy" *ChemPhysChem*, **2018**, 19, 895 – 906.

[74]. Angulo, B. Langpap, A. Blume, T. Biet, B. Meyer, N R. Krishna, H. Peters, M. M Palcic, T. Peters., "Blood group B galactosyltransferase: insights into substrate binding from NMR experiments" *J Am Chem Soc*, **2006**, 128(*41*), 13529-38.

[75]. Neuhaus, D., & Williamson, M. P.. The Nuclear Overhauser Effect in Structural and Conformational Analysis ,**1989** ,(2nd ed)

[76]. Ni, F. Recent developments in transferred NOE methods Progress in NMR Spectroscopy,.,**1994**,26,517-606.

[77]. M. Mayer, B. Meyer. "Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy" *Angew. Chem Int Ed Engl*, **1999**, 38(*12*), 1784-1788.

[78]. Di Carluccio C, Forgione MC, Martini S, Berti F, Molinaro A, Marchetti R, Silipo A. Investigation of protein-ligand complexes by ligand-based NMR methods. *Carbohydr Res*. **2021**, 503:108313.

[79]. M. Mayer, B. Meyer. "Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor" *J. Am. Chem. Soc*., **2001**, 123(*25*), 6108–611.

[80]. Angulo, P. M. Enríquez-Navas, P. M. Nieto. "Ligand–Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates chemistry" Chemistry – *A European Journal,* **2010**, 26(*16*), 7803-7812

[81]. C. Fernandez, G. Wider. "TROSY in NMR studies of the structure and function of large biological macromolecules" *Current Opinion in Structural Biology*, **2003**,13, 570–580.

[82]. Ardá, A. Canales, F. J. Cañada, J. Jiménez-Barbero. "Chapter 1: Carbohydrate–Protein Interactions: A 3D View by NMR, in Carbohydrates in Drug Design and Discovery" ,**2015**, 1-20.

[83]. M. P. Williamson. "Chemical Shift Perturbation" *Modern Magnetic Resonance*, **2018**, 995-1012.

[84]. Di Carluccio, C.; Cerofolini, L.; Moreira, M.; Rosu, F.; Padilla-Cortés, L.; Gheorghita, G. R.; Xu, Z.; Santra, A.; Yu, H.; Yokoyama, S.; et al. Molecular Insights into O-Linked Sialoglycans Recognition by the Siglec-Like SLBR-N (SLBRUB10712) of Streptococcus Gordonii. *ACS Cent. Sci*. **2024**, 10 (*2*), 447–459.

[85]. Lewars, E. G. Computational Chemistry, Introduction to the Theory and Applications of Molecular and Quantum Mechanics, *Springer*, Third Edition, **2016**.

[86]. Hermann, J.; Spencer, J.; Choo, K.; Mezzacapo, A.; Foulkes, W. M. C.; Pfau, D.; Carleo, G.; Noé, F. Ab Initio Quantum Chemistry with Neural-Network Wavefunctions. *Nat. Rev. Chem*. **2023**, 7 (*10*), 692–709.

[87]. D.J. Rigden. "Ab Initio Protein Structure Prediction" (ed.) From Protein Structure to Function with Bioinformatics, **2009**, 3-25.

[88]. Kathleen M. Morgan and Joshua H. Baraban. Thermochemical Studies of Small Carbohydrates *The Journal of Organic Chemistry* **2024** 89 (*3*), 1769-1776.

[89]. Valery Lutsyk and Wojciech Plazinski. Exploring Ring Conformation in Uronate Monosaccharides: Insights from Ab Initio Calculations and Classical Molecular Dynamics Simulations. *The Journal of Physical Chemistry*, **2024** 128 (*2*), 472-491.

[90]. M. C. Payne, M. P. Teter, D. C. Allan, T. Arias, J. Joannopoulos. "Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients" *Rev. Mod. Phys*., **1992**, 64(*4*), 1045.

[91]. K. M. Dorst, G. Widmalm, Conformational Preferences at the Glycosidic Linkage ofSaccharides in Solution as Deduced from NMR Experimentsand MD Simulations: Comparison to Crystal Structures. *Chem. Eur. J*., **2024**, 30, e202304047.

[92]. Nieto-Fabregat, F.; Lenza, M. P.; Marseglia, A.; Di Carluccio, C.; Molinaro, A.; Silipo, A.; Marchetti, R. Beilstein. Computational toolbox for the analysis of protein–glycan interactions. *J. Org. Chem*., **2024**, 20, 2084–2107.

[93]. K. Vanommeslaeghe, O. Guvench, A.D. Jr MacKerell. "Molecular mechanics" *Current pharmaceutical design*, **2014**, 20(*20*), 3281–3292.

[94]. N. S. Pagadala, K. Syed, J. Tuszynski. “Software for molecular docking: a review” *Biophysical reviews*, **2017**, 9(*2*), 91–102.

[95]. Hosen, M. A., El Bakri, Y., Rehman, H. M., Hashem, H. E., Saki, M., & Kawsar, S. M. A. A computational investigation of galactopyranoside esters as antimicrobial agents through antiviral, molecular docking, molecular dynamics, pharmacokinetics, and bioactivity prediction. *Journal of Biomolecular Structure and Dynamics*, **2023**,42(*2*), 1015–1030.

[96]. G. M. Morris, D. S. Goodsell, R. Huey, A.J. Olson. “Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4” *Journal of computer-aided molecular design*, **1996**,10 (*4)* 293–304.

[97]. J. Li, A. Fu, L. Zhang. “An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking” *Interdisciplinary sciences, computational life sciences*, **2019**, 11(*2*) 320–328.

[98]. Samanta, P.; Mishra, S.K.; Pomin, V.H.; Doerksen, R.J. Docking and Molecular Dynamics Simulations Clarify Binding Sites for Interactions of Novel Marine Sulfated Glycans with SARS-CoV-2 Spike Glycoprotein., *Molecules,* **2023**, 28, 6413.

[99]. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson. “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility” *J Comput Chem*., **2009**, 30(*16*) 2785–2791.

[100]. K. Crampon, A. Giorkallos, M. Deldossi, S. Baud, L. A. Steffenel. “Machine-learning methods for ligand–protein molecular docking” *Drug Discovery Today*, **2022**, 27(*1*), 51-164.

[101]. W.F. de Azevedo, R. Dias. “Molecular Docking Algorithms” *Bioorg. Med. Chem*., **2008**,16(*20*), 9378.

[102]. Morris et al. “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function” *Journal of Computational Chemistry,* **1998,** 19(*14*), 1639-1662.

[103]. D. C. Young. Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems. S. A. Hollingsworth, R.O. Dror. “Molecular Dynamics Simulation for All” *Neuron*, **2018**, 99(*6*) 1129–1143.

[104]. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, M. Ghazimirsaeed, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, Z. Huang, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, Z. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, C. Zhao, Q. Zhu, and P.A. Kollman **2024**, Amber 2024, University of California, San Francisco.

[105]. D.A. Case, H.M. Aktulga, K. Belfon, D.S. Cerutti, G.A. Cisneros, V.W.D. Cruz eiro, N. Forouzesh, T.J. Giese, A.W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M.C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M.R. Machado, H.M. Nguyen, K.A. O’Hearn, A.V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T.E. Cheatham III, D.R. Roe, A. Roitberg, C. Simmerling, D.M. York, M.C. Nagan*, and K.M. Merz Jr.* AmberTools. J. *Chem. Inf. Model*. **2023**, 63, 6183-6191.

[106]. N.L. Allinger, M. Rahman, H. Lii. "A molecular mechanics force-field (Mm3) for alcohols and ethers" *J. Am. Chem. Soc*., **1990**,112,8293–8307.

[107]. Callum J. Dickson, Ross C. Walker, and Ian R. Gould. Lipid21: Complex Lipid Membrane Simulations with AMBER. *Journal of Chemical Theory and Computation* **2022** 18 (*3*), 1726-1736.

[108]. Kirschner, K.N., Yongye, A.B., Tschampel, S.M., Daniels, C.R., Foley, B.L, Woods, R.J. *J. Comput. Chem*. **2008**, 29, 622-655.

[109]. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. **2015**; 11(*8*), 3696-713.

[110]. Pekka Mark and Lennart Nilsson Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K *The Journal of Physical Chemistry A*, **2001** 105 (*43*), 9954-9960.

[111]. S. Jo, T. Kim, V.G. Iyer. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem*. **2008**, 29, 1859-1865.

[112]. S. Perez, E. Fadda, O.N. Makshakova. "Computational Modeling in Glycoscience" Comprehensive Glycoscience, 2nd edition **2020**.

[113]. Yamakawa, N.; Yasuda, Y.; Yoshimura, A.; Goshima, A.; Crocker, P. R.; Vergoten, G.; Nishiura, Y.; Takahashi, T.; Hanashima, S.; Matsumoto, K.; et al. Discovery of a New Sialic Acid Binding Region That Regulates Siglec-7. *Sci. Rep*. **2020**, 10*(1)* 1–13.

[114]. Chang, L.; Liang, S.; Lu, S.; Tseng, H. C.; Tsai, H.; Tang, C.; Sugata, M.; Chen, Y.; Chen, Y.; Wu, S.; et al. Basis and Role of Siglec-7 Ligand Expression on Chronic Lymphocytic Leukemia B Cells. *Front. Immunol.* **2022**, 13 ,1–.

[115]. Yadav S, Rammohan G. Meningococcal Meningitis. *IDCases* **2020**, 21, e00897.

[116]. Colicchio R, Pagliuca C, Ricci S, Scaglione E, Grandgirard D, Masouris I, Farina F, Pagliarulo C, Mantova G, Paragliola L, Leib SL, Koedel U, Pozzi G, Alifano P, Salvatore P. Virulence Traits of a Serogroup C Meningococcus and Isogenic cssA Mutant, Defective in Surface-Exposed Sialic Acid, in a Murine Model of Meningitis. *Infect Immun*. **2019**, 87(*4*). e00688-18.

[117]. Harrison LH, Pelton SI, Wilder-Smith A, Holst J, Safadi MA, Vazquez JA, Taha MK, LaForce FM, von Gottberg A, Borrow R, Plotkin SA.The Global Meningococcal Initiative: recommendations for reducing the global burden of meningococcal disease. Vaccine. .2011 29(18):3363-71

[118]. Harrison, O. B.; Claus, H.; Jiang, Y.; Bennett, J. S.; Bratcher, H. B.; Jolley, K. A.; Corton, C.; Care, R.; Poolman, J. T.; Zollinger, W. D.; et al. Description and Nomenclature of Neisseria Meningitidis Capsule Locus. *Emerg. Infect. Dis.,* **2013**, 19. 4.

[119]. Jones C, Virji M, Crocker PR. Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. *Mol Microbiol*, **2003**. 49(*5*). 1213-25.

[120]. Sindrewicz, P., Li, X., Yates, E.A. et al. Intrinsic tryptophan fluorescence spectroscopy reliably determines galectin-ligand interactions. *Sci Rep.,* **2019**, 9, 11851.

[121]. Di Carluccio, C.; Forgione, R.E.R.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B.A.; et al. Molecular recognition of sialoglycans by streptococcal Siglec-like adhesins: Toward the shape of specific inhibitors. *RSC Chem. Biol*. **2021**, 2, 1618–1630.

[122]. Gonzalez-Gil A, Schnaar RL. Siglec Ligands. *Cells*. **2021**, 10(*5*), 1260.

[123]. Di Carluccio C, Forgione MC, Martini S, Berti F, Molinaro A, Marchetti R, Silipo A. Investigation of protein-ligand complexes

by ligand-based NMR methods. *Carbohydr Res*., **2021**, 503; 108313.

[124]. Marchetti, R.; Perez, S.; Arda, A.; Imberty, A.; Jimenez-Barbero, J.; Silipo, A.; Molinaro, A. "Rules of Engagement" of Protein–Glycoconjugate Interactions: A Molecular View Achievable by Using NMR Spectroscopy and Molecular Modeling. *ChemistryOpen*., **2016**, 5(*4*), 274–296.

[125]. Nieto-Fabregat, F.; Lenza, M. P.; Marseglia, A.; Di Carluccio, C.; Molinaro, A.; Silipo, A.; Marchetti, R. Beilstein *J. Org. Chem*. **2024**, 20, 2084–2107.

[126]. Soares, C. O.; Grosso, A. S.; Ereño-Orbea, J.; Coelho, H.; Marcelo, F. Molecular Recognition Insights of Sialic Acid Glycans by Distinct Receptors Unveiled by NMR and Molecular Modeling. *Front. Mol. Biosci*. **2021**, 8, 1–16.

[127]. Kuttel, M. M.; Timol, Z.; Ravenscroft, N. Cross-Protection in Neisseria Meningitidis Serogroups Y and W Polysaccharides: A Comparative Conformational Analysis. *Carbohydr. Res*. **2017**, 446–447.

[128]. Pirone, L., Lenza, M.P., Di Gaetano, S., Capasso, D., Filocaso, M., Russo, R., Di Carluccio, C., Saviano, M., Silipo, A., Pedone, E., 2024. Biophysical and Structural Characterization of the Interaction between Human Galectin-3 and the Lipopolysaccharide from Pseudomonas aeruginosa. *Int. J.Mol. Sci.,* **2024**. 25, 2895.

[129]. Kirschner, K.N., Yongye, A.B., Tschampel, S.M., Daniels, C.R., Foley, B.L, Woods, R.J. *J. Comput. Chem*. **2008**, 29, 622-655.

[130]. Attrill, H., Imamura, A., Sharma, R.S., Kiso, M., Crocker, P.R., van Aalten, D.M.F. Siglec-7 Undergoes a Major Conformational Change When Complexed with the α-(2,8)-Disialylganglioside GT1b. *J Biol Chem.,* **2006**, 281, 32774-327.

[131]. Forrester, M. A., Wassall, H. J., Hall, L. S., Cao, H., Wilson, H. M., Barker, R. N., & Vickers, M. A. Similarities and differences in surface receptor expression by THP-1 monocytes and differentiated macrophages polarized using seven different conditioning regimens. *Cellular immunology,* **2018**.,332, 58–76.

[132]. Parikh SR, Campbell H, Beebeejaun K, Ribeiro S, Gray SJ, Borrow R, Ramsay ME, Ladhani SN. Meningococcal Group W Disease in Infants and Potential Prevention by Vaccination. *Emerg Infect Dis*., **2016**, 22(*8*).

[133]. Emily Sotheran, Courtney R Lane, Kristy Horan, Kerrie Stevens, Christine Guglielmino, Susan Bradbury, Karina Kennedy, Louise Cooley, Belinda McEwan, Charlene M Kahler, Shakeel Mowlaboccus, David J Speers, Robert Baird, Kevin Freeman, Lex Leong, Morgyn Warner, Deborah A Williamson, Jodie McVernon, Monica Lahra, Amy V Jennison, Benjamin P Howden, Patiyan Andersson, Genomic Surveillance of Invasive Meningococcal Disease During a National MenW Outbreak in Australia, Open Forum *Infectious Diseases*, **2024,** 14(*6*),e249.

[134]. Di Carluccio, Cristina et al. Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22 *iScience*, **2020**, 24(1), 10199.

[135]. Rivero-Calle, I.; Raguindin, P. F.; Gómez-Rial, J.; Rodriguez-Tenreiro, C.; Martinón-Torres, F. Meningococcal Group B Vaccine for the Prevention of Invasive Meningococcal Disease Caused by Neisseria Meningitidis Serogroup B. *Infect. Drug Resist*. **2019**, 12, 3169–3188.

[136]. Villanueva-Cabello, T. M.; Gutiérrez-Valenzuela, L. D.; Salinas-Marín, R.; López-Guerrero, D. V.; Martínez-Duncker, I. Polysialic Acid in the Immune System. *Front. Immunol*. **2022**, 12, 1–21.

[137]. Nagae, Masamichi et al. "Crystal Structure of Anti-polysialic Acid Antibody Single Chain Fv Fragment Complexed with Octasialic

Acid.", *Journal of Biological Chemistry*, **2013**, 288(*47*), 33784 – 33796.

[138]. Brisson et al, Helical Epitope of the Group B Meningococcal α(2-8)-Linked Sialic Acid Polysaccharide., *Biochemistry*, **1992**, 31, 4996-5004.

[139]. Avila, M., Ojcius, D. M., & Yilmaz, O. The oral microbiota: living with a permanent guest. *DNA and cell biology*, **2009**, 28(*8*), 405–411.

[140]. Zijnge, V., van Leeuwen, M. B., Degener, J. E., Abbas, F., Thurnheer, T., Gmür, R., & Harmsen, H. J. (2010). Oral biofilm architecture on natural teeth. *PloS one*, **2010**, 5(*2*), e9321.

[141]. Kolenbrander, P. E., Palmer, R. J., Jr, Periasamy, S., & Jakubovics, N. S. Oral multispecies biofilm development and the key role of cell-cell dis-tance. *Nature reviews. Microbiology,* **2010**, 8(*7*), 471–480.

[142]. Brennan, C. A., & Garrett, W. S. Fusobacterium nucleatum - symbi-ont, opportunist and oncobacterium. *Nature reviews. Microbiology*, **2019**, 17(*3*), 156–166.

[143]. Sveen, K., & Hofstad, T. Synergistic effect on blastogenesis in murine spleen cells of lipopolysaccharide, lipid A, and acid-degraded polysaccharide from Fusobacterium nucleatum. *FEMS microbiology immunology*,**1991**, 3(*1*), 47–52.

[144]. Williams MD, Kerber CA, Tergin HF. Unusual presentation of Lemierre's syndrome due to Fusobacterium nucleatum. *J Clin Microbiol*. **2003**, 41(*7*), 3445-8.

[145]. Zepeda-Rivera, M., Minot, S.S., Bouzek, H. et al. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature*, **2024**, 628, 424–432.

[146]. Ou S, Wang H, Tao Y, Luo K, Ye J, Ran S, Guan Z, Wang Y, Hu H, Huang R. Fusobacterium nucleatum and colorectal cancer:

From phenomenon to mechanism. *Front Cell Infect Microbiol*. **2022**, 12, 1020583.

[147]. Garcia-Vello, P.; Di Lorenzo, F.; Lamprinaki, D.; Notaro, A.; Speciale, I.; Molinaro, A.; Juge, N.; De Castro, C. Structure of the O-Antigen and the Lipid A from the Lipopolysaccharide of Fusobacterium Nucleatum ATCC 51191. *ChemBioChem* **2021**, 22 (*7*), 1252–1260.

[148]. Lamprinaki, D.; Garcia-Vello, P.; Marchetti, R.; Hellmich, C.; McCord, K. A.; Bowles, K. M.; Macauley, M. S.; Silipo, A.; De Castro, C.; Crocker, P. R.; et al. Siglec-7 Mediates Immunomodulation by Colorectal Cancer-Associated Fusobacterium Nucleatum Ssp. Animalis. *Front. Immunol*. **2021**, 12, 1–14.

[149]. Boelaars, K.; van Kooyk, Y. Targeting Myeloid Cells for Cancer Immunotherapy: Siglec-7/9/10/15 and Their Ligands. *Trends in Cancer*, **2024**, 10 (*3*), 230–241..

[150]. O'Reilly, M. K.; Paulson, J. C. Multivalent Ligands for Siglecs, 1st ed.; *Elsevier Inc*, **2010**, 478.

[151]. Gonzalez-Gil, A.; Schnaar, R. L. Siglec Ligands. *Cells*, **2021**, 10(*5*), 1–21.

[152]. Huifeng Yu, Anabel Gonzalez-Gil, Yadong Wei, Steve M Fernandes, Ryan N Porell, Katarina Vajn, James C Paulson, Corwin M Nycholat, Ronald L Schnaar, Siglec-8 and Siglec-9 binding specificities and endogenous airway ligand distributions and properties, *Glycobiology*, **2017**, 27(*7*), 657–668,

[153]. Choi, H.; Ho, M.; Adeniji, O. S.; Giron, L.; Bordoloi, D.; Kulkarni, A. J.; Puchalt, A. P.; Abdel-Mohsen, M.; Muthumani, K. Development of Siglec-9 Blocking Antibody to Enhance Anti-Tumor Immunity. *Front. Oncol*., **2021**, 11, 1–10.

[154]. Cheong, K. A.; Chang, Y. S.; Roh, J. Y.; Kim, B. J.; Kim, M. N.; Park, Y. M.; Park, H. J.; Kim, N. D.; Lee, C. H.; Lee, A. Y. A Novel

Function of Siglec-9 A391C Polymorphism on T Cell Receptor Signaling. *Int. Arch. Allergy Immunol.* **2011**, 154 (*2*), 111–118.

[155]. Ibarlucea-Benitez, I.; Weitzenfeld, P.; Smith, P.; Ravetch, J. V. Siglecs-7/9 Function as Inhibitory Immune Checkpoints in Vivo and Can Be Targeted to Enhance Therapeutic Antitumor Immunity. *Proc. Natl. Acad. Sci*., **2021**, 118 (*26*), e2107424118.

[156]. Atxabal, U.; Nycholat, C.; Pröpster, J. M.; Fernández, A.; Oyenarte, I.; Lenza, M. P.; Franconetti, A.; Soares, C. O.; Coelho, H.; Marcelo, F.; et al. Unraveling Molecular Recognition of Glycan Ligands by Siglec-9 via NMR Spectroscopy and Molecular Dynamics Modeling. *ACS Chem. Biol*. **2024**, 19 (*2*), 483–496.

[157]. 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.

[158]. Meyer, S. J., Linder, A. T., Brandl, C., & Nitschke, L. B. Cell Siglecs-News on Signaling and Its Interplay With Ligand Binding. *Frontiers in immunology*, **2018**, 9, 2820.

[159]. Zhang, P., Lu, X., Tao, K., Shi, L., Li, W., Wang, G., & Wu, K. Sig-lec-10 is associated with survival and natural killer cell dysfunction in hepatocel-lular carcinoma. *The Journal of surgical research*, **2015**, 194(*1*), 107–113.

[160]. Barkal, A.A., Brewer, R.E., Markovic, M. et al. CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature*, **2019**, 57, 2392–396.

[161]. Hazra R, Chattopadhyay S, Mallick A, Gayen S, Roy S. Unravelling CD24-Siglec-10 pathway: Cancer immunotherapy from basic science to clinical studies. *Immunology*. **2024**.

[162]. Li, X., Tian, W., Jiang, Z. et al. Targeting CD24/Siglec-10 signal pathway for cancer immunotherapy: recent advances and future directions. *Cancer Immunol Immunother*., **2024,** 73, 31.

[163]. Smith, B.A.H., Bertozzi, C.R. The clinical impact of glycobiology: targeting selectins, Siglecs and mammalian glycans. *Nat Rev Drug Discov.,* **2021***,* 20, 217–243

[164]. Duan, S.; Paulson, J. C. Siglecs as Immune Cell Checkpoints in Disease. *Annu. Rev. Immunol*. **2020**, 38, 365–395.

[165]. Antigenic liposomes displaying CD22 ligands induce antigen-specific B cell apoptosis Matthew S. Macauley, Annette von Drygalski, James C. Paulson Published June 3, 2013 *J Clin Invest*., **2013**;123(*7*), 3074-3083.

[166]. DarongYang., The Potential of Siglec Receptors in Cancer Immunotherapy *Translational Surgical Oncology*, 1(3), 76-8976.

[167]. Joelle Texereau, Jean-Daniel Chiche, William Taylor, Gerald Choukroun, Beatrice Comba, Jean-Paul Mira, The Importance of Toll-Like Receptor 2 Polymorphisms in Severe Infections, *Clinical Infectious Diseases*, **2005**, 41(*7*), S408–S415.

[168]. Matamoros-Recio, A.; Franco-Gonzalez, J. F.; Perez-Regidor, L.; Billod, J. M.; Guzman-Caldentey, J.; Martin-Santamaria, S. Full-Atom Model of the Agonist LPS-Bound Toll-like Receptor 4 Dimer in a Membrane Environment. Chem. - *A Eur. J*. **2021**, 27 (*62*), 15406–15425.

[169]. Yang, RB., Mark, M., Gray, A. et al. Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature*, **1998,** 395, 284–288.

[170]. Kang, J. Y.; Nan, X.; Jin, M. S.; Youn, S. J.; Ryu, Y. H.; Mah, S.; Han, S. H.; Lee, H.; Paik, S. G.; Lee, J. O. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. *Immunity* **2009**, 31(*6*), 873–884.

[171]. Gautam, Jitendra K. et al., Structural and Functional Evidence for the Role of the TLR2 DD Loop in TLR1/TLR2 Heterodimerization and Signaling. *Journal of Biological Chemistry*, **2006**, 281(40), 30132 – 30142.

[172]. Nackiewicz, D., Dan, M., He, W. et al. TLR2/6 and TLR4-activated macrophages contribute to islet inflammation and impair beta cell insulin gene expression via IL-1 and IL-6. *Diabetologia*, **2014,** 57, 1645–1654.

Acknowledgements

These are going to be the most polyglot acknowledgments ever.

First of all, I'd like to thank everyone who walked with me during the past three years of my PhD. One way or another, you have all contributed to helping me reach this point. I'd like to begin with my warmest thanks to my supervisor: Alba Silipo. Grazie, Alba. Come advisor e come persona. Mi hai insegnato a crescere e maturare, mi hai supportato, e con serietà, firmezza, confronti e anche tante risate, mi hai aiutato a prendere le scelte più approriate. Grazie per il tuo umore, ma soprattutto per la tua empatia nei momenti di difficoltá. Prometto non dimenticare mai più l'effetto esoanomerico. A Roberta Marchetti, per essere stata non solo co-supervisor, ma anche un supporto durante tutto questo tempo, ho finito per capirti a volte solo da uno sguardo; grazie. A Tony, per spronare a tutti a formarsi, conoscere mondo e a prendere le cose con amore. A Flaviana, per la tua resilienza, per tante caramelle prese dallo studio e per aver sempre quell sorriso che ti caratterizza sia come persona che come scienziata.

Marta, estos tres años sin ser consciente de cuánto puedo ser negada para las plantas no hubiesen sido los mismos. Gracias por escucharme, por haber permanecido juntas todo este tiempo y por tu calma. Gracias por crecer juntas, te voy a llevar siempre conmigo. Fuiste y eres casa. Kiwi, Lima y yo te vamos a recibir siempre. Molly, I hope you can read this big thank you with a warm cup of tea and biscuits while we laugh for no particular reason.

Continuo per il satellite: Imma, Anna, Adele. Grazie per essere state le sorelle grandi di SSCN, tanti caffè e consigli. Alessandro, sei una roccia. Sei molto più che un collega di lavoro, ma sei un amico. Grazie per tutto questo tempo, risate e anche piantini; Stefania, che sarebbe stato di me senza tutti i tuoi consigli. Quante volte mi hai dato calma e supporto. Con chi mostrerò il mio napoletano?

Ti voglio bene. Raffaele e Silvia, grazie per essere sempre così tranquilli e divertenti. Ferran: acuérdate de que se dice afilalápiz!, gracias por tu humor. Angela M.: gracias. Cristina, sei stata sempre una persona speciale per me, un esempio a seguire. Pia, grazie per i coonsigli, anche personali, durante questa tesi. Marcello grazie, per essere cosi spontaneo e naturale sempre, sei un piccolo gioiello. Alessia, sei davvero una persona speciale, grazie per la tua sensibilità di tutti i giorni, ti porto con me. Emmanuela, sei energia, è incredibile, un piccolo terremoto, grazie. Roberta e Ornella, le shigellose. Non so se avrei potuto arrivare a questo punto senza le mille risate con voi. Luca, dammi un poco della tua calma che ti caratterizza Roberta piccolina: e io cosa avrei fatto senza le conversazioni infinite.; grazie per tanta dolcezza. Se manca qualcuno, vi ringrazio e vi chiedo scusa. Tutti, davvero tutti avete contribuito a questo mio traguardo. Napoli è diventata casa grazie a voi, SSCN, e anche a tutti gli amici e amiche, senza di voi tutto ciò non sarebbe stato lo stesso. Luna, amica mia, mi hai dato davvero l'aria più fresca quando più ne avevo bisogno. So che ci rivedremmo.

To all my ESRs from GLYTUNES, what an amazing life journey we've passed through all together.

A Sonsoles, gracias por tu cariño, me has tratado como una segunda madre. A todos los chicos del CSIC de Madrid: Marina, Joan, Olmo, Irene: sois increíbles.

A todos os meus amigos de sempre: María Bugallo, Lara, María Bastida, Iris, Iván, Rubén, Carla, Uxía, Suevia, Micho. Todos vós. As conversacións nos bares cando vos visito, o sentir que sodes casa dende e pra sempre. Síntome como cando tomas unha sopa ben quentiña nunha xornada ben fría. Axudástesme a crer en min, a ver a vida con outros ollos, a saír do habitual e coñecer novas metas, a motivarme a saír dos meus proprios bucles cando máis o precisei. Grazas. Por todo, pero sobre todo por ter tido parte deste meu logro.

A mi pareja, Nacho. Una de las personas que mejor me conoce. El pragmatismo que te define y la fuerza que tienes para buscar soluciones a todo me hacen tenerte como modelo, pues me ayudas a crecer. Por si fuera poco, también eres capaz de hacerme reír incluso en los momentos más inesperados. Me haces feliz. Tú siempre has creído en mí. Eres hogar, eres roca, eres seguridad. Gracias por llegar cuando más lo necesitaba y por tu apoyo incondicional, a sabiendas de que estos momentos de escritura sumados a la distancia no hayan sido nada fáciles para ninguno. Te mereces todo lo bueno. Gracias por ser y estar.

Ahora vienen los agradecimientos a mi familia: mi abuelo, mi abuela y mi tía. No hubiese podido llegar a donde estoy hoy si no hubieseis sido las piezas fundamentales de mi vida. Habéis creído en mí cuando nadie más lo hacía. Me motivasteis a que pudiese escoger un destino de vida del cual sentirme partícipe; que fuese mío. ¿Quién nos iba a decir a nosotros que hoy iba a escribir unos agradecimientos de una tesis doctoral después de todo? Qué fuertes que sois. Merecéis no solo un párrafo, si no un libro donde pueda contar por todo lo que hemos pasado juntos. Os quiero. A Kiwi y Lima también, aunque no sepan leer. Han sido mi motivación diaria en Nápoles.

Por último, y no por ello menos importante: a mamá. A pesar de que no esté con nosotros desde hace unos años, ha sido y es mi motor. Me hubiese gustado que todo esto lo hubiese podido disfrutar conmigo, pero me ha dado una fuerza inmensa.

Mamá, esta tesis va por ti. Te quiero para siempre.