Università degli Studi di Napoli Federico II

Ph.D. Program in

Computational and Quantitative Biology

XXXVIII Cycle

Thesis for the Degree of Doctor of Philosophy

When *Non-Self* Looks Like *Self*: Decoding Lectins–Ligand Recognition across Endogenous and Exogenous Glycans

by

Alessandro Antonio Masi

Advisor: Prof. Alba Silipo

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

IMPERATOR ROMANORVM SEMPER AVGVST

Alla memoria delle vittime innocenti di Gaza, affinché il dolore che hanno subito non sia mai dimenticato e il loro grido di dolore resti inciso nella coscienza dell'umanità.

La scienza resti sempre al servizio della vita, mai della distruzione.

IMPERATOR ROMANORVM SEMPER AVGVST

Ph.D Thesis presented
for the fulfilment of the Degree of Doctor of Philosophy in Computational and Quantitative Biology

When *Non-Self* Looks Like *Self*: Decoding Lectins–Ligand Recognition across Endogenous and Exogenous Glycans

by

Alessandro Antonio Masi

October 2025

Approved as to style and content by

Prof. Alba Silipo, Advisor

Università degli Studi di Napoli Federico II
Ph.D Program in Computational and Quantitative Biology
XXXVIII cycle – Chairman; Prof. Michele Ceccarelli

IMPERATOR ROMANORVM SEMPER AVGVSTVS FRIDERICVS

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 30th, 2025

Alessandro Antonio Masi 

IMPERATOR ROMANORVM SEMPER AVGVSTVS FRIDERICVS

ContentsList of Acronyms i

The following acronyms are used throughout the thesis.

AA	Acetylated Alditols
AMG	Acetylated Methyl Glycosides
AMR	Anti-Microbial Resistance
CSP	Chemical Shift Perturbation
CSM	Chemical Shift Mapping
CRD	Carbohydrate Recognition Domain
Cer	Ceramide
CNS	Central Nervous System
ConA	Concanavalin A
COSY	Correlation Spectroscopy
CRM	Cross-Reactive Material
DFT	Density Functional Theory
EP	Evolutionary Programming
EPS	Exopolysaccharide
ESI	Electrospray Ionization
Fab	Fragment Antigen-Binding
Fc	Fragment Crystallizable
FT-ICR	Fourier Transform Ion Cyclotron Resonance
GBPs	Glycan-Binding Proteins
GBS	Guillain–Barré Syndrome
gg	*gauche–gauche*
gt	*gauche–trans*
HF	Hartree–Fock
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single Quantum Correlation
Hz	Hertz
Ig	Immunoglobulin
ITAM	Immunoreceptor Tyrosine-Based Activation Motif
ITC	Isothermal Titration Calorimetry
ITIM	Immunoreceptor Tyrosine-Based Inhibitory Motif
LGA	Lamarckian Genetic Algorithm
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
mAbs	Monoclonal Antibodies
MALDI	Matrix-Assisted Laser Desorption/Ionization

MAMP	Microbe-Associated Molecular Pattern
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MS	Mass Spectrometry
NK	Natural Killer
nMS	Native Mass Spectrometry
Nm	*Neisseria meningitidis*
Ng	Neisseria gonorrhoeae
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OAc	O-Acetylated
PAMPs	Pathogen-Associated Molecular Patterns
PBC	Periodic Boundary Conditions
PCP	Phenol–Chloroform–Petroleum Ether
PDB	Protein Data Bank
PRRs	Pattern-Recognition Receptors
QM	Quantum Mechanics
RCC	Renal Cell Carcinoma
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
ROESY	Rotating-Frame Overhauser Effect Spectroscopy
SAMPs	Self-Associated Molecular Patterns
Siglec	Sialic Acid-Binding Immunoglobulin-Like Lectin
SNFG	Symbol Nomenclature for Glycans
STD	Saturation Transfer Difference
TFA	Trifluoroacetic Acid
TLRs	Toll-Like Receptors
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse Relaxation Optimized Spectroscopy
ToF	Time-of-Flight

The following abbreviations for amminoacid/sugar are used throughout the thesis.

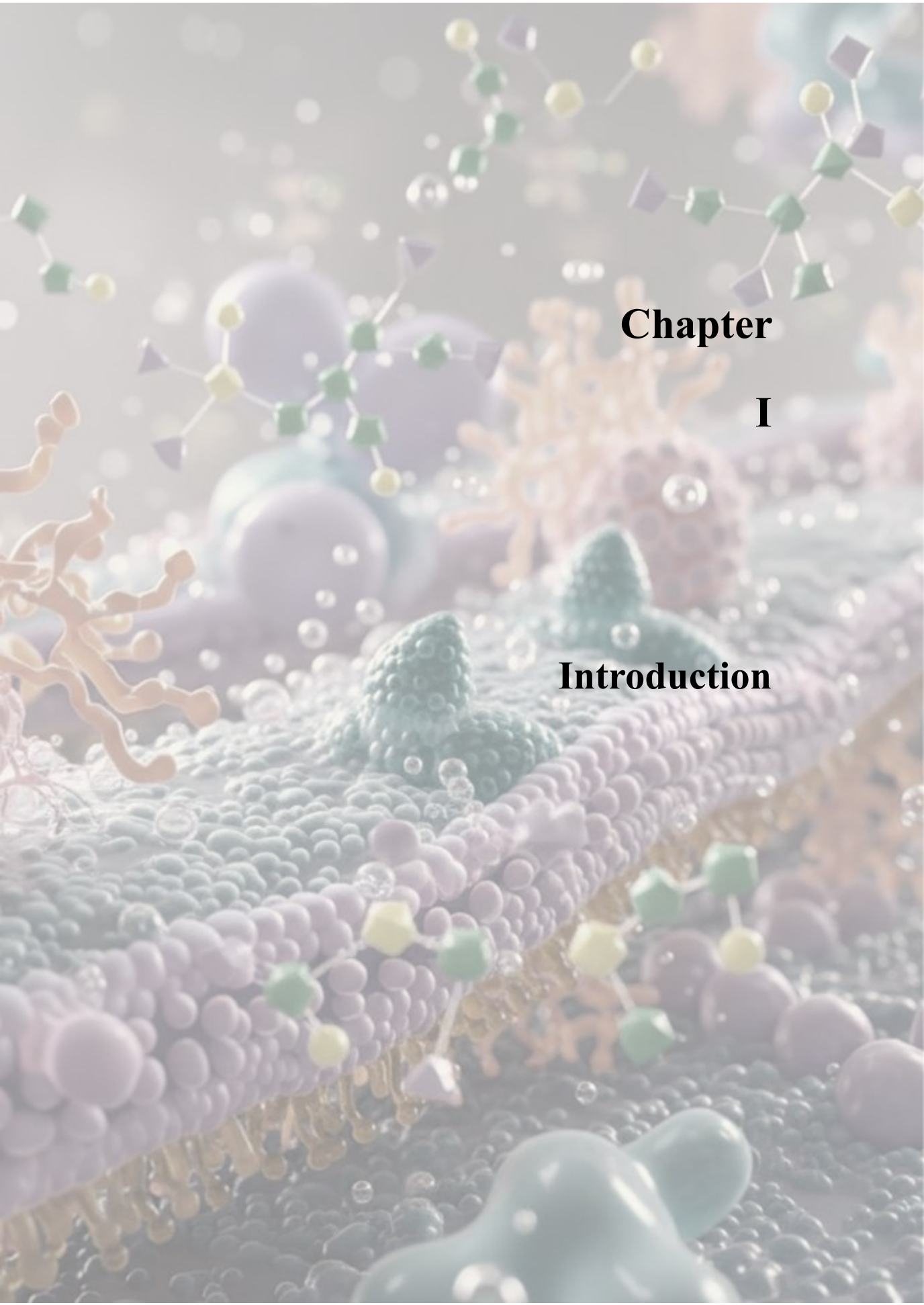
Glc	Glucose
GlcNAc	N-acetylglucosamine
Man	Mannose
Gal	Galactose
GalNAc	N-acetylgalactosamine
Neu5Ac	N-acetylneuraminic acid (5-N-acetylneuraminic acid)
Sia	Sialic acid
Neu5Gc	N-glycolylneuraminic acid
Kdn	2-keto-3-deoxy-D-glycero-D-galacto-nononic acid
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
Hep	L-glycero-D-manno-heptopyranose
sLe^x	Sialyl Lewis X
AAT	2-acetamido-4-amino-2,4,6-trideoxy-α-L-fucose
Asn	Asparagine
Fuc	Fucose
Ser	Serine
Thr	Threonine
Arg	Arginine
Lys	Lysine
Tyr	Tyrosine
Glu	Glutamic acid
Trp	Tryptophan
Asp	Aspartic acid
Ala	Alanine
Ile	Isoleucine

Abstract

Glycans are central immune modulators and are principal communicators of the host–microbe interface. Sialylated glycoconjugates such as gangliosides and bacterial lipooligosaccharides (LOS) are some that surface as principal drivers of *self/non-self* discrimination. Recognition by sialic acid-binding immunoglobulin-like lectins (Siglecs), particularly Siglec-7, is an evolutionary immune checkpoint mechanism hijacked by host and pathogens. This thesis investigates, at the molecular level, structural and functional binding interactions of Siglec-7 with a range of endogenous and exogenous sialoglycans, most notably disialylated gangliosides and bacterial LOS from *Neisseria* and *Fusobacterium* species. A multidisciplinary approach including glycan synthesis, structural biology (NMR, mass spectrometry), thermodynamic characterization (ITC, fluorescence titrations), and computational modeling (docking and molecular dynamics) has been employed. Eliciting important findings are the elucidation of conformational requirements vital for Siglec-7 binding, selective recognition of cancer-associated gangliosides such as GD3 and DSGb5, and comprehension of mechanisms of bacterial mimicry via sialylated LOS to escape immune surveillance. Additionally, the study examines the interaction of LOS core epitopes with monoclonal antibody 2C7, offering structural rationale of relevance to vaccine and therapeutic design. In summary, this thesis offers novel molecular insights into immune modulation by glycans and demonstrates the therapeutic potential of inhibiting glyco-immune checkpoints in infectious diseases as well as cancer.

Keywords: Computational chemistry, Sialoglycans, Siglecs, Molecular recognition, NMR, nMS

AVGVSTVS FRIDERICVS IMPERATOR ROMANORVM SEMPER



Chapter I

Introduction

I. Introduction

1.1 Glycans: Structure, Classification and Biological Function in Immune Recognition

Glycans, often referred to as carbohydrates, are complex biomolecules composed of monosaccharide units that are linked via glycosidic bonds. Alongside proteins, nucleic acids and lipids, they are one of the four major classes of biological macromolecules and are characterized by extraordinary structural diversity. This heterogeneity arises from the variety of monosaccharide building blocks, the ability to form linear or branched chains, and the different types of glycosidic linkages and non-carbohydrate modifications they can incorporate (e.g. acetyl, sulphate and phosphate groups)[1].

In biological systems, glycans rarely occur as free entities. They are typically found covalently attached to proteins (glycoproteins) or lipids (glycolipids), forming what are known as glycoconjugates, which localize predominantly at the cells surfaces. It is evident that this strategic positioning makes them essential players in cell–cell communication, receptor recognition, and modulation of the immune system[1, 2].

The glycan layer coating the surface of eukaryotic cells – known as the glycocalyx – acts as a molecular interface with the environment and plays a pivotal role in recognition events involving pathogens and immune cells[3]. Glycosylation, the enzymatic process by which glycans are attached to proteins or lipids, is not simply a structural decoration. Their importance has been demonstrated to exert a critical influence on a number of key biological processes, including the folding of proteins, their solubility, resistance to proteolytic degradation, and biological activity[4, 5]. For instance, glycosylation is a crucial factor in dictating the stability and half-life of immunoglobulins, and it is imperative for their effector functions, including complement activation and antibody-dependent cell-mediated toxicity[6].

Structurally, glycans are composed of monosaccharides residues—such as glucose, galactose, mannose, fucose, N-acetylglucosamine (GlcNAc), and N-acetylgalactosamine (GalNAc)—which are covalently linked via glycosidic bonds to form oligosaccharide or polysaccharide chains.

Each glycan chain can be classified based on the type of linkage between monosaccharides (e.g., α- or β-anomers), the position of linkage (e.g., 1→2, 1→3, 1→4, 1→6), and the presence of branching points.

This leads to a complex, branched architecture in which a single glycan core can be diversified into a multitude of glycoforms through sequential and variable monosaccharide additions.

A typical example of glycan diversity is the N-linked glycan core structure, $(GlcNAc)_2(Man)_3$, which serves as a universal scaffold that can be further elaborated into *high-mannose*, *hybrid*, or *complex* types (Figure 1A). These variations are the result of the differential addition of monosaccharides, which creates distinct glycoforms with unique biological properties.

In contrast to N-linked glycans, O-linked glycans exhibit greater structural variability and flexibility, both in terms of their core structures and patterns of elongation. O-glycosylation typically begins with the attachment of an N-acetylgalactosamine (GalNAc) residue to the hydroxyl group of Serine (Ser) or Threonine (Thr) side chains in the polypeptide backbone. This initial monosaccharide linkage marks the starting point of what is commonly referred to as mucin-type O-glycosylation, due to its abundance in mucins and other secreted or membrane-associated glycoproteins. These structures are commonly classified into eight distinct types, designated as core 1 to core 8 (Figure 1B). Each of these structures differs in their linkage and elongation patterns[7, 8]. These cores can then be extended with additional sugars, such as galactose (Gal), glucosamine (GlcNAc), fucose (Fuc), and sialic acid (Sia), contributing to high variability and tissue-specific expression patterns[9, 10]. Moreover, glycans—whether N- or O-linked—can be linearly elongated or branched,

with branching points often occurring at Man or GlcNAc residues. This branching architecture increases the valency and avidity of interactions with glycan-binding proteins, such as lectins, thereby enhancing the specificity and strength of molecular recognition events[9]. Glycans frequently incorporate non-carbohydrate substituents such as acetyl, sulphate, or phosphate groups, that possess the characteristic to modulate not only the chemical properties (e.g., charge, polarity), but also their conformational behaviour and binding profiles within biological systems[11].

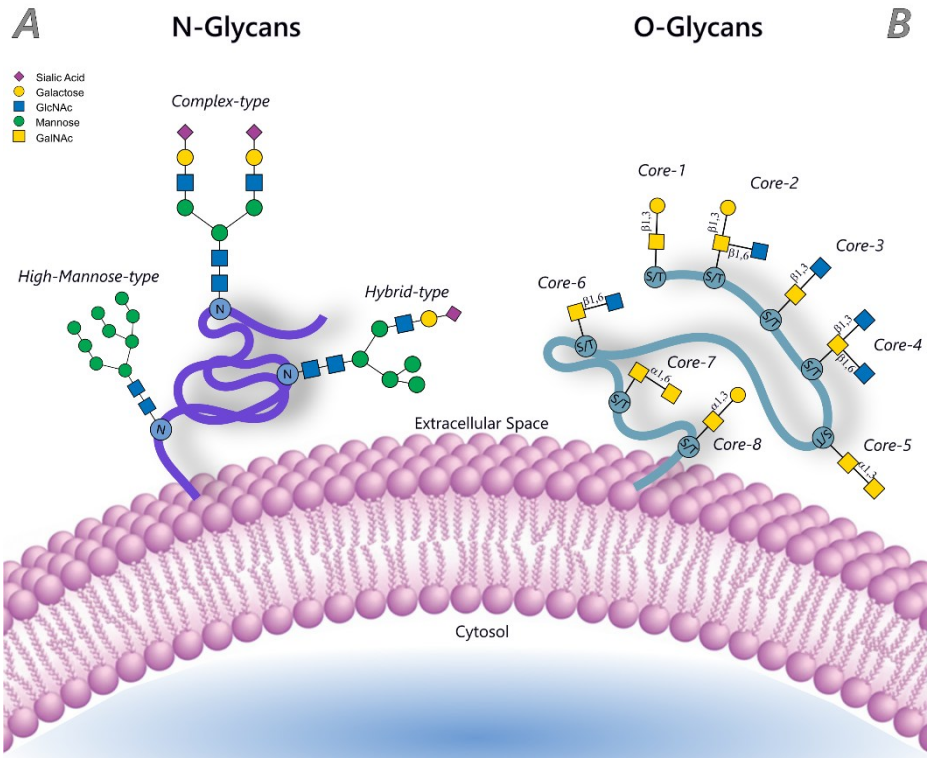


Figure I.1. N-glycan/O-glycan core structures and classification. A) Biosynthesis of N-glycans gives rise to three main structures: high-mannose, complex, and hybrid glycans. High-mannose forms, characterized by extensive mannose residues, are frequently displayed on tumor cells. Complex glycans are branched and typically terminate with α2,6- or α2,3-linked sialic acids on galactose, mediating lectin recognition. Hybrid glycans combine features of both high-mannose and complex types and are linked to immune regulation and disease states. B) Core mucin-type glycans are defined by a GalNAc residue attached to Ser (S) or Thr (T) residues.

1.1.1 Sialylated Glycans

A particularly important family of sugar residues is that of sialic acids (Sias), a group of negatively charged nine-carbon sugars typically found at the terminal positions of glycan chains on glycoproteins and glycolipids[8]. In humans, the most prevalent form is N-acetylneuraminic acid (Neu5Ac), followed by N-glycolylneuraminic acid (Neu5Gc) and 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN)[12]. Sialic acids are typically linked via α2,3 or α2,6 bonds to galactose (Gal) or GlcNac, and in case of polysialylated glycans via α2,8 or α2,9 linkages, established by sialyltransferases, while their removal is mediated by sialidases or neuraminidases[13].

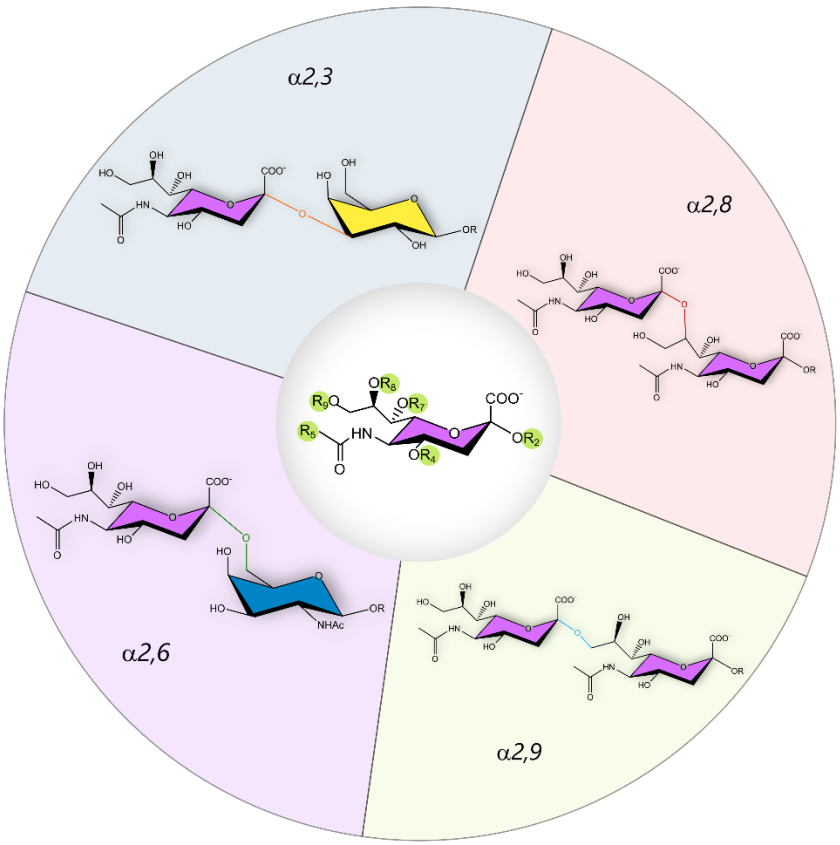


Figure I.2 Sialic acid chemical structure. In figure is shown one of the most predominant form, N-acetylneuraminic acid (Neu5Ac) and it's possible glycosidic linkages. At physiological pH, the carboxylate group confers the typical negative charge of sialic acids. Substitutions can occur at multiple positions: R1 may form intramolecular lactones or

intermolecular lactams at C-5; R2 can establish α-linkages with Gal (C-3/4/6), GalNAc (C-6), GlcNAc (C-4/6), other sialic acids (C-8/9), or 5-O-Neu5Gc; R4 may be H, acetyl, fucose, or galactose; R5 can host amino, N-acetyl, N-glycolyl, hydroxyl, O-acetyl- or O-methyl-glycolyl groups; R7 is either H or acetyl; R8 may be H, acetyl, methyl, sulfate, another sialic acid, or glucose; R9 can carry H, acetyl, lactyl, phosphate, sulfate, or sialic acid.

The terminal position and negative charge of sialic acids confer a significant regulatory capacity on these molecules, by modulate cell adhesion, receptor binding, and immune recognition, often functioning as ligands for lectins, including Siglecs and selectins[14]. Their distinctive biophysical characteristics render them pivotal mediators of host–pathogen interactions, as numerous viruses and bacteria utilize sialylated structures for attachment and entry into host cells[15]. It is noteworthy that humans have lost the capacity to synthesize Neu5Gc due to an ancestral mutation in the *CMAH* gene. However, Neu5Gc can still be metabolically incorporated from dietary sources such as red meat and cow's milk, and has been detected in various human tumours, including melanoma, colon, and breast cancer, where it contributes to immune modulation and tumour progression[16, 17]. In addition to their structural functions, sialic acids also serve as fundamental molecular markers of the "*self*" for the immune system. Their widespread presence at the terminal positions of glycans on the surface of host cells acts as Self-Associated Molecular Patterns (SAMPs), molecular signals that distinguish endogenous components from potential threats[18]. By presenting these sialylated motifs, healthy host cells signal immune tolerance, thereby preventing inappropriate activation of innate immune receptors. This immunological coding becomes particularly relevant in the context of host-pathogen interactions, where many microbes mimic or acquire sialylated structures to camouflage themselves as "*self*" and evade immune surveillance. This dual role of sialic acids, both as mediators of recognition and as shields against immune attack, lays the foundation for understanding how innate immune

receptors, such as Pattern Recognition Receptors (PRRs), respond to glycan-based signals in distinguishing between *self* and non-*self*[19].

During immune surveillance, PRRs expressed on the surface of innate immune cells, including macrophages, dendritic cells, and neutrophils, recognize conserved microbial structures known as Microbe-Associated Molecular Patterns (MAMPs). A significant portion of these MAMPs are glycan-based and this recognition activates signaling cascades that promote inflammatory responses, cytokine release, and the recruitment of adaptive immunity[20, 21]. Selective expression and modification of glycans during embryonic development and tissue differentiation also underlie their role as developmental markers. Changes in glycan patterns are indicative of cellular transformation, stress, or malignancy, making them valuable biomarkers in diagnostics and targets for therapeutic intervention. However, this system can also be disrupted. Cancer cells often present abnormal glycosylation patterns, particularly altered sialylation, which help them escape immune surveillance, spread more easily, and resist programmed cell death[22]. In addition, certain pathogens have been observed to mimic host glycans, a phenomenon referred to as "molecular mimicry," as a strategy to evade detection by the immune system[23].

1.2 Siglecs as Immune Glycan-Binding Lectins: Recognition of Sialylated Self and Non-Self glycoconjugates in pathogenic events

Glycan-binding proteins (GBPs) are a diverse group of carbohydrate-recognizing molecules that mediate essential biological processes by binding to glycans exposed on the cell surface. Their interactions are involved in a wide range of cellular events, including cell-cell communication, immune modulation, and inflammatory responses[24-26]. Within this broad category, lectins represent a particularly important class of GBPs. These proteins are found across all kingdoms of life and engage in glycan recognition to orchestrate immune responses, facilitate

intercellular adhesion, and modulate signaling cascades. Based on structural and functional criteria, lectins are typically divided into several families, such as C-type, P-type, S-type (galectins), and I-type lectins, the latter being characterized by the presence of immunoglobulin-like domains[27, 28]. Among the I-type lectins, a prominent subfamily is that of the sialic acid-binding immunoglobulin-like lectins, commonly referred to as Siglecs. These are predominantly expressed on the surface of cells of the immune system, including natural killer cells, monocytes, and subsets of lymphocytes. Siglecs are distinguished by their specific affinity for sialylated glycans, a feature that enables them to finely discriminate between *self* and non-*self* structures. This recognition capacity positions them as key regulators of immune homeostasis, where they act both as sensors of endogenous sialic acid patterns and as mediators of tolerance or activation depending on context[29, 30]. In humans, 14 functional Siglec proteins capable of recognizing sialylated glycans have been identified, while the total number rises to 17 when including all members of the family described in hominoid primates[31]. Phylogenetically, the Siglec family is divided into two major groups: the conserved Siglecs—including Siglec-1 (sialoadhesin), Siglec-2 (CD22), Siglec-4 (myelin-associated glycoprotein), and Siglec-15—which are evolutionarily stable and have orthologs conserved across most mammalian species, and the CD33-related Siglecs, which have undergone rapid evolutionary diversification.

The CD33-related subgroup includes receptors such as Siglec-3 (CD33), Siglec-5, -7, -8, -9, -10, and -11, which share high sequence homology (50–99%) within their V-set domain but show considerable interspecies variability, particularly between rodents and primates[32].

These genes are organized in a tightly linked gene cluster on human chromosome 19q13.3–q13.4, reflecting their common evolutionary origin through gene duplication events. In murine models, the functional counterparts of human Siglecs can be identified based on expression profiles and immunological roles. For instance, Siglec-E, which is found

on murine neutrophils, monocytes, and dendritic cells, shows functional similarity to human Siglec-9. Likewise, Siglec-F, expressed primarily on eosinophils, is considered the closest functional homolog of human Siglec-8, despite notable sequence differences[33].

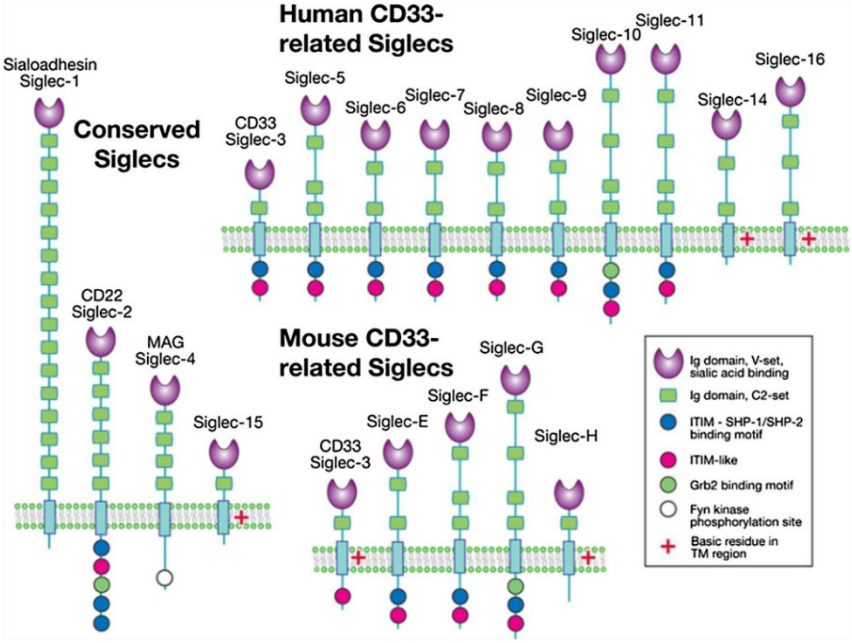


Figure I.3. Classification of Siglec proteins. Phylogenetic and structural classification of Siglecs into conserved and CD33-related subgroups in human and mice, highlighting domain composition, signalling motifs, and transmembrane features[32]. Image adapted from Ref.32.

In physiological contexts, Siglecs participate in fine-tuning immune cell activation, phagocytosis, cytokine release, and apoptotic clearance. These receptors act as sentinels of cellular self-identity and as checkpoints that limit excessive inflammation, thereby preventing autoimmunity[19, 33].

In pathological settings such as cancer or chronic infections, many Siglecs are co-opted by tumours or pathogens, which exploit their immunosuppressive signalling to escape immune surveillance[8, 33]. Over the past two decades, the Siglec family has garnered increasing attention as a therapeutic target, particularly in oncology and immunotherapy, due to their central role in immune modulation and tolerance induction[34, 35]. Their

ligand specificity, structural variability, and differential expression across leukocyte populations position them as promising candidates for immune checkpoint inhibition or sialic-acid mimetic design[36-38].

1.2.1 Siglecs: Structure and Function

All Siglecs share a common architecture consisting of an N-terminal V-set immunoglobulin domain, which is responsible for sialic acid binding, followed by varying numbers (from 1 to 16) of C2-set Ig-like domains connected by disulphide bridge[27], a transmembrane segment, and a cytoplasmic tail often containing signalling motifs such as ITIM (Immunoreceptor Tyrosine-based Inhibitory Motif) or ITAM-like (Immunoreceptor Tyrosine-based Activator Motif) sequences[27, 29].

The N-terminal V-set domain is localized in the extracellular interface and represents the functional core of Siglec-mediated ligand recognition. This domain adopts a β-sandwich fold formed by antiparallel β-strands, structurally homologous to the variable region of immunoglobulins. This domain is composed of nine β-strands (designated A to G) arranged in a characteristic topology act as the primary binding site for sialylated ligands, facilitating selective recognition through the conservation of structural motifs[39]. Sialic acid has been observed to bind to a shallow groove located on the face of the F strand with the carboxylate group of the ligand forms a critical salt bridge with a highly conserved arginine (Arg) residue, essential for sialoglycan recognition[30].

Another key contributor to glycan recognition is the formation of CH–π interactions between aromatic side chains of the receptor and the axial C–H bonds of the glycan's hydrophobic face[40]. This interaction is made possible by a disulfide bond between the B and E β-strands, which maintains separation of the β-sheets and exposes aromatic residues on the A and G strands[41]. These exposed residues, particularly Trp and Tyr, form stabilizing contacts with the lateral glycerol chain and N-acetyl group of sialic acid[29]. Furthermore, the FG loop (also referred as CC′) not only

contributes to glycan binding but has also been shown to exhibit structural flexibility, particularly in the context of extended or branched glycan chains[41]. This adaptability enhances the capacity of Siglecs to engage ligands with variable topologies and degrees of sialylation. A distinguishing structural characteristic of the V-set domain of Siglecs is the presence of five surface-exposed loops, designated as BC, C′C″, C″D, DE, and FG. These loops collectively influence the configuration of the sialic acid binding site and contribute to ligand specificity by modulating binding affinity and selectivity. As demonstrated by case studies of Siglec-7 complexed with GT1b ganglioside, a significant conformational rearrangement of this loop occurs upon ligand engagement. This supports an induced-fit binding mechanism that facilitates the recognition of structurally diverse sialylated glycoconjugates, including glycolipids, mucins and microbial oligosaccharides[42].

The C2-set domains, while not directly involved in ligand binding, contribute to the overall stability and spacing of the extracellular portion of Siglecs. In sialoadhesin (Siglec-1), for instance, the extended array of 16 C2-set domains forms a rod-like structure of ~40 nm, projecting the binding site far from the cell membrane, an adaptation likely favouring trans interactions with sialylated ligands on opposing cells. The extracellular portion of Siglecs is connected to the intracellular region via a single-pass transmembrane domain. This architecture connects the extracellular domains to the cytoplasmic tail, which acts as an intracellular signal transducer. In most inhibitory Siglecs, the cytosolic tail contains one or more ITIMs. Upon ligand binding, these ITIMs are phosphorylated and recruit SH2 domain-containing tyrosine phosphatases, mainly SHP-1 and SHP-2, resulting in attenuation of activation signals in immune cells[43].

In contrast, a subset of Siglecs, such as Siglec-14, Siglec-15, and Siglec-16, do not contain intrinsic ITIMs, instead exerting activation functions through their association with adaptor proteins that harbor immunoreceptor

tyrosine-based activation motifs (ITAMs), such as DAP12. This interaction is mediated by electrostatic coupling between a positively charged residue within the transmembrane domain of Siglec and a negatively charged residue in the adaptor protein[44]. Through this configuration, these activating Siglecs can initiate downstream pro-inflammatory signalling cascades[29].

1.2.2 Recognition between Siglecs and sialoglycans

The specific recognition of sialylated glycans by Siglec molecules is facilitated by their unique interactions, which are primarily mediated by the N-terminal V-set Ig-like domain. These sialoglycans are present in abundance on the surface of vertebrate cells as terminal residues of N- or O-linked glycans, glycolipids, or glycosaminoglycans, and they play a key role in defining the molecular "*self*" landscape[19, 36]. A noteworthy aspect of Siglec biology is its capacity to engage in both *cis* and *trans* interactions with sialylated ligands.

In *cis* interactions, Siglecs bind to sialoglycans present on the same cell membrane where they are expressed. This phenomenon is attributable to the high local density of sialylated glycoconjugates on the surface of immune cells, which leads to the masking of the Siglec binding site. Masked Siglecs are conformationally restricted, thereby preventing them from interacting in a *trans* manner and preventing non-specific or aberrant signalling. This mechanism functions as a protective regulatory system, ensuring that Siglec-mediated inhibition is activated only in appropriate biological contexts[14]. Conversely, when sialylated ligands are present in high local density on an opposing surface, such as another cell or a microbial structure, trans interactions become predominant. These events

unmask Siglec receptors, allowing for productive activation and subsequent activation of downstream signalling pathways (Figure 4).

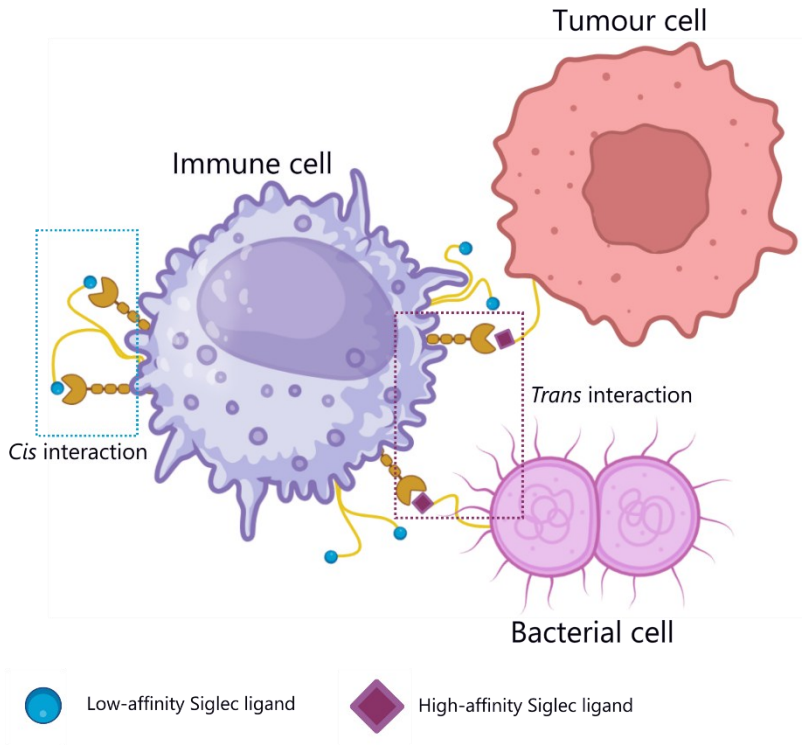


Figure I.4. Spatial relationship between Siglecs and their ligands. Siglecs can bind sialylated ligands present on the same cell membrane (*cis* interaction) or on the surface of other cells or particles (*trans* interaction).

The equilibrium between *cis* and *trans* interactions is modulated by numerous factors, including ligand affinity and valence, its spatial presentation, and glycan architecture (e.g., α2,3- vs. α2,6-linked sialic acid or polysialylation pattern)[45]. This dynamic interaction is biologically relevant in the context of *self* and *non-self* discrimination. Siglecs are believed to act as innate immune checkpoints by recognizing *self*-associated molecular patterns (SAMPs), which are typically rich in sialic acid. When such SAMP are detected on host cells, Siglec interaction attenuates immune activation, thereby maintaining homeostasis[18].

However, numerous pathogens have developed strategies to exploit this system. For instance, *Neisseria meningitidis*, *Fusobacterium nucleatum*, and *Campylobacter jejuni* decorate their surface lipooligosaccharides (LOS) or capsular polysaccharides (CPS) with sialic acids that mimic host glycans, thereby allowing them to evade immune detection by binding to inhibitory Siglecs on innate immune cells[46-48]. This form of molecular mimicry has been demonstrated to facilitate immune evasion, colonization, and persistence within the host[49, 50].

A comparable mechanism has been observed in the context of cancer biology. Tumour cells frequently undergo aberrant glycosylation, leading to the overexpression of sialylated glycoconjugates, including both glycoproteins and gangliosides. These structures decorate the cell surface and modulate the immune system. For example, disialylated gangliosides, such as GD3, DSGb3, and DSGb5, have been observed to be overexpressed in various malignancies, including melanoma, glioblastoma, breast cancer and renal cell carcinoma (RCC)[51, 52].

The expression of these disialylated gangliosides contributes to the creation of a SAMP on the cancer cells surfaces, effectively masking the cells from immune attack by exploiting *trans*-interactions with Siglec receptors. This strategy not only promotes tumour survival but also facilitates metastasis by impairing immune surveillance.

1.2.3 Siglec-7

Siglec-7 is a member of the CD33-related family of Siglecs, a group of type I transmembrane receptors predominantly expressed on cells of the innate immune system. Among these, Siglec-7 functions as an inhibitory receptor, primarily expressed on natural killer (NK) cells, but also detectable on monocytes, eosinophils, dendritic cells, and certain subsets of T lymphocytes[33]. From a structural standpoint, the extracellular region of Siglec-7 is composed of a V-set Ig-like domain, responsible for sialic acid recognition, followed by two C2-set Ig-like domains that act as spacers, maintaining the receptor in an extended conformation on the cell surface. The V-set domain shares the canonical fold found in other Siglecs, including a conserved Arg124 residue, which engages the carboxylate

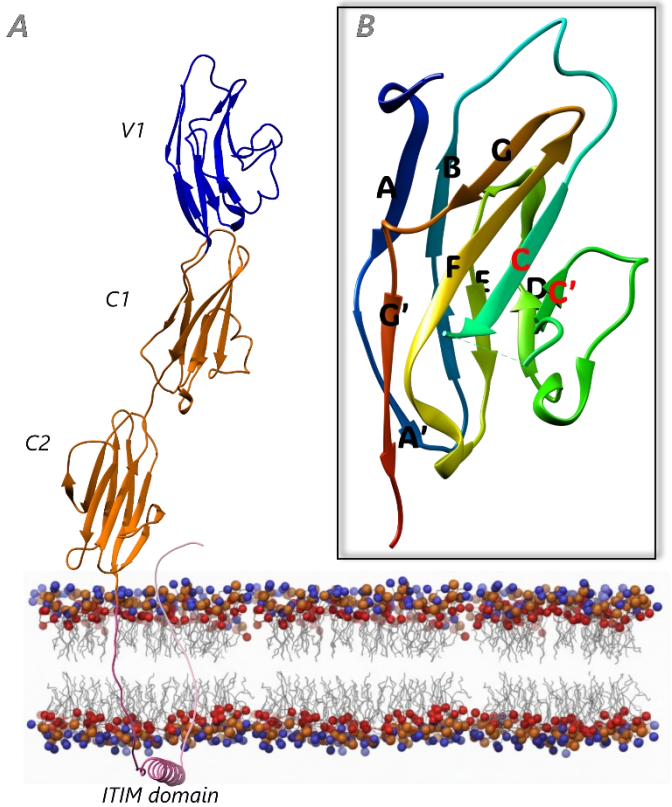


Figure I.5 3D representation of Siglec-7 structure (PDB ID:2HRL). A) The membrane is shown in grey (sticks and balls), the ITIM domain in pink, the two C2-set domains in orange, and the V-set domain in blue. B) Close-up of the V-set domain highlighting the β-strand nomenclature.

group of sialic acid via a salt bridge—an interaction that represents a conserved recognition mechanism across the Siglec family. Additional residues such as Tyr26, Lys131, and Trp132 contribute to ligand binding by establishing hydrogen bonds and hydrophobic contacts with the sialic acid moiety.

Crystallographic studies have revealed that Siglec-7 preferentially binds α2,8-linked[42] disialylated glycans, such as those found in gangliosides GD3 and DSGb5, which are terminally sialylated glycolipids abundantly expressed in tumour tissues. Upon ligand binding, the CC′ loop (residues Arg67 to Trp78) within the V-set domain undergoes a significant conformational rearrangement. This plasticity is thought to enable the receptor to accommodate structurally diverse sialoglycans across different biological contexts. Notably, a second ligand-binding site has been proposed, involving Arg67, which may act in tandem with the canonical Arg124 interaction site, further enhancing the receptor's ability to distinguish sialylated ligands[53].

The intracellular tail of Siglec-7 contains two ITIMs. Upon binding to sialylated ligands on target cells, signal transduction suppresses downstream activation pathways within NK cells, leading to inhibition of the release of cytotoxic granules and proinflammatory cytokines, such as IFN-γ[33]. However, this inhibitory signalling can be exploited by cancer cells that overexpress α2,8-linked disialylated gangliosides—notably GD3 and DSGb5—that effectively engage Siglec-7, thereby creating an immunosuppressive microenvironment[42]. Furthermore, pathogens such as *Campylobacter jejuni* and *Fusobacterium nucleatum* present sialylated structures that are capable of binding Siglec-7. In the case of *C. jejuni*, the bacterium expresses GQ1b-like lipooligosaccharide (LOS) epitopes that mimic host disialylated glycans and are recognized by Siglec-7, contributing to the immunopathogenesis of Guillain-Barré syndrome (GBS)[54]. Similarly, *F. nucleatum*, a known onco-bacterium, has been shown to engage Siglec-7 through its sialylated LOS, promoting local

immunosuppression and facilitating tumour progression in colorectal cancer and other malignancies[48].

Given its central role in immune checkpoint regulation, Siglec-7 has emerged as a promising target for immunotherapeutic intervention. Strategies aimed at blocking Siglec-7–ligand interactions, either through monoclonal antibodies or glycomimetic inhibitors, are currently under investigation for cancer immunotherapy and the modulation of pathogen-induced immunosuppression.

1.3 Gangliosides: Classification, Tissue Expression and Immunological Relevance

Gangliosides are a subclass of glycosphingolipids characterized by one or more sialic acid residues attached to an oligosaccharide chain that is covalently linked to a ceramide lipid anchor. These amphipathic molecules integrate a hydrophobic ceramide moiety that anchors into the plasma membrane and a hydrophilic carbohydrate domain that extends into the extracellular space.[22, 55, 56]

The substantial structural variability exhibited by lipid and glycan components renders gangliosides as structurally intricate molecules. The glycan portion is built upon a lactosylceramide core (Galβ1–4Glcβ1–1′Cer), which serves as the scaffold for further glycosylation and sialylation. The variability in the number, position, and linkage of sialic acids gives rise to the formation of several structurally defined series (e.g. Gala-, Neogala-, Globo-, Ganglio-)[57]. Among these, the ganglio-series is predominant in vertebrates and is defined by the conserved tetrasaccharide backbone (Galβ1–3GalNAcβ1–4Galβ1–4Glcβ1–Cer)[55], subdivided into three distinct categories: A, B, and C series. This classification is based on the number of sialic acids contained within the gangliosides. However, the nomenclature most widely adopted in the literature is the Svennerholm system[58], in which each ganglioside is assigned a shorthand code beginning

with “G” for ganglioside, followed by a capital letter denoting the number of sialic acid residues—0 if sialic acid is not present, M for mono-, D for di-, T for tri-, and so on—and a number reflecting its migration during thin-layer chromatography; For example, GM3 refers to a ganglioside with a single α2,3-linked sialic acid on a galactose residue, while GD3 carries two sialic acids in α2,3 and α2,8 bonds.

Gangliosides are expressed in a ubiquitous manner in mammalian tissues; however, their distribution is neither uniform or static. However, it is subject to stringent regulation, with distinct patterns observed for different tissue and cell types; This regulation is believed to reflect the unique functional demands of each biological context. In detail, they are present in the greatest quantity and structural variety in the central nervous system (CNS)[59], constitute approximately 10–12% of the total membrane lipids in neuronal cells and up to 5–10% of the total sialic acid content in the brain's gray matter[60]. It is important to note that their expression is associated with critical neurodevelopmental processes, including synaptogenesis, axon guidance, and neuronal differentiation making them indispensable for nervous system function[59].

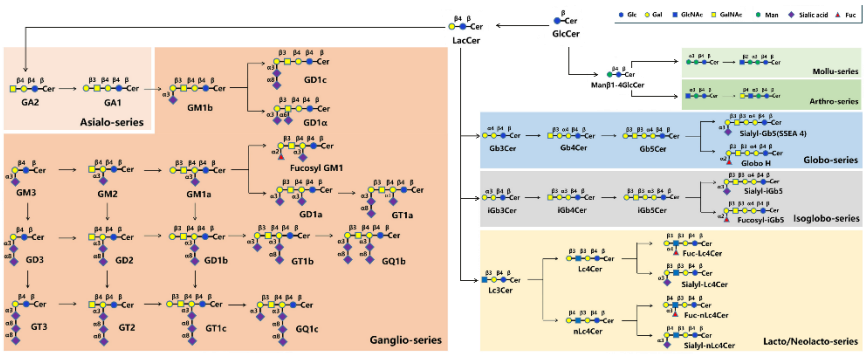


Figure I.6 Overview of glycosphingolipid (GSL) classes defined by their structural features. Seven major categories are distinguished according to the glycosyl group attached to the oligosaccharide chain. Abbreviations: Sa, sphinganine; FA, fatty acid; Cer, ceramide; Gal, galactose; Glc, glucose; GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; Man, mannose. Image adapted from ref. 52.

Outside of the nervous system, gangliosides are also found in epithelial cells, immune cells, and endothelial barriers, though at lower levels. In immune cells such as T lymphocytes and dendritic cells, gangliosides participate in membrane raft formation and receptor organization, processes essential for antigen presentation and immune signaling[8].

Pathological conditions are frequently associated with altered ganglioside expression. In cancer, for instance, a distinct glycosylation signature often emerges. Gangliosides such as GD3 and GD2 are overexpressed in melanoma, neuroblastoma, and small cell lung carcinoma, while DSGb5 has been detected in Renal Cell Carcinoma (RCC)[52]. From an immunological perspective, the sialylated termini of gangliosides serve as high-affinity ligands for glycan-binding proteins, especially Siglecs, Galectins, and Selectins. These interactions influence the thresholds of immune activation, often tipping the balance toward tolerance in the tumour microenvironment. The affinity and specificity of these interactions are not only dictated by the number and type of sialic acid residues but also by the conformation, flexibility, and spatial presentation of the glycan chain. This dynamic interplay regulates processes such as leukocyte trafficking, NK cell inhibition, and macrophage polarization.[31, 61]

1.4 Bacterial Glycoconjugates and the Architecture of LPS/LOS in Gram-Negative Bacteria

The outer surface of bacterial cells, particularly in Gram-negative species, is rich in structurally diverse glycoconjugates that play critical roles in mediating cell-cell communication, regulate the membrane charge, and host–pathogen interactions. These microbial glycoconjugates are teichoic acids (TAs), capsular polysaccharides (CPS), exopolysaccharides (EPS), and most notably lipopolysaccharides (LPS) and Lipooligosaccharides (LOS), which are unique to Gram-negative bacteria[62]. Due to their capacity to be identified by the host immune system as foreign entities, these

structures are designated as pathogen-associated molecular patterns (PAMPs). These PAMPs elicit the activation of innate immunity through the engagement of pattern recognition receptors (PRRs), including Toll-like receptors (TLRs), Lectins, and inflammasomes[63].

An area of particular interest in this work is the investigation of lipopolysaccharides (LPS) that are complex amphipathic macromolecules that constitute the major components of the outer leaflet of the outer membrane in Gram-negative bacteria. Functionally, they serve as structural elements essential for membrane integrity and as potent mediators in host–pathogen interactions. Their multifaceted role as PAMPs makes them central to the activation of the innate immune response in the host[64].

From a structural perspective, LPS are typically composed of three distinct regions (Figure 7).

- **Lipid A**: Represents the hydrophobic anchor of LPS, embedding the molecule into the bacterial outer membrane. It is composed of a bis-phosphorylated β-(1→6)-linked disaccharide of glucosamine, acylated by a variable number of saturated or hydroxylated fatty acids. The typical hexa-acylated lipid A, found in *Escherichia coli*, is considered the canonical endotoxin, capable of potently activating the TLR4/MD-2 complex on macrophages and other innate immune cells[62, 65].
- **Core oligosaccharide region (core OS)**: Covalently linked to lipid A, which is subdivided into inner and outercore regions. The innercore is relatively conserved among Gram-negative bacteria and typically contains L-glycero-D-manno-hepto-pyranose (L,D-Hep) residues and 3-deoxy-D-manno-oct-2-ulopyranosonic acid (Kdo), the latter being a hallmark sugar of Gram-negative organisms[66, 67].The outercore, in contrast, exhibits greater structural variability and is composed of hexoses such as glucose, galactose, and their amino derivatives. This region bridges the conserved inner structure with the more variable O-antigen.

- **O-Antigen**: When present, forms the outermost portion of LPS and consists of a linear or branched polysaccharide chain of repeating oligosaccharide units. These units are highly variable among bacterial strains, conferring antigenic diversity that facilitates immune evasion. The O-antigen length and structure are not only species-specific but also dynamically regulated in response to host immune pressure. Some Gram-negative pathogens, including *Neisseria meningitidis* and *Neisseria gonorrhoeae*, lack this portion, yielding a truncated LPS form known as lipooligosaccharide (LOS)[68]. The immunogenic portions of LPS, particularly the core OS and O-antigen, are critical in determining serotype and strain-specific immunity, making them attractive targets for vaccine development.

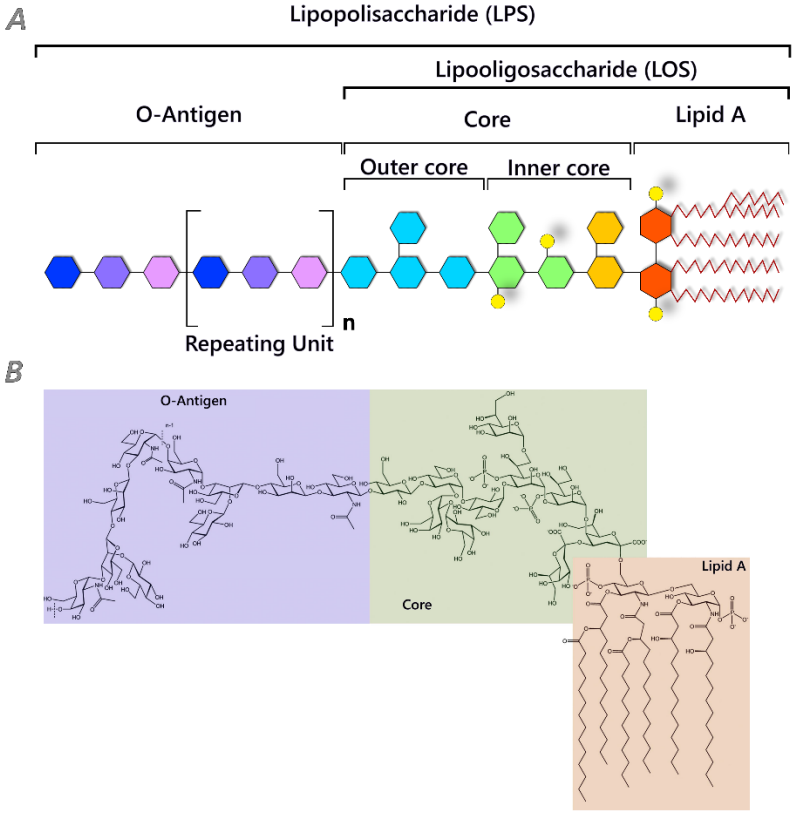


Figure I.7. Representation of LPS structure and composition in the Gram-negative membrane. A) LPS consists of three main components: lipid A, the core region, and the O-antigen; when the O-antigen is absent, the structure is referred to as LOS. B) an LPS structure example of E. coli strain O6.

1.5 Properties of Antibodies

Antibodies, also known as immunoglobulins (Ig), are soluble glycoproteins, essential in host defense and immune regulation.

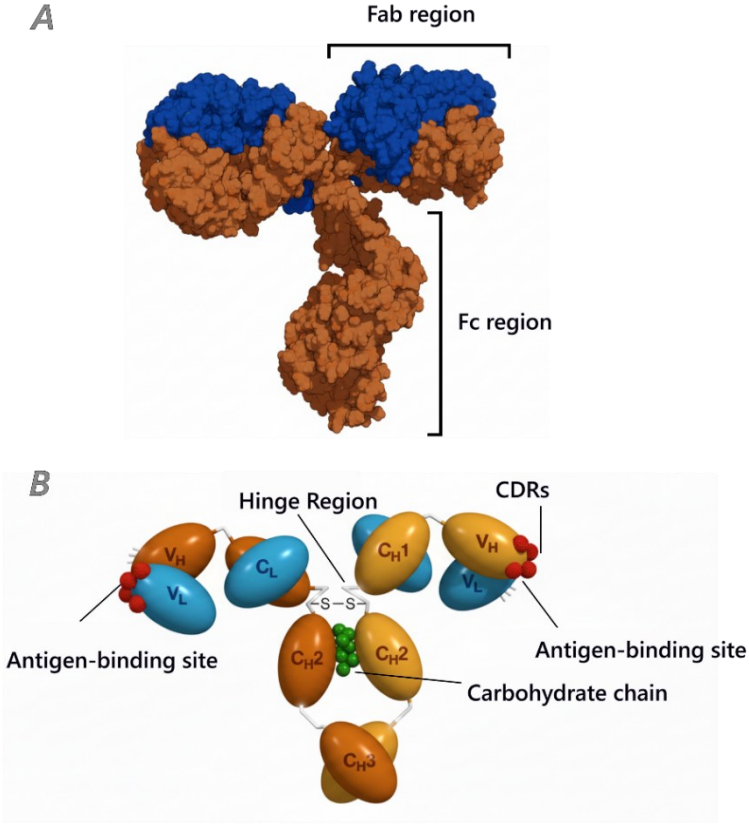


Figure I.8. Stucture of IgG antibody. The variable domains of the heavy and light chains (V_H and V_L) contain the complementarity-determining regions responsible for antigen recognition. The Fab portion includes the constant domains C_H1 and C_L, while the Fc region of the heavy chain is composed of C_H2 and C_H3.

Primarily produced by B lymphocytes, they can be secreted into extracellular fluids or exposed on immune cell surfaces (as with IgM) in response to antigen detection. Structurally, antibodies are Y-shaped heterotetramers composed of two heavy chains (H-chain, ~50kDa) and two light chains (L-chain, ~25kDa) that are covalently linked through interchain disulfide bonds. This quaternary structure defines the immunoglobulin's specificity and effector function.

The two arms of the Y, known as the Fab fragment (fragment antigen-binding), contain antigen-recognition sites. Each Fab fragment consists of the variable (V) and constant (C) regions of the heavy and light chains. The *Fc* (fragment crystallizable) region, in the central part of the Y, consists of the constant domains of the heavy chains. It is responsible for the antibody's biological activity, including its interaction with Fc receptors (*FcRs*) and activation of the complement cascade.

Each immunoglobulin chain is composed of multiple immunoglobulin domains, which are compact β-sheet folds of approximately 110 amino acids. The variable region of the heavy and light chains contains three hypervariable loops known as complementarity-determining regions (CDRs). These are designated CDR-H1, CDR-H2, and CDR-H3 for the heavy chain and CDR-L1, CDR-L2, and CDR-L3 for the light chain. These loops form the antigen-binding site at the surface of the Fab arms and are flanked by more conserved framework segments that preserve the structural integrity of the variable domains[69, 70].

CDR-H3 is particularly important due to its high degree of sequence and conformational variability, which makes it a key determinant of antigen specificity. Structural studies have demonstrated that alterations in CDR-H3's amino acid composition and length can significantly modify the shape and binding capacity of the antibody binding pocket[71].

Immunoglobulins are classified into different isotypes based on the constant region of their heavy chains. In humans, five major classes exist: IgG, IgA, IgM, IgD, and IgE. Each isotype plays a distinct role in immune responses. For instance, IgG is the most abundant in serum and mediates long-term immunity and opsonization.

Beyond isotypes, antibodies are also classified based on allotypes and idiotypes. Allotypes refer to allelic variants in the constant regions of

Ig genes, differing among individuals of the same species—these are relevant in contexts like transfusions or pregnancy. Idiotypes, on the other hand, are unique antigenic determinants within the variable region, defining the individual specificity of each antibody and often used in vaccine design and tumor diagnostics[72].

A major milestone in immunology was the development of monoclonal antibodies (mAbs), which are laboratory-produced antibodies that recognize a single epitope with high specificity. Introduced by Köhler and Milstein in 1975 via hybridoma technology[73], mAbs are generated by fusing an antigen-specific B cell with a myeloma cell to obtain an immortal cell line capable of continuous antibody production.

These antibodies are produced on a large scale and then purified for diagnostic and therapeutic applications. Among the potential immunotherapeutic applications of mAbs is their use in counteracting the global spread of antimicrobial-resistant (AMR). In this context, monoclonal antibodies (mAbs) have been shown to neutralize specific bacterial virulence factors without exerting selective pressure on the microbial genome, one of the main factors determining antibiotic resistance. This phenomenon can be attributed to their capacity for highly selective modes of action, which enable binding specifically to bacterial antigens such as surface proteins, adhesins, toxins, or polysaccharides. This precision not only limits off-target effects on the host microbiota but also reduces the risk of resistance development.

1.6 Objectives

This thesis aims to dissect, at the molecular level, the intricate recognition mechanisms that govern the interactions between host immune receptors and sialylated/non-sialylated glycoconjugates. Particular attention is given to the role of Siglec receptors and their engagement with endogenous (self) and exogenous (non-self) sialylated ligands, including human glycoconjugates and bacterial mimics.

Glycans are central mediators of host–microbe communication, shaping immune responses through structurally diverse and highly context-specific molecular codes. Among these, gangliosides and lipooligosaccharides (LOS) represent two classes of surface glycoconjugates—of human and bacterial origin, respectively—that modulate immune system. Notably, certain pathogenic bacteria such as *Neisseria gonorrhoeae*, *Neisseria meningitidis*, and *Fusobacterium nucleatum* have evolved the ability to express LOS/LPS that engaging inhibitory Siglecs and evading immune surveillance through a process of molecular mimicry. This work combines glycochemical characterization, structural techniques, biophysical, and computational approaches, integrating Nuclear Magnetic Resonance spectroscopy (NMR), Mass Spectrometry (MS), Isothermal Titration Calorimetry (ITC), and Molecular Dynamics (MD) Simulations experiments. The structural and thermodynamic parameters of Siglec–glycan interactions are investigated in both free-state and complex-bound conditions, aiming to clarify how fine structural features affect binding specificity and immunological outcome. The thesis results is structured into the following chapter:

- Chapter 3: *Investigation underlying the recognition between Siglec-7 and endogenous ligands.*

Gangliosides are endogenous sialylated glycosphingolipids involved in cell signaling and immune modulation. Among them, GD3 acts as a tumor-associated antigen and a high-affinity ligand for the inhibitory receptor Siglec-7 on NK cells. Through NMR, biophysical, and computational analyses, the interaction between Siglec-7 and disialylated gangliosides was characterized, highlighting the role of glycan topology and conformational rigidity in determining binding affinity. Compared to GD3, the increased flexibility of Gb3 derivatives resulted in entropically less favorable interactions. These findings contribute to understanding Siglec–glycan recognition mechanisms and their implications in tumor immune evasion.

- Chapter 4: *Mechanistic insights into Siglec-7 recognition of exogenous ligands.*

  This paragraph also investigates the molecular recognition of exogenous bacterial glycoconjugates by Siglec-7, focusing on an clinically relevant pathogen: *Fusobacterium nucleatum*. In the first case, We explored the interaction between Siglec-7 and the sialylated LPS of *F. nucleatum* subsp. *polymorphum* 10953, a Gram-negative bacterium implicated in colorectal cancer. In this case, Siglec-7 was found to interact with the bacterium's lipopolysaccharide (LPS), specifically targeting its O-antigen domain, which contains neuraminic acid and 2-acetamido-4-amino-2,4,6-trideoxy-α-L-fucose (AAT). Through an integrated structural and computational approach, we identified a novel Siglec-7 binding epitope shaped by internal sialic acid and AAT residues. Structural modeling revealed a dynamic, wing-like motion of the O-antigen, allowing alternative engagement by Siglec-7's BC and CC′ loops. Our findings unveil a previously unknown immune evasion strategy employed by *F. nucleatum*

and underscore the therapeutic potential of targeting Siglec-7–LPS interactions in cancer-associated microbial pathogenesis.

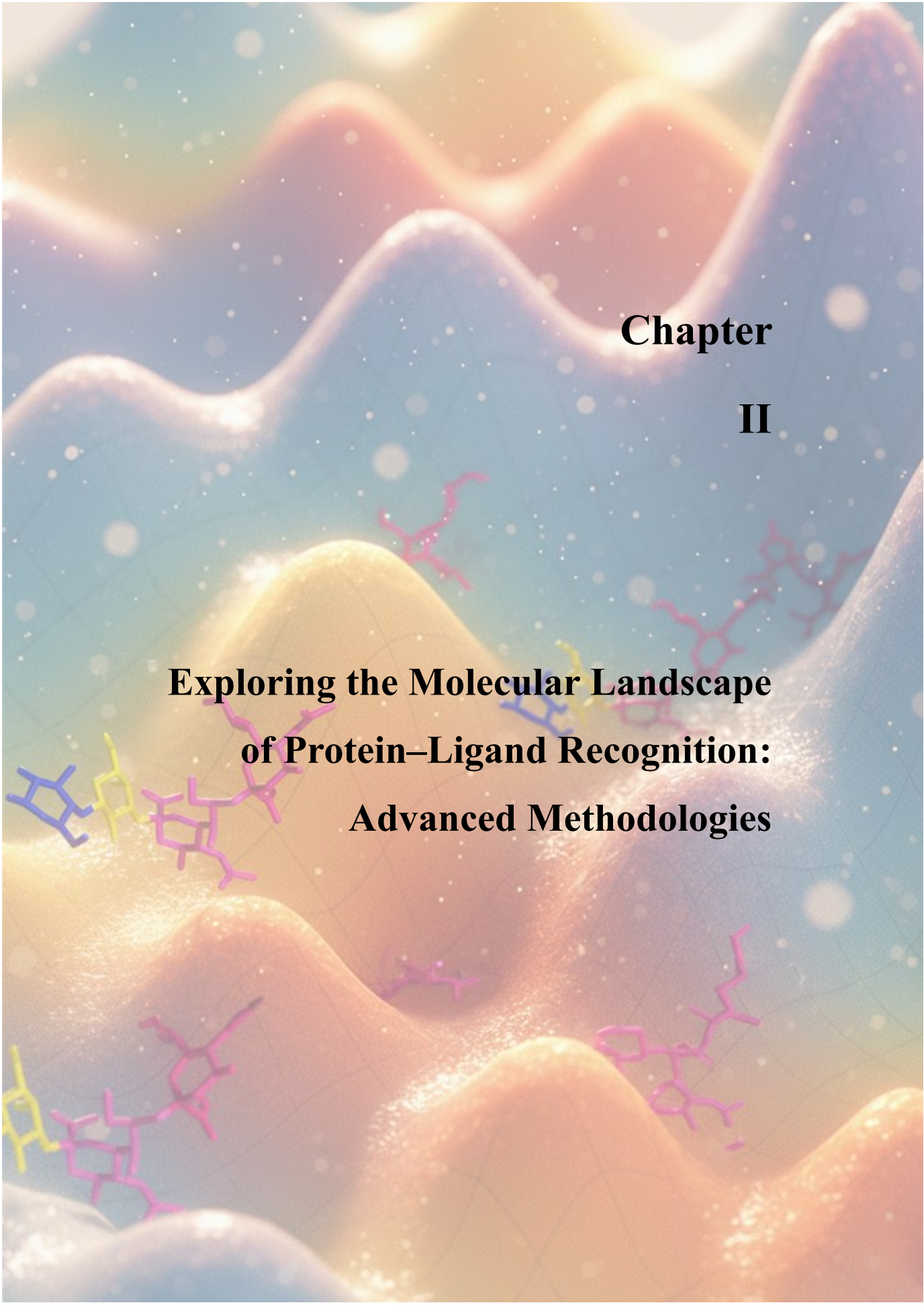
- Chapter 5: *Functional screening of Siglec-7 glycan ligands*

  To expand our understanding of Siglec-7 binding preferences beyond conformational structure and substitution pattern, a systematic screening of structurally diverse glycans is currently being conducted. This set of ligands includes tumor-associated carbohydrate antigens such as 6'Sulfo-Sialyl Lewis^{x} (6'S-sLe^{x}), ganglioside derivatives (including DSGb5 derivatives, DSGb4 and GD3 variants), Core-2 O-glycans in both sulfated and non-sulfated forms, as well as CPS/LPS/LOS extracted from different pathogenic bacterial strains. These approach combine fluorescence titration experiments and native mass spectrometry (nMS) to provide quantitative insight into Siglec-7 glycan recognition landscape and chemical-physical informations. Binding stoichiometry and binding constants are determined using label-free approaches, enabling comparative analysis of affinity profiles across chemical and structural distinct ligands. Early findings indicate that the differences in sialylation patterns, sulfation substitution, and extended glycan flexibility significantly affect Siglec-7 binding energetics. This effort contributes to a broader goal of mapping fine structural determinants of glycan–Siglec interactions, with potential applications in the rational design of synthetic ligands for immune modulation or glycan-based diagnostics.

- Chapter 6: *Neisseria gonorrhoeae LOS recognition by antibody 2C7.*

  As part of the broader investigation into exogenous glycans capable of modulating immune responses via recognition by host lectins, *Neisseria gonorrhoeae* emerged as a particularly relevant pathogen. Its surface lipooligosaccharide (LOS) displays

antigenic determinants that are specifically recognized by immune effectors, including monoclonal antibodies. Notably, *N. gonorrhoeae* represents a pressing public health concern due to its rapidly increasing resistance to conventional antibiotics. In light of this, antibody-based therapies have gained traction as promising alternatives to traditional treatments. This paragraph focused on dissecting the interaction between the monoclonal antibody 2C7 and the core oligosaccharide (OS) of the *N. gonorrhoeae* strain 15253 LOS, aiming to define the structural principles underlying selective antigen recognition. We employed both the OS and a synthetic fragment (Tetra-1) that mimics the antigenic epitope, exploring their potential as vaccine candidates. Western blot assays confirmed the selective recognition of the native OS by mAb 2C7, motivating further structural analyses. A combination of Nuclear Magnetic Resonance (NMR), mass spectrometry, computational modeling, and physicochemical assays was used to elucidate the interaction features. These findings not only deepen our understanding of the molecular determinants of mAb 2C7 specificity and affinity but also contribute to the rational development of next-generation immunotherapeutic strategies against multidrug-resistant *N. gonorrhoeae*.



Chapter II

Exploring the Molecular Landscape of Protein–Ligand Recognition: Advanced Methodologies

AVGVSTVS FRIDERICVS IMPERATOR ROMANORVM SEMPER

II. Exploring the Molecular Landscape of Protein–Ligand Recognition: Advanced Methodologies

The biological functions of glycoconjugates at the host–microbiota and host–pathogen interfaces are largely mediated by their capacity to engage in highly specific interactions with glycan-binding proteins. These recognition events are fundamental to both homeostatic and pathological processes, and their investigation is essential to understand mechanisms of immune modulation, microbial adhesion, and molecular mimicry. A thorough characterization of these interactions – especially at the structural and thermodynamic levels – is therefore essential for the rational development of glycan-based diagnostics and therapeutics, including structure-guided drug design. It is evident that the intrinsically low binding affinity, conformational flexibility, and structural complexity of carbohydrates give rise to considerable analytical challenges in the study of glycan-protein interactions. Therefore, that a singular technique is incapable of providing a comprehensive elucidation of these systems; consequently, a multidisciplinary approach that integrates wet-lab methods, high-resolution structural biology, biophysical assays, and computational simulations is indispensable.

Among the most informative techniques are nuclear magnetic resonance (NMR) and mass spectrometry (MS), which are widely utilised for the structural elucidation of bacterial glycans in both their free state and when bound to protein receptors. The application of these methodologies yields complementary data on glycan topology, conformation, and epitope presentation. Direct measurements of binding parameters are best obtained using isothermal titration calorimetry (ITC), Native Mass Spectrometry (nMS) and Fluorescence assay, which provide quantitative insights into the thermodynamics and kinetics of glycan–lectin interactions.

The following discussion will explore the question of whether the concept of 'the other' is applicable to the context of the study. In parallel, X-ray crystallography provides a means to resolve high-resolution structures of protein–glycan complexes, though its application is often limited by technical challenges such as crystallization or sample preparation. In order to circumvent these limitations and to explore the dynamic nature of these interactions, there has been an increasing integration of computational tools, including molecular dynamic (MD) simulations, docking and quantum chemical calculations, into glycoscience workflows.

2.1 Bacterial LOS extraction and purification

The study of bacterial glycoconjugates, including lipopolysaccharides (LPS) and lipooligosaccharides (LOS) from Gram-negative organisms, begins with their meticulous extraction from microbial biomass, ensuring the preservation of their native chemical structures of extracted components. Due to their structural differences, particularly in saccharide chain length and polarity, LPS and LOS exhibit distinct hydrophobic profiles. LOS, which lacks the extended O-antigenic polysaccharide, is comparatively more hydrophobic than LPS. Typically, LOS is isolated first using a phenol-chloroform-petroleum ether (PCP) extraction.[74] In this method, lyophilized bacterial cells are treated with a mixture of petroleum ether, chloroform, and 96% aqueous phenol in an 8:5:2 volume ratio. The LOS partitions into the phenol-rich phase, from which it is subsequently precipitated by water addition. The subsequent step involves the execution of the hot phenol-water method, which involves the treatment of the bacterial pellet with equal volumes of water and phenol at a temperature of 70°C[75]. After phase separation,

both aqueous and phenolic fractions are dialyzed extensively and lyophilized to yield crude extracts.

It is important to note that even minor structural variations—such as the number of repeat units, sugar composition, or specific substitutions along the glycan chain—can significantly influence the solubility and extraction efficiency of LPS/LOS molecules. Therefore, the choice and optimization of extraction protocols are critical for downstream structural analyses.

Following extraction, enzymatic hydrolysis is applied to remove residual nucleic acids and proteins. The samples are sequentially treated with DNase, RNase, and protease. To eliminate other contaminating polysaccharides, such as glycogen, amylase and pullulanase are added—preferably before protease treatment—to enhance purification.

The quality and composition of the extracted glycoconjugates are then assessed using SDS–PAGE, followed by silver staining. LPS typically produces a ladder-like pattern on the gel, reflecting the heterogeneity of O-antigen chain lengths. In contrast, LOS, due to its shorter oligosaccharide chain and lower molecular mass, migrates as a compact band near the bottom of the gel.

If required, additional purification steps such as ultracentrifugation or column chromatography can be employed to obtain highly pure preparations suitable for structural characterization and chemical derivatization.

2.2 Structural characterization of LOS/LPS

Following extraction and preliminary purification, lipopolysaccharides (LPS) and lipooligosaccharides (LOS) undergo a series of chemical treatments and derivatization steps aimed at revealing their detailed molecular architecture. The goal of structural

characterization is to comprehensively define the saccharidic and non-saccharidic components of these complex glycoconjugates.

Specifically, the analytical workflow is designed to determine:

- The qualitative and quantitative composition of constituent monosaccharides.
- The absolute configuration (D- or L-form) of each sugar unit,
- The size of the pyranose or furanose rings and the nature of glycosidic linkages.
- The sequential order of monosaccharides in the oligosaccharide chain.
- The anomeric configuration (α or β) of glycosidic bonds.
- The presence, identity, and position of non-carbohydrate substituents, such as phosphate groups, acyl chains, or amino acid modifications.

Due to the chemical complexity and heterogeneity of bacterial glycans, no single technique is sufficient to provide all this information. Instead, a multi-disciplinary approach is adopted, combining advanced mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy, often supplemented by gas chromatography (GC-MS) and chemical derivatization protocols.

Mass spectrometry offers high sensitivity and resolution for the detection of oligosaccharide masses and fragmentation patterns, allowing precise assignment of molecular weights, sugar composition, and sequence information. NMR spectroscopy, on the other hand, provides atomic-level detail on the three-dimensional structure of saccharides in solutions. It allows determination of anomeric configurations, ring conformations, and glycosidic linkage positions through analysis of ^{1}H, ^{13}C, and 2D correlation spectra (e.g., COSY, TOCSY, HSQC, HMBC). NMR is also indispensable for identifying subtle features such as acetylation, phosphorylation, and unusual sugar

moieties—including heptoses or Kdo (3-deoxy-D-manno-oct-2-ulosonic acid)—commonly found in the inner core of LOS molecules. In some cases, derivatization steps (e.g., methylation for linkage analysis or peracetylation for improved spectral resolution) are necessary to enhance the interpretability of mass or NMR spectra. Additionally, enzymatic or chemical hydrolysis may be employed to simplify complex polysaccharide mixtures into smaller oligosaccharide fragments that are more amenable to structural elucidation.

2.2.1 Gas Chromatography-Mass Spectrometry (GC-MS)

Gas chromatography coupled with mass spectrometry (GC-MS) represents a well-established and highly sensitive technique employed in the structural characterization of bacterial glycans. However, one of the primary requirements for GC analysis is that analytes must be volatile—a condition not naturally fulfilled by intact LPS/LOS molecules or their constituent monosaccharides. Therefore, appropriate chemical derivatization of the samples is an essential prerequisite before GC-MS analysis.

To render the sample suitable for gas-phase separation and detection, the glycan components are typically subjected to controlled hydrolysis to break down the polysaccharide backbone into individual monosaccharides. These are then converted into volatile derivatives, which can be analyzed based on their retention time and fragmentation patterns, allowing for the identification and quantification of specific sugar residues.

2.2.1.1 Acetylated alditol (AA)

The acetylated alditol method (AA) is a widely employed technique for the identification of neutral monosaccharides, including aldoses,

ketoses, and amino sugars. This derivatization strategy has been found to be particularly useful for the characterization of the sugar composition of LPS/LOS samples. The workflow starts with acid hydrolysis, typically employing trifluoroacetic acid (TFA), which efficiently cleaves glycosidic linkages and releases free monosaccharides. After hydrolysis, a reduction step is performed using sodium borohydride, which converts the reactive carbonyl groups of sugars into stable alditols. It is imperative to ensure that this step is completed with the greatest possible care to avoid the occurrence of ring-opening or degradation during subsequent steps in the process. Finally, the hydroxyl groups of the reduced sugars are acetylated using acetic anhydride in pyridine, producing volatile alditol acetates. These derivatives are then suitable for GC-MS analysis, where their retention times and fragmentation patterns allow for unambiguous identification and relative quantification of the constituent sugars.

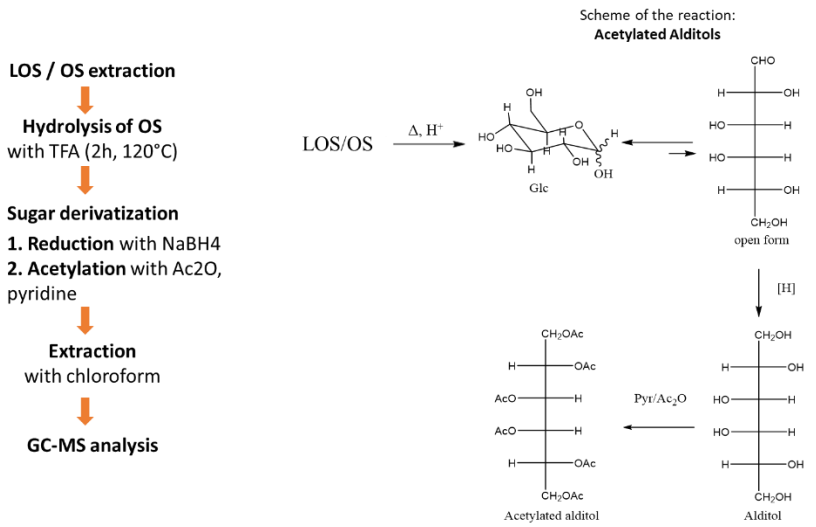


Figure II.1. Schematic representation of the chemical transformation of glycans into acetylated alditols. Native oligosaccharides are subjected to reduction and subsequent acetylation, yielding fully derivatized alditol acetates. This process stabilizes the monosaccharide units and enhances their volatility and detectability, allowing accurate compositional and structural analysis by GC–MS[76].

2.2.1.2 Acetylated Methyl Glycoside (AMG)

The acetylated methyl glycoside method (AMG) is a versatile analytical approach for the detection of a broad range of monosaccharides, including hexoses, amino sugars, uronic acids, and ulosonic acids such as Kdo. In comparison with the AA method, AMG is particularly well-suited for the analysis of acidic and acid-labile sugars, which are often underrepresented in other derivatization protocols. The procedure starts with a methanolysis reaction in anhydrous conditions, typically using a solution of hydrogen chloride in methanol (MeOH/HCl). This step serves two functions: firstly, it hydrolyses glycosidic linkages, and secondly, it converts the released monosaccharides into their O-methyl glycoside derivatives. It is imperative that this reaction is meticulously regulated, as the efficiency of methanolysis has the potential to act as a limiting factor, particularly in the context of labile glycosidic bonds or sensitive sugar residues. After the methanolysis process, the resulting methyl glycosides are subjected to acetylation using acetic anhydride in pyridine. This modification of the hydroxyl groups is intended to enhance both volatility and chromatographic performance. The final acetylated methyl glycosides are dissolved in an organic solvent, such as acetone, and the resulting sample is then ready for analysis by GC-MS. The identification of sugar is achieved through a comparison of retention times and mass fragmentation patterns with those of reference standards.

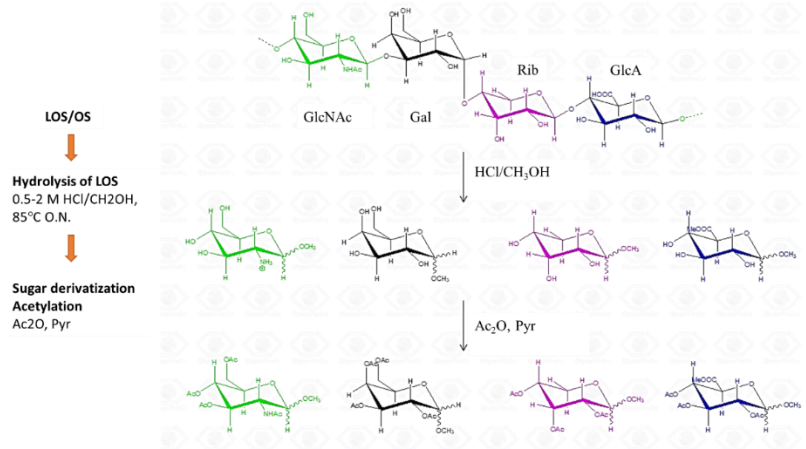


Figure II.2. Scheme of the chemical workflow used to obtain acetylated methyl glycosides from LOS/OS samples. The procedure involves controlled hydrolysis under acidic conditions to release the constituent monosaccharides, followed by methanolysis to generate the corresponding methyl glycosides. Subsequent acetylation yields fully derivatized sugar products suitable for structural analysis by GC–MS and related analytical techniques[76].

2.2.2 NMR

Nuclear Magnetic Resonance spectroscopy remains one of the most effective tools for carbohydrate structural analysis. Another significant benefit is that NMR has a capability to analyze directly within aqueous solutions usually under conditions near physiological conditions to offer useful information regarding carbohydrate structures, conformation and geometry at atomic as well as molecular levels. It is most applicable to lipooligosaccharides (LOS) and lipopolysaccharides (LPS) since they retain high solubility due to water after removal of their lipid component.

Essentially, nuclear magnetic resonance (NMR) is based upon the absorption of electromagnetic radiation corresponding to a transition among nuclear spin states subjected to a static magnetic field (B_0). Only nuclei with a non-zero spin quantum number (I), generating a magnetic moment with the ability to interact with B_0. In the field of glycans, the most informative nuclei are ^{1}H, ^{13}C, and ^{31}P, with I = ½

having the ability to orient either parallel or antiparallel to magnetic field. An exhausting analysis of NMR principles is outside the scope of this thesis; therefore, focus will be placed upon experimental methodology providing the most complete insight into oligosaccharide and polysaccharide structures within solution.

The direct NMR analysis of intact LOS or LPS is difficult because they tend to be present as micelles with limited solubility. To eliminate these problems, mild hydrolysis with acetic acid (usually 1%) is routinely carried out to break the acid-labile linkage between the Kdo residue and non-reducing GlcNAc of lipid A. It then resolves the insoluble lipid A, which precipitates, from the soluble saccharide component amenable to NMR analysis. Other de-lipidation procedures include sequential hydrazinolysis/alkaline hydrolysis to eliminate amide- and ester-linked fatty acids. Resulting oligosaccharides better withstand NMR characterization.

One-dimensional ^{1}H and ^{13}C NMR spectra provide valuable information regarding monosaccharide compositions, types of ring structure, anomeric configuration, and substitution patterns. To complete resonance assignments, however, it is necessary to conduct two-dimensional NMR experiments. Homonuclear strategies such as COSY (Correlation Spectroscopy) and TOCSY (Total Correlation Spectroscopy) utilize scalar couplings to assign spin systems of each sugar rings. Heteronuclear experiments such as HSQC (Heteronuclear Single Quantum Coherence) and HMBC (Heteronuclear Multiple Bond Correlation) allow correlations between protons with directly bonded or more remote heteronuclei to provide long-range connectivity useful for defining sequence of glycosidic linkages. Coupling constants are useful to define structural properties. For instance, $^{3}J_{H1,H2}$ coupling constants higher than 8 Hz are associated with β-anomeric configurations found within pyranose rings among both gluco- and galacto- forms, a value lower than 3 Hz indicates α-

anomers. Similarly, one-bond $^{1}J_{C1,H1}$ coupling constant is a further discriminator among anomeric configurations where coupling constants lower than 165 Hz indicate β-anomers and higher than 170 Hz indicate α-anomers.

Table II.1 Relevant 1H and 13C chemical shift value for carbohydrates analysis

δ (ppm)	**^{1}H**
8.5-7.5	Amide resonances
5.5-4.2	Anomeric protons
4.5-2.8	Sugar ring protons
2.6-1.8	a-methylene protons of deoxy sugar
2.0-1.0	Methyl protons of 6-deoxy sugars and of the acetyl group
2.0-0.8	Methyl groups
δ (ppm)	**^{13}C**
180-160	Carbonyl carbons
105-95	Anomeric carbons
80-60	Sugar ring carbons
60-45	Nitrogen bearing carbon signals
~30	Aliphatic methylene carbons of deoxy sugars
20-17	Methyl carbons of deoxy sugars, acetyl group

Through-space correlations utilize Nuclear Overhauser Effect Spectroscopy (NOESY) and Rotating-frame Overhauser Effect Spectroscopy (ROESY). These techniques allow protons aligned in close spatial proximity, usually within about 4-5 Å to be identified, thus providing information about ring conformation and glycosidic connectivity. NOE-based correlations are especially useful for verifying anomeric center configuration. For example, β-configured gluco and galacto units typically present strong intra-residue NOEs between H3/H5 and H1, whereas α-anomers present such correlations between H1/H2. Inter-residue NOEs involving anomeric proton and protons from an adjacent residue are critical to determining sequence and linkage positions. In addition, ^{1}H, ^{31}P-HSQC experiments are a

standard part of finding phosphorylation locations within oligosaccharide backbones.

Combination of different NMR strategies—1D and 2D experiments, homonuclear- and heteronuclear-based analyses, as well as scalar- and dipolar-based correlations—enables systematic rebuilding of the overall structural "puzzle" related to complex oligo- and polysaccharides. Such holistic approach emphasizes a central position of NMR spectroscopy among analytical techniques within glycobiology because such a methodology provides not only frozen static structural information but also dynamic information about conformational equilibria and solution molecular flexibility.

2.3 Exploration of protein-ligand interaction

Amongst several such methodologies employed for molecular recognition studies, ligand-based Nuclear Magnetic Resonance (NMR), native mass spectrometry (nMS), isothermal titration calorimetry (ITC) and fluorescence spectroscopy deserve special consideration. It is evident that these techniques have gained much prominence in the field of carbohydrate-protein interaction studies. This is primarily due to their capacity to achieve a balance between high levels of analytical sensitivity and reduced consumption of samples. Moreover, they facilitate access to precise, reproducible information about both structural and thermodynamic features of binding.

Ligand-based NMR has the advantage that carbohydrates in solution can be directly investigated to allow atomic-level information to aid identification of protons and structural patterns that have critical roles to play in the process of recognition[77]. In such a scenario, native mass spectrometry is a complementary method by allowing observation of intact non-covalent complexes directly with retention of their

structural integrity under such conditions to enable accurate evaluations of their stoichiometry and calculation of their equilibrium binding constants. Fluorescence spectroscopy also has a rapid and highly sensitive measurement to assess binding interactions in solution where changes to intrinsic emission from aromatic residues such as tryptophan can be transcribed into quantitative estimations of affinity. They collectively constitute a consistent experimental framework. Each brings a unique but supplementary standpoint: NMR defines fine structural details of the ligand, nMS detects overall binding equilibria, and fluorescence provides a sensitive interaction strength measure. Their simultaneous application therefore guarantees that carbohydrate–protein recognition complexity can be treated with both accuracy and efficiency.

2.3.1 Ligand-based NMR

NMR spectroscopy provides powerful methodologies for the detection and characterization of binding processes between small ligands (L) and their macromolecular receptors (P). In its simplest description, the interaction follows a bimolecular equilibrium with second order kinetics:

$$[P] + [L] \underset{k_{off}}{\overset{k_{on}}{\leftrightarrows}} [PL]$$

[Equation 2.1]

where the association (k_{on}) and dissociation (k_{off}) rate constants define the dissociation constant ($K_D = k_{off} / k_{on}$). The magnitude of K_D, described below, has direct implications for NMR experiments:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{on}}{k_{off}}$$

[Equation 2.2]

Bound state systems that exhibit weak affinity exhibit higher K_D values, requiring larger excesses of ligand to achieve saturation, whereas strong binders with low K_D values can significantly alter exchange dynamics. When the exchange between free and bound states is slow on the NMR timescale, distinct resonances for the two species are observed. Conversely, under fast exchange conditions, the signals collapse into a single peak of resonance, that represents the average of free and bound chemical shift forms. Both regimes can be exploited to monitor protein–ligand interactions, either by following ligand resonances or by detecting perturbations in protein signals.

Within this framework, ligand-based approaches have become particularly valuable for carbohydrate–protein systems, where weak and transient interactions are common.

2.3.1.1 Chemical shift perturbation (CSP)

Chemical shift perturbation (CSP), also referred to as chemical shift mapping (CSM), is one of the most widely applied NMR-based approaches to investigate protein–ligand interactions. The method is typically performed on uniformly ^{15}N-labeled proteins, for which two-dimensional ^{1}H–^{15}N HSQC spectra provide a fingerprint of the amide environment of the entire protein backbone. Upon titration with a ligand, residues that participate in or are influenced by binding undergo detectable changes in their chemical shifts, which can be systematically monitored and quantified. The experimental workflow involves the acquisition of an HSQC spectrum of the apo protein as a reference, followed by a series of spectra recorded after incremental additions of the ligand, ideally until saturation of the binding site is achieved. A critical prerequisite for reliable CSP analysis is that both the protein and the ligand are prepared in identical buffer conditions, as chemical shifts are highly sensitive to variations in pH, ionic

strength, and temperature. Even small differences in these parameters can significantly affect the position of amide resonances, leading to potential artifacts in the analysis. The behavior of the chemical shifts depends strongly on the exchange regime between free and bound states. In the fast exchange regime, the observed resonance is the population-weighted average of the free and bound states, resulting in a progressive linear shift of the peak position as ligand concentration increases. In the slow exchange regime, instead, separate signals corresponding to free and bound states may be detected, with a concomitant decrease in intensity of the free resonance as the bound population increases; in some cases, new peaks appear only under high ligand excess. More complex patterns can arise when a ligand interacts

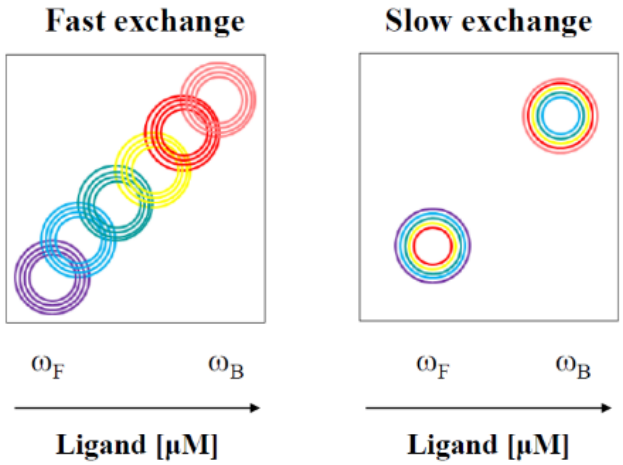


Figure II.3 Chemical shift perturbation detected by protein NMR. Progressive addition of ligand induces variations in the ^{15}N–HSQC spectrum, with the nature of the shift depending on the exchange regime between free and bound states.

with multiple protein sites of differing affinities, leading to non-linear trajectories in the chemical shift changes.

Mapping the residues that exhibit significant perturbations allows the identification of the regions of the protein most directly involved in ligand binding. However, it is important to note that not all changes

reflect direct contacts with the ligand. Perturbations may also report on long-range conformational rearrangements or allosteric effects, in which residues distant from the binding pocket experience environmental changes upon ligand association. Such observations are particularly valuable, as they provide insights not only into the binding interface but also into the conformational dynamics and allosteric communication within the protein structure.

2.3.1.2 Saturation Transfer Difference (STD) NMR

Saturation transfer difference (STD) NMR is one of the most widely used methods for mapping the epitopes of ligands that interact in solution with large macromolecules[78]. The principle relies on selective saturation of protein resonances, which is subsequently transferred to bound ligands via intermolecular NOE and spin diffusion. By subtracting two mono-dimensional NMR spectra:

- Off-resonance spectra: irradiation is applied at a frequency where neither protein nor ligand protons resonate, typically far from the spectral region of interest (around 40 ppm).
- On-resonance spectrum: irradiation is applied at frequencies corresponding to protein resonances (e.g., aliphatic region 0 to –1 ppm or aromatic region around 7.5 ppm), resulting in selective saturation of the protein magnetization. The initial magnetization transferred to the protein is transmitted to the bound ligand by intermolecular $^{1}H–^{1}H$ cross-relaxation. After dissociating from the complex, the ligand returns to the bulk solution maintaining the saturated state. Since the longitudinal relaxation rate (R_1) of ligand protons is not so fast as compared to the dissociation rate (k_{off}), the saturation that has been transferred persists after the release. As a result of repeated exchange events, saturated free ligand molecules build up inside the solution, manifested by the

reduction of the intensity of their NMR signals due to the acquired magnetization.

The intensity of each proton signal is quantified through the STD intensities:

$$\eta_{STD} = \frac{I_0 - I_{sat}}{I_0}$$

[Equation 2.3]

where I_0 is the signal intensity in the off-resonance spectrum and I_{sat} is the corresponding intensity in the on-resonance spectrum. The value of η_{STD} ranges from 0 (no transfer, non-interacting proton) to 1 (complete saturation transfer).

The STD spectrum is obtained by subtracting these two spectra. Ligand protons that receive magnetization from the protein appear as positive signals, with the intensity reflecting their proximity to the binding site.

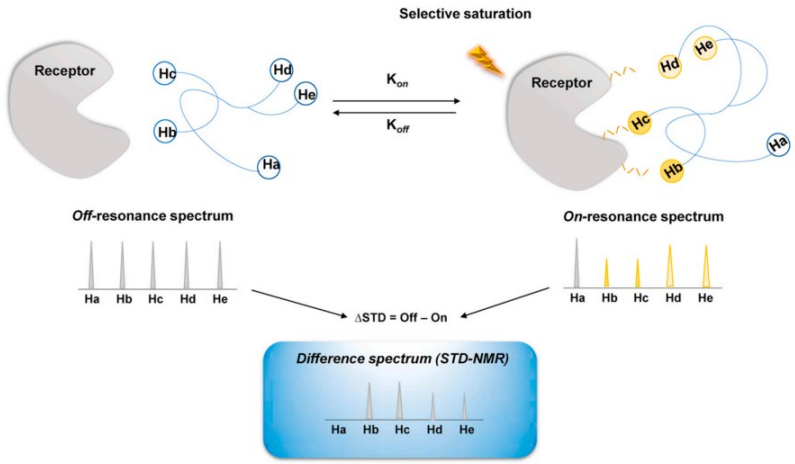


Figure II.4 Saturation Transfer Difference (STD) NMR principle. Selective saturation of protein resonances leads to magnetization transfer to bound ligand protons (Hb, Hc, He), while unbound protons remain unaffected. Comparison of the on-resonance and off-resonance spectra, followed by subtraction, yields the STD difference spectrum, in which the relative intensities of ligand signals reflect their proximity to the receptor binding site, enabling epitope mapping[78].

Depending on the degree of saturation transferred to each proton, it is possible to construct an epitope map of the ligand. The most intense STD signal is conventionally set to 100% STD effect, and the remaining signals are normalized accordingly to yield relative STD% values[79]. To minimize possible artifacts arising from differences in relaxation times, STD spectra are typically acquired at multiple saturation times (t_{sat}, usually ranging from 0.5 to 5 s), generating STD build-up curves. The initial growth rates are derived by extrapolating STD intensities towards the limit of zero saturation time, where ligand re-binding and relaxation processes are negligible. To quantify the effect, the STD amplification factor (A_{STD}) is introduced, defined as[80]:

$$A_{STD} = \frac{I_0 - I_{sat}}{I_0} \times \frac{[L]_0}{[P]_0} = \eta_{STD}\varepsilon$$

[Equation 2.4]

Where $\eta_{STD} = I_0 - I_{sat}/I_0$ is the relative STD effect, and $\varepsilon = [L]_0/[P]_0$ accounts for the excess of ligand over protein. A_{STD} is calculated for each proton involved in the interaction at each saturation time. The data are then fitted to a mono-exponential function:

$$STD\ (t_{sat}) = STD_{sat} \times (1 - e^{(-k_{sat}t_{sat})})$$

[Equation 2.5]

The relative intensities of these signals provide a semi-quantitative measure of proximity, enabling the identification of binding epitopes. The requirement for an excess of ligand (typically in the 1:20–1:200 protein-to-ligand ratio range) ensures that the bound fraction is sufficient to produce detectable effects, while the K_D range accessible to STD generally falls between 10^{-6} and 10^{-3} M. Epitope mapping can be refined by acquiring spectra at different saturation times, generating build-up curves that distinguish true proximity effects from relaxation

artefacts. Such experiments yield amplification factors that, once normalized, define the epitope map of the ligand with high resolution.

2.3.1.3 Transferred NOESY (TrNOESY)

Nuclear Overhauser Effect (NOE) experiments provide cornerstones in NMR for establishing the three-dimensional structure of molecules in solution. The transferred NOE effect exploits the distinct relaxation behaviors of a ligand in its free and bound state, providing valuable information on the conformational rearrangements that occur upon binding. From an NMR perspective, the NOE arises as a consequence of homonuclear dipolar interactions between protons in spatial proximity to each other, typically less than 5–6 Å, and NOE strength strongly depends on molecular size and solution tumbling, given by the product $\omega_0\tau c$ where ω_0 is the Larmor precession frequency and τc is the correlation time. Rapidly isotropically tumbling small molecules, i.e., molecules of molecular weights below ~2 kDa, have short correlation times and, usually positive NOEs resulting. These can, in certain combinations of molecular weight, field strength, and experimental condition, vanish and become, to an extent, negative in sign. Larger macromolecules, in this instance, proteins, have slower tumbling and long correlation times and, consequently, negative NOEs.[78]

When a small ligand binds to a receptor protein, it transiently acquires the relaxation properties of the macromolecule. If the exchange between free and bound states is fast on the NMR timescale, the NOE pattern of the bound form is effectively transferred to the free ligand resonances. This phenomenon, known as the transferred NOE (tr-NOE), allows the ligand to be observed spectroscopically as if it were part of the protein–ligand complex, thereby carrying structural information about the bioactive conformation.

An effective way to discriminate between NOEs arising from free and bound ligands is through the analysis of build-up rates, defined as the time required for an NOE to reach its maximum intensity. Since NOE enhancement depends on molecular tumbling, transferred NOEs reach their maximum at much shorter mixing times (τmix in the range of 50–100 ms) than unbound ligands, which typically require four- to ten-fold longer times. This feature provides a practical diagnostic for distinguishing signals arising from the bound state.

From a structural perspective, NOE correlations also offer valuable information about carbohydrate configuration and linkage. For instance, in β-configured saccharides such as glucose, galactose, or mannose, the anomeric H-1 proton shows strong intra-residue correlations with H-3 and H-5, whereas in α-anomers H-1 primarily correlates with H-2. Furthermore, NOE spectra can reveal inter-residue contacts, such as correlations between an anomeric proton and the protons of an adjacent sugar unit, which are crucial for determining glycosidic connectivity.

The quantitative treatment of tr-NOE data relies on the construction of build-up curves, obtained by integrating cross-peak intensities at different mixing times. The intensity of a cross-peak at a given mixing time (t) can be described by a biexponential function:

$$f(t) = a \times (e^{-ct})(1 - e^{-bt})$$

[Equation 2.6]

where *f(t)* is the integrated cross-peak intensity, and a, b, and c are adjustable parameters. The initial slope of the build-up curve, determined as the derivative at t=0, is:

$$f(0) = a \times b$$

[Equation 2.7]

This slope is proportional to the cross-relaxation rate between two protons. Once the cross-relaxation rates are determined, inter-proton distances can be calculated using the isolated spin pair approximation:

$$r_{ij} = r_{ref} \times (\frac{\sigma_{ref}}{\sigma_{ij}})^{1/6}$$

[Equation 2.8]

Where r_{ij} is the distance between protons i and j, r_{ref} is a known reference distance, σ_{ref} is the cross-relaxation rate of the reference, and σ_{ij} is the experimentally determined cross-relaxation rate.

By combining these analyses, tr-NOE experiments provide precise information on the conformation of a ligand in its bound state. This aspect is particularly relevant in the study of complex glycans, where function is tightly linked not only to chemical structure but also to conformational flexibility. A major challenge in the characterization of bacterial oligo- and polysaccharides lies in the fact that multiple conformations can coexist, with only a subset recognized by the receptor as the bioactive form. For instance, hexopyranose rings often interconvert between the two lowest-energy chair conformations, $^{1}C_{4}$ and $^{4}C_{1}$, while uronic acids may populate additional low-energy states such as the $^{2}S_{0}$ skew-boat. In monosaccharides carrying an exocyclic hydroxymethyl group, the orientation of this substituent introduces further conformational variability, generally described by three staggered rotamers—gauche–gauche (gg, –60°), trans–gauche (tg, 180°), and gauche–trans (gt, +60°). When considering oligo- and polysaccharides, additional degrees of freedom are introduced by the torsion angles around the glycosidic bonds, φ (H1–C1–O–Cx′) and ψ (C1–O–Cx′–Hx′). In the case of 1→6 linkages, the ω torsion angle (O′6–C′6–C5′–O′5) further contributes to defining the three-dimensional arrangement of the glycan chain.

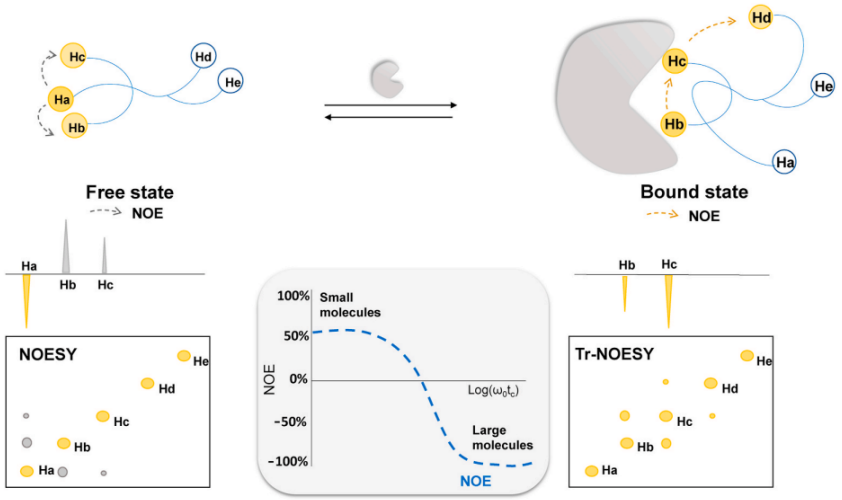


Figure II.5 Schematic representation of the NOE effect. Free ligands typically display positive NOEs, with cross-peaks of opposite sign to the diagonal in NOESY spectra, whereas upon binding they adopt negative NOEs, behaving like the macromolecule in tr-NOESY[78].

2.3.2 Mass Spectrometry (MS)

Mass spectrometry (MS) is a highly advanced analytical technique that has become a prominent tool in the field of glycoscience, due to its exceptional sensitivity, resolution, and versatility. The mass-to-charge ratio (*m/z*) of ionized molecules is measured by MS, thus providing key information on the molecular weight, composition, and structural characteristics of a wide variety of biomolecules, including glycans and glycoconjugates. The capacity to characterize both free oligosaccharides and complex macromolecular assemblies has led to MS becoming a fundamental tool in the field of structural glycobiology, complementing spectroscopic and computational approaches.

A typical mass spectrometer consists of three essential components: an ionization source, a mass analyzer, and a detector. The ionization source is responsible for transferring the analytes into the gas phase as charged species. The most common sources used include electrospray

ionization (ESI), matrix-assisted laser desorption/ionization (MALDI), and atmospheric pressure chemical ionization (APCI). After ionization, the analytes are directed into the mass analyzer, where they are separated based on their *m/z* ratio. There are several types of analyzer available, including Time-of-Flight (ToF), quadrupole, ion trap, Orbitrap, and Fourier transform ion cyclotron resonance (FT-ICR). Each of these types of analyzer offers specific advantages in terms of resolution and mass range detection. Finally, the detector registers the ions and converts the signal into a mass spectrum, which is subsequently interpreted to derive qualitative and quantitative information[81].

2.3.2.1 Native Mass Spectrometry (nMS)

While classical MS techniques are of value for the characterization of monosaccharide composition and the fine details of glycans, following chemical derivatization and fragmentation in many instances, intact, non-covalent protein-glycan complexes require more delicate methodologies. In such instances, native mass spectrometry (nMS) has proven to be a sensitive and versatile tool for direct probing of non-covalent interactions near physiological conditions. As opposed to the conventional MS method, which may be linked to denaturing conditions, nMS preserves the native structure of biomolecules. Through this, intact complexes are observed in the gas phase without compromising their functional stability. This approach is based on soft ionization by electrospray ionization (ESI), which facilitates the translation of small proteins and their ligands from the solution to the gas phase without breaking non-covalent contacts (H-bonds, hydrophobic interactions, van der Waals forces, salt bridges). An aqueous solution containing the protein and its ligand is nebulized from a high-voltage capillary in the process, resulting in an aerosol of

charged droplets. The solvent slowly evaporates, reducing the volume of the droplets to the Rayleigh limit. Coulomb fission at this point releases multi-charged molecular ions, like the protein-ligand complex.

To ensure the conservation of the complex during the transfer to the gas phase, efficient chemical and instrumental conditions need to be utilized. Low ionic strength, neutral pH and ammonium acetate volatile buffers are among the most frequently used to prevent the use of non-volatile salts, such as NaCl or phosphates, whose potential to create artifacts and complicate the interpretation of spectra has been reported[82]. Instrumental parameters (potentials, temperatures, desolvation gases) are also made "*soft*" to avoid an optimal transfer of energy to the complex[82]. A further advantage of nMS is that it can give quantitative information on the interaction dynamics between proteins and ligands. In well-controlled conditions, relative signal intensity of signals in the mass spectrum – that of free protein and bound complex – can be directly related to concentrations of species in solution. From this, it's possible to derive the dissociation constant (K_D). This is calculated through the performance of titration experiments or single-point analysis under excess ligand.

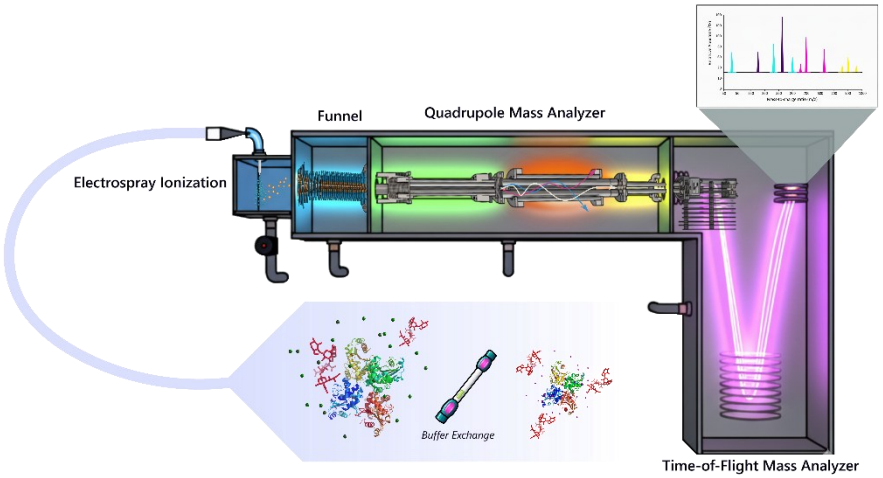


Figure II.6 Native mass spectrometry workflow. Protein–ligand complexes are introduced by electrospray ionization (ESI) under non-denaturing conditions, allowing preservation of non-covalent interactions. Following desolvation and buffer exchange, intact complexes enter the mass analyzer (quadrupole and time-of-flight), where their mass-to-charge distribution is measured. The resulting spectra directly report on stoichiometry and binding equilibria, providing quantitative information on protein–ligand interactions while maintaining near-physiological conditions.

The analytical framework applied in this work builds upon the approach first formalized by Gabelica for DNA–ligand systems[83], which we have here adapted to the study of protein–glycan interactions. This methodology provides a rigorous and widely accepted basis for extracting equilibrium constants from native mass spectra, ensuring both reproducibility and comparability with previous studies. In the context of the assessment of dissociation (K_D) or association constants (K_a) for protein–ligand complexes through the utilization of native mass spectrometry (nMS), it is imperative to acknowledge the significance of the consideration of two factors. Firstly, the peaks that characterize a mass spectrum are defined not only by their *m/z* ratio, but also by their integrated area. While the measured masses directly reveal the stoichiometry of all complexes present in solution at equilibrium – an advantage over spectrophotometric methods that infer stoichiometry indirectly from

binding curves – the relative abundances of each species can be quantitatively assessed by evaluating the peak areas in the spectrum.

In the context of ideal conditions, it is hypothesized that the relative areas of the free and bound species are proportional to their concentrations in solutions when ionized. This proportionality is expressed as follows:

$$\frac{[\mathrm{X}]}{[\mathrm{Y}]} = R \times \frac{\mathrm{A}_{(\mathrm{X})}}{\mathrm{A}_{(\mathrm{Y})}}$$

[Equation 2.9]

This relationship assumes a proportionality constant, R = 1, signifying that both species (X and Y) undergo ionization with equal efficiency. This assumption is only valid when the protein and its complex form differ minimally in terms of mass, charge distribution, shape, or hydrophobicity. These conditions are typically satisfied when small ligands bind without causing major conformational rearrangements in the target protein. As there is currently no universally applicable method to accurately determine R for each individual system, most models for K_D calculation in nMS rely on the simplifying assumption that R = 1. To account for these potential error sources, it is recommended to calculate the binding constant for each charge state observed independently. The observation of consistent results across charge states suggests that the R = 1 assumption is approximately valid. However, discrepancies may indicate bias in ionization efficiency or complex instability during ion transfer. The implementation of internal validation serves to enhance the robustness of affinity estimates derived from nMS data.

The analytical workflow proceeds as follows. For each protein–ligand mixture, peaks corresponding to species of the form [Protein + n Ligand]$^{-z}$ are annotated, with the highest *n* observed defining the maximum stoichiometry (*N*). Background subtraction is applied to remove noise while preserving peak shape, and the spectrum recorded

at the optimum voltage for each charge state is selected for analysis. Peak corresponding to each $[1{:}n]^{-z}$ species (n = 0 to N) are then integrated (correspond to $A_{(1:n)}$ value) and used to calculate the concentrations of free Protein (1:0) and of each complex according to:

$$[1{:}n] = C_{protein} \times \frac{A_{(1:n)}}{\sum_{n=0}^{N} A_{(1:n)}}$$

[Equation 2.10]

The free ligand concentration for each charge state is obtained from the mass balance equation:

$$[L] = C_L - \sum_{n=1}^{N} n \times [1{:}n]$$

[Equation 2.11]

From the values of all species in free state, stepwise equilibrium of association ($K_{a,n}$) and dissociation ($K_{D,n}$) constants are calculated for each charge state:

$$K_{a,n} = \frac{1}{K_{d,n}} = \frac{[1{:}n]}{[L][1{:}(n-1)]}$$

[Equation 2.12]

Constants derived from different charge states may not fully coincide. This divergence reflects an inherent limitation of the electrospray process, since peak intensities in the gas phase do not always correspond exactly to the relative concentrations present in solution. Outlier values were discarded only when this decision was supported by appropriate statistical analysis. The final reported constant was expressed as the means of the values obtained across the different charge states, and the associated standard deviation was used as a measure of the uncertainty intrinsic to the native mass spectrometry method.

2.3.3 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) has become an essential tool for investigating biomolecular interactions in solution, as it directly measures the heat released or absorbed during complex formation. In a typical experiment, the ligand is incrementally injected into a cell containing the protein, and the instrument records in real time the power required to maintain constant temperature. Each injection generates a calorimetric signal corresponding to the enthalpic contribution of binding, and integration of these signals produces the thermal isotherm that describes the entire association process.

The analysis of ITC data provides access to fundamental parameters that define molecular recognition. In addition to the binding constant (K_b) or its reciprocal dissociation constant (K_D), the technique yields the Gibbs free energy of binding ($\Delta_bG°$) binding enthalpy ($\Delta_bH°$), the number of binding sites (n), and, indirectly, the binding entropic contribution ($\Delta_bS°$), through the thermodynamic relationship:

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

[Equation 2.13]

Together, these parameters not only quantify the overall strength of the interaction but also allow the dissection of its enthalpic and entropic components. A central concept in ITC analysis is the Wiseman parameter ($c = K_a \cdot [M]_{cell}$), which depends on the concentration of macromolecule in the cell. Reliable experiments are obtained when c falls within an intermediate range (typically between 10 and 1000); outside this window, the binding isotherm becomes too shallow or too steep, and the resulting parameters are less accurate[84].

One of the strengths of ITC is that it provides, in a single experiment, a complete thermodynamic description of the interaction, without the need for fluorescent probes or chemical modifications. At the same time, the method is not without limitations: it requires relatively high concentrations of pure material, and the measured heat may contain

contributions from buffer ionization or dilution effects, which must be carefully controlled. Despite these challenges, the ability of ITC to directly quantify the driving forces of binding makes it a uniquely powerful approach for the study of complex biological systems.

In the context of protein–glycan recognition, ITC is particularly valuable because it disentangles the observed affinity into its energetic components. For example, the technique can distinguish between binding stabilized by direct interactions such as hydrogen bonds or hydrophobic contacts, and binding favored by the entropic gain associated with the release of water molecules. In this way, ITC not only confirms the formation of a complex but also provides insight into its energetic nature. When combined with complementary techniques such as NMR spectroscopy, fluorescence titrations, and molecular dynamics simulations, ITC contributes a fundamental and orthogonal perspective to a comprehensive understanding of glycan recognition by proteins.

2.3.4 Fluorescence Titration

Fluorescence spectroscopy is a powerful and sensitive technique for studying molecular recognition processes in solution. Its principle is based on the ability of certain chromophores to absorb light at a specific wavelength and subsequently emit light at a longer wavelength. Among amino acids, tryptophan is the main intrinsic fluorophore in proteins, and its emission properties are highly sensitive to the polarity of the local environment. Therefore, monitoring changes in tryptophan fluorescence within the binding site following the addition of a ligand provides valuable information on binding events and conformational rearrangements. In steady-state fluorescence titrations, the progressive addition of a ligand to a protein solution often results in quenching or enhancement of the emission signal as a function of concentration. Such changes may be related to

the formation of a protein-ligand complex[85]. The advantage of this approach lies in its high sensitivity, which allows the detection of subtle structural perturbations even at low ligand concentrations. Quantitatively, the interaction can be modeled as a bimolecular equilibrium, as described in Equations 2.1-2.2.

$$\frac{F}{F_0} = \frac{1 + \varphi K_b[L]}{1 + K_b[L]}$$

[Equation 2.14]

Assuming a single-site (1:1) interaction between the protein (P) and the ligand (L), the binding isotherm model relates the observed fluorescence signal to the fraction of complex formed as a function of the free concentration of [L].

$$[L] = \frac{-(1 + [P_T]K_b - [L_T]K_b) + \sqrt{(1 + [P_T]K_b - [L_T]K_b)^2 + 4K_b[L_T]}}{2K_b}$$

[Equation 2.15]

The formulation in Equation 2.14 expresses the same dependence but in terms of total ligand [L_T]. While in the first case, the unknown parameter is the amount of free ligand, in the second case, the amount of total ligand (L_T) and the amount of total protein (P_T) are known inputs chosen at the beginning of the titration, while the equilibrium binding constant K_b and the quantum yield ratio of the complex relative to the free protein are parameters that can be obtained experimentally during titration.

In practice, L_T and P_T are fixed experimental parameters, while K_b and the quantum yield ratio of the complex relative to the free protein are unknown to be determined through curve fitting of the experimental titration data. This approach enables not only the detection of binding but also the quantitative characterization of protein–ligand affinities with high precision. In the simplest 1:1 binding model, the variation in fluorescence intensity with ligand concentration can be fitted to

obtain the equilibrium constant. In addition to providing equilibrium constants at a given temperature, fluorescence titrations performed at different temperatures allow the extraction of thermodynamic parameters associated with binding through the van't Hoff formalism. The equilibrium constant K_b is linked to the standard Gibbs free energy change by the fundamental relationship:

$$\Delta_b G^\circ = -RT \ln K_b$$

[Equation 2.16]

where R is the universal gas constant and T the absolute temperature. Furthermore, by considering the temperature dependence of the equilibrium constant, the van't Hoff equation can be expressed as:

$$\ln K_b = -\frac{\Delta_b H^\circ}{RT} + \frac{\Delta_b S^\circ}{R}$$

[Equation 2.17]

In this linearized form, a plot of $ln\ K_b$ versus $1/T$ yields a straight line, whose slope corresponds to $-\Delta_b H^\circ/R$ and intercept to $\Delta_b S^\circ/R$. Thus, the enthalpic and entropic contributions to binding can be independently estimated.

This analysis provides an important extension to the interpretation of fluorescence titration data. While K_b values define the affinity of a given interaction, the van't Hoff approach disentangles the relative contributions of enthalpy and entropy, thereby revealing the energetic driving forces of the binding process. For protein–glycan systems, this can be particularly informative, as interactions are often governed not only by direct contacts such as hydrogen bonds, but also by entropic effects arising from solvent reorganization and conformational changes.

2.4 Computational Methods

Computational chemistry has progressively become an essential component of structural biology and drug design. The rapid evolution of high-performance computing resources, together with increasingly sophisticated algorithms and software packages, has allowed researchers to tackle problems that were once inaccessible. These approaches are now routinely employed not only to accelerate the identification of potential therapeutic molecules, but also to complement experimental techniques by providing atomistic and dynamic perspectives on biomolecular systems.

In the field of carbohydrate chemistry, computational methodologies have proven particularly valuable. The intrinsic conformational flexibility of glycans and their complex interactions with proteins present challenges that are not always easily resolved experimentally. Molecular modeling, however, offers the possibility to explore conformational landscapes, predict bioactive conformations, and evaluate energetic contributions that underlie molecular recognition. Importantly, when integrated with NMR or mass spectrometry data, these methods enhance the reliability of structural models by providing independent validation and mechanistic insight.

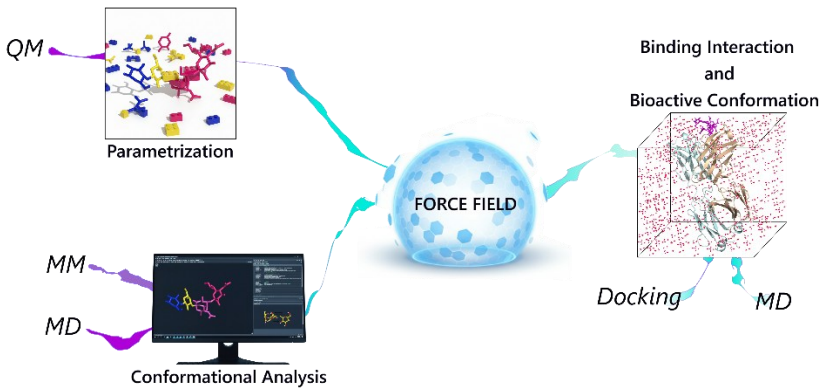


Figure II.7 Schematic representation of the computational approaches applied to investigate carbohydrates in their free and protein-bound states. Parametrization and force-field optimization enable accurate molecular dynamics simulations, which provide access to glycan conformational landscapes in solution and to binding-induced structural rearrangements in complexes. These complementary analyses yield

The following sections will outline the main computational strategies relevant to the study of protein–glycan systems. We will first introduce quantum mechanical methods and their hybrid extensions (QM/MM), which provide electronic-level descriptions and form the basis for force field development. This will be followed by molecular docking, which offers a rapid means to predict possible binding modes between glycans and their protein partners. Finally, molecular mechanics and molecular dynamics simulations are described, which allow the investigation of conformational behavior and binding processes in explicit solvent and over biologically relevant timescales.

2.4.1 Quantum mechanics and Molecular Mechanics (QM/MM)

Quantum mechanical (QM) methods represent the most rigorous level of molecular modeling, as they are based on the explicit solution of the Schrödinger equation. In principle, this equation describes the total energy of a molecular system as the sum of its kinetic and potential energy contributions:

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

[Equation 2.18]

Where $\hat{H}$ is the Hamiltonian operator of the system, Ψ the wavefunction, and E the associated energy value. The wavefunction contains complete information on the electron distribution within a molecule, from which properties such as charge density, dipole moment, and regions of nucleophilic or electrophilic reactivity can be derived. In practice, however, solving the Schrödinger equation exactly is only feasible for very small systems, and even with modern computational resources, *ab initio* approaches remain limited by high computational costs. To make calculations possible, a number of approximations have been introduced, such as the Born–Oppenheimer

approximation, which separates nuclear and electronic motion, or simplified models like the Hückel method for π-systems. Among the most applied approaches are Hartree–Fock, Post–Hartree–Fock (including Møller–Plesset perturbation theory and Coupled Cluster methods), and density functional theory (DFT), which has become widely employed due to its favorable balance between accuracy and computational efficiency. Despite their limitations in system size, QM methods have found an essential role in biomolecular studies, particularly in the parameterization of classical force fields. Energy values, geometries, vibrational frequencies, and electrostatic potential-derived charges obtained from QM calculations are commonly used to refine empirical parameters applied in molecular mechanics and molecular dynamics simulations. To bridge the gap between quantum accuracy and the ability to handle large biomolecular systems, QM/MM hybrid methods have been developed. In these approaches, the system is divided into two regions: a small, chemically relevant portion (such as an active site, a sugar moiety, or a reactive group) is treated quantum mechanically, while the remainder of the system is described using molecular mechanics. This combination allows one to capture the electronic effects governing bond making and breaking events, while still considering the structural and environmental contributions of the entire biomolecule at a reduced computational time/cost. In the context of glycobiology, QM and QM/MM approaches are particularly valuable for describing the conformational preferences of monosaccharides, evaluating glycosidic torsion energy profiles, and refining force fields tailored for carbohydrate simulations. Furthermore, they serve as a fundamental link between experimental data and computational predictions, ensuring that the models used in molecular dynamics simulations accurately represent the underlying electronic structure.

2.4.2 Molecular Docking

Molecular docking is a computational strategy widely employed to predict how small molecules, such as glycans or drug candidates, interact with biological macromolecules, typically proteins. The central principle is to evaluate the ability of a ligand to fit into a receptor binding site by exploring a large number of possible orientations and conformations, commonly referred to as poses. These poses are then assessed by scoring functions, which approximate the binding free energy and allow the ranking of the predicted complexes. Docking approaches rely heavily on the principles of molecular mechanics. Force fields, originally derived from classical mechanics, are used to estimate the binding energy between ligand and receptor by evaluating both bonded and non-bonded contributions. In this framework, van der Waals, electrostatic, torsional, stretching, and bending terms are considered to ensure that the calculation reflects the physical behavior of molecules. As a result, docking can provide not only a prediction of binding geometry but also an estimate of the relative stability and affinity of the ligand–receptor complex.

The performance of a docking calculation is determined by two fundamental components: the search algorithm and the scoring function. The search algorithm explores the conformational and orientational space of the ligand, while the scoring function ranks the poses according to their predicted binding energy. Three main types of search algorithms are commonly used:

- *Shape-matching* algorithms, which compare the geometric complementarity between ligand and protein surfaces to identify plausible binding sites.
- *Systematic searches*, typically applied in flexible-ligand docking, where all degrees of freedom are explored exhaustively.

- *Stochastic algorithms*, such as Monte Carlo (MC) simulations or evolutionary programming (EP), which introduce random perturbations to efficiently sample the conformational space.

The choice of algorithm is particularly critical when the ligands are glycans. Due to their multiple rotatable glycosidic bonds, glycans exhibit a high degree of conformational flexibility, significantly increasing the search space and computational cost. Advanced docking strategies often integrate conformational libraries or use molecular dynamics simulations as a pre-sampling step to address this challenge.

The evaluation of binding poses relies on different types of scoring functions. Force-field-based scoring directly computes interaction energies using classical terms, while knowledge-based functions derive potentials from statistical analyses of structural databases. Finally, empirical scoring functions combine multiple contributions, such as hydrogen bonding, hydrophobic interactions, and solvation effects, adjusted to reproduce experimental affinities as closely as possible.

Accurate docking calculations require a reliable three-dimensional structure of the receptor, which can be obtained from X-ray crystallography, NMR spectroscopy, homology modeling, or prediction tools such as AlphaFold. Prior knowledge of the binding site significantly improves accuracy: in AutoDock, for instance, this information is used to define the grid box that restricts ligand sampling to the relevant region. If the binding site is unknown, larger grid volumes can be employed, in a process known as *blind docking*, where the ligand is allowed to explore the entire protein surface. While this approach provides valuable exploratory information, it inevitably increases computational costs, reduces accuracy, and is therefore followed by a second round of focused docking on the most promising regions.

Among the many available docking programs, AutoDock has gained widespread use due to its efficiency and flexibility. AutoDock employs an evolutionary programming algorithm, specifically the Lamarckian Genetic Algorithm (LGA), which combines a genetic algorithm with a local search procedure. In this approach, ligand variables such as translation, rotation, and torsional angles are treated as "*genes*" defining a *genotype*, while the corresponding atomic coordinates represent the *phenotype*. The search process involves genetic operations such as mutation and crossover in genotypic space, while energy minimization refines the phenotype. The Lamarckian aspect lies in the inheritance of local improvements by subsequent generations, allowing efficient convergence toward low-energy conformations. AutoDock employs a semi-empirical free energy scoring function, which integrates intermolecular interaction terms with ligand internal energies. The pose with the lowest calculated free energy is generally considered the most favorable binding conformation.

2.4.3 Molecular Dynamics simulation (MD)

At the atomic scale, the behavior of a particle is formally described by quantum mechanics. However, explicit quantum treatment of large macromolecules rapidly becomes computationally intractable, since it requires the calculation of molecular orbitals and electron distributions for every atom in the system. For this reason, Molecular Mechanics (MM) provides an alternative, classical framework that allows efficient sampling of the conformational space of biomolecules.

In MM, molecules are represented as assemblies of atoms treated as hard spheres with defined van der Waals radii and connected by elastic bonds, angles, and torsional interactions. Electronic motions are not explicitly considered; instead, the positions of the nuclei are described

according to Newtonian mechanics under the Born–Oppenheimer approximation, assuming fixed electron distributions around the nuclei. The potential energy E_{tot} of a system in MM is expressed as the sum of bonded and non-bonded contributions:

$$E_{tot} = E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{elec}$$

[Equation 2.19]

where the bonded terms describe bond stretching, angle bending, and torsional rotations (often modeled by Hooke's law), while the non-bonded terms represent van der Waals interactions and electrostatics. These parameters are empirically derived and assembled into force fields, which provide the energy functions necessary to describe molecular geometries, conformational equilibria, and physical properties. Importantly, most force fields are developed through a combination of quantum mechanical calculations and experimental data fitting.

For carbohydrate simulations, the choice of an appropriate force field is critical. Among the most widely used are AMBER and CHARMM, both of which include parametrizations for monosaccharides and glycosidic linkages. Within AMBER, the GLYCAM_06j force field is specifically tailored for glycans, covering both D- and L-enantiomers, mono- and oligosaccharides, and a variety of glycosidic linkages, including N-glycosylation. A notable advantage of GLYCAM_06j is the consistent treatment of the anomeric carbon atom type (Cg) in both α- and β-anomers, enabling accurate simulation of ring flipping and equilibrium between axial and equatorial conformers. AMBER also provides specialized parameter sets for proteins (ff14SB), nucleic acids (OL15 for DNA, OL3 for RNA), lipids (lipid21), water models (TIP3P), and general organic molecules (GAFF2).

Building on MM, Molecular Dynamics (MD) extends the framework to simulate molecular motions as a function of time. In MD, the

positions and velocities of all atoms are propagated by numerically integrating Newton's second law of motion:

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

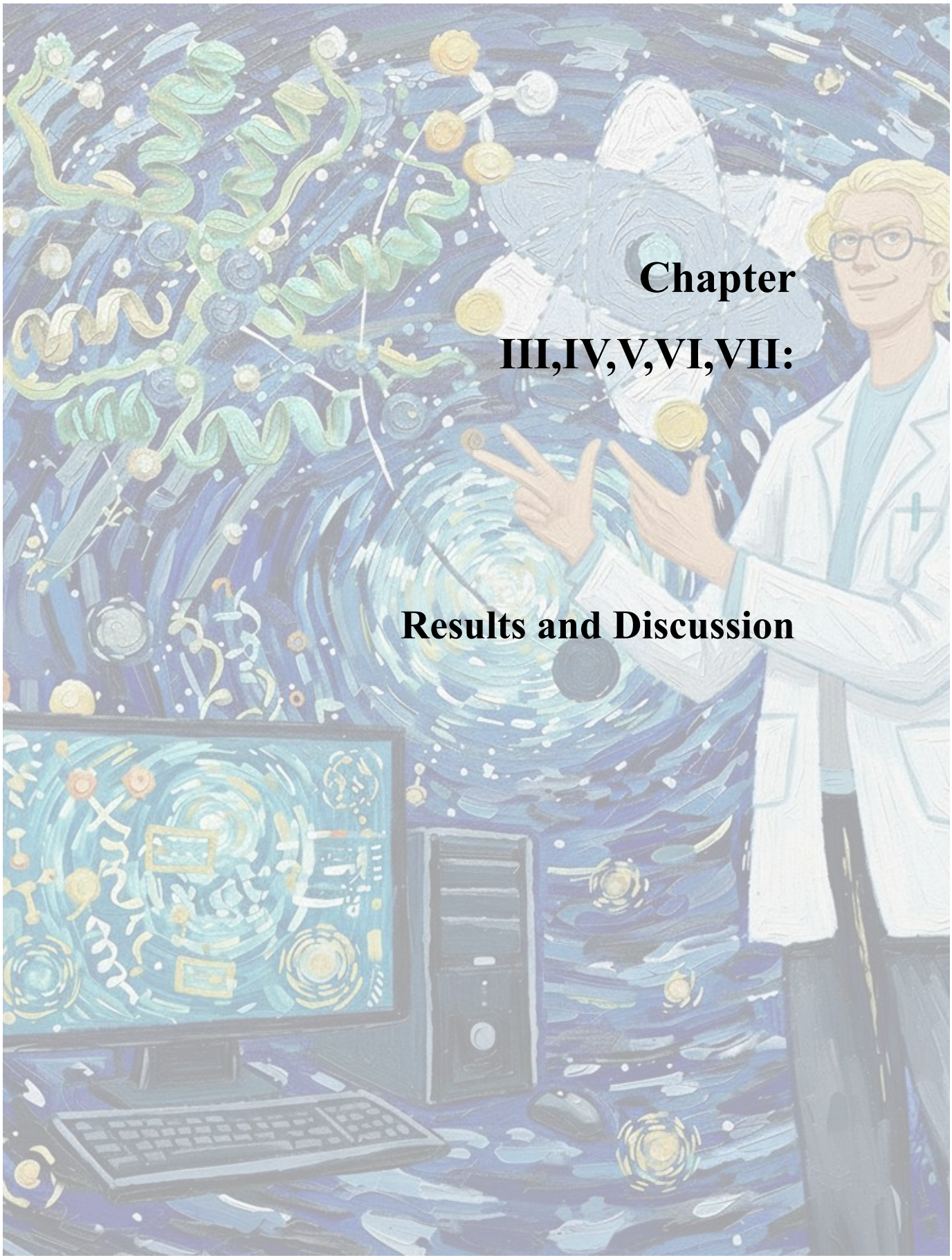
[Equation 2.20]

where m_i is the mass of atom i, x_i its Cartesian position, and F_{xi} the force acting upon it. Repeated over very small timesteps (typically 1–2 fs), this iterative process generates a trajectory, which describes the time evolution of the entire system at atomic resolution.

A typical MD simulation involves several preparation and production stages. The first is energy minimization, in which unfavorable contacts are relaxed to reach a local minimum on the potential energy surface. Common algorithms include the steepest descent and conjugate gradient methods. Next, the system is heated by gradually increasing atomic velocities according to a Maxwell–Boltzmann distribution, removing steric clashes between solute and solvent. This is followed by an equilibration phase, during which the system is allowed to relax under controlled thermodynamic conditions (temperature, pressure, energy, and volume), until stable values are reached. Finally, the production run generates the trajectory data to be analyzed.

The environment plays a crucial role in these simulations. Carbohydrates, in particular, are highly hydrophilic due to the abundance of hydroxyl groups capable of forming hydrogen bonds. Therefore, explicit solvation is typically employed, placing the solute in a box of water molecules and applying periodic boundary conditions (PBC) to mimic the bulk solvent. The most commonly used water model in biomolecular simulations is TIP3P. Long-range electrostatics are handled through Particle Mesh Ewald (PME) summation, which efficiently computes infinite-range Coulombic interactions by splitting them into short- and long-range components.

MD simulations have proven especially valuable in glycobiology. They allow the characterization of conformational landscapes of free glycans and the refinement of structural models of glycan–protein complexes. When integrated with experimental techniques such as NMR spectroscopy, MD trajectories provide a powerful framework for validating conformational ensembles and gaining insight into dynamic properties. In this way, molecular dynamics represents a bridge between static structural data and the dynamic processes that underlie molecular recognition events.



Chapter III,IV,V,VI,VII:

Results and Discussion

III. Chapter III: Investigation underlying the recognition between Siglec-7 and endogenous ligands

RESEARCH ARTICLE

ADVANCED SCIENCE Open Access

www.advancedscience.com

Insights into Siglec-7 Binding to Gangliosides: NMR Protein Assignment and the Impact of Ligand Flexibility

*Cristina Di Carluccio, Luis Padilla-Cortés, Marta Tiemblo-Martìn, Giulia Roxana Gheorghita, Rosario Oliva, Linda Cerofolini, Alessandro Antonio Masi, Celeste Abreu, Hsin-Kai Tseng, Antonio Molinaro, Pompea Del Vecchio, Ondřej Vaněk, Chun-Cheng Lin, Roberta Marchetti, Marco Fragai, and Alba Silipo**

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/advs.202415782

DOI: 10.1002/advs.202415782

3.1 Introduction

Sialic acids (Sias) are a family of negatively charged nonulosonic acids attached to the terminal portion of N-glycans, O glycans, and glycosphingolipids via α2-3, α2-6 and/or α2-8 glycosidic linkages, resulting in an immense diversity[1]. Moreover, sialoglycans recognition can be translated into a specific biological response when they are targeted by host receptors[2]. Among these, Siglecs, sialic acid-binding immunoglobulin-like receptors, are I-type lectins that share an N-terminal Ig domain that recognizes sialic acid–containing glycans[3]. They are expressed mainly by immune cells, including B-cells, NK cells, eosinophils, neutrophils, and dendritic cells and play a crucial role in distinguishing between self and non-self via glycan recognition. Gangliosides are glycosphingolipids containing one or more sialic acids, whose lipidic part is composed of ceramide. Ceramides are made up of sphingosine and a fatty acid connected by an amide bond. The ganglio series (GgCer) tetrasaccharide core is formed by Galpβ1–3GalpNAcβ1 4Galpβ1–Cer, with the length varying depending on the specific ganglioside. In addition to variations in the elongation of the core structure, gangliosides are characterized also by the number and positions of sialic acids which categorize them into 0, M, D, T, Q, and P (zero to five sialic acids) and

a, b, and c (one, two, or three sialic acids on internal Gal residue) series. Specifically, in GD3, the "ganglio" core consists of four saccharide residues, denoted by the letter G, with the letter D indicating the presence of two sialic acid residues[4]. They are endogenous Siglec ligands[5], structurally composed of an extra cellular carbohydrate moiety linked to ceramide, a hydrophobic lipid portion embedded in the membrane. Recently, the study of gangliosides has made significant advances, particularly concerning their role in pathological events. Even though they are broadly distributed, gangliosides predominate in the brain[6], and in particular, disialyl gangliosides crucially intervene in cellular processes such as neurotransmission, interaction with regulatory proteins, cell–cell recognition, and modulation of signal trans duction pathways[7,8]. In the nervous system, GD1a and GT1b are recognized by Siglec-4 (or myelin-associated glycoprotein MAG) to support axon-myelin interactions essential for long term axonal survival. Changes in gangliosides distribution and content are associated with malignancies; certain cancers also produce and shed gangliosides that have immunosuppressive effects. Disialylated gangliosides in cancers affect cell behavior, such as proliferation, migration, invasion, adhesion, and angiogenesis, as well as tumor immunosuppression[9]. Siglec-7 is an inhibitory immune receptor expressed on human natural killer (NK) cells and subsets of myeloid and dendritic cells, and rep resents a potential new target for tumor immunotherapy[10,11]. Sialoglycans expressed on cancer cells engage Siglec-7 and inhibit NK cell-mediated killing, thus making Siglec-7 and cognate ligands novel glyco-immune checkpoints for cancer immunotherapy[12,13]. Siglec-7 preferentially binds internally branched α2,6-linked disialylated gangliosides, such as DSGb5, disialosyl Lc4 (DSLc4), and α2,8-linked gangliosides such as GD2, GD3, and GT1b[14,15]. GD3 (NeuAcα2-8NeuAcα2-3Galβ1 4Glcβ1-1′Cer) is crucial for brain development but its levels are decreased in adults[16]. Conversely, GD3 is considered a crucial tumor-associated antigen, being upregulated in pathological conditions, such as cancers and neurodegenerative disorders[17,18]. Once sialoglycans on cancer cells' surface target Siglec-7, the cancer cells evade immune detection and

proliferate[19,20]. Recent findings indicate that cell surface GD3 with classical ceramide supports Siglec-7 binding, whereas GD3 with ceramide alterations, including an extra hydroxyl group on sphingosine C4 (phytoceramide) or a 2-hydroxyl group on the fatty acid amide, does not effectively engage with Siglec-7. Understanding the dynamics of these interactions holds great promise for providing insights into disease mechanisms and potentially opening the door to developing diagnostic and therapeutic strategies[21]. This can have significant implications in immunotherapy, where targeting these interactions may lead to novel therapeutic strategies against cancer, which often exploits sialoglycans to immune evasion mechanisms. To deeply understand Siglec-7 binding to closely related sialoglycan structures, and with the long-term goal of developing potential ligands within cancer immunotherapy, we undertook a comprehensive study of Siglec-7 recognition and binding to the sugar portions of GD3[22] and two Gb3 (globotriaosylceramide) derivatives[23], such as DSGb3α3 andDSGb3α6. For this purpose, we employed a combination of multidisciplinary and complementary methods, consisting of fluorescence, high-resolution ligand and protein-based Nuclear Magnetic Resonance (NMR) experiments, isothermal titration calorimetry (ITC), and computational approaches, including docking, molecular dynamics (MD) simulations, and RedMat 3D structure evaluation program[24] to obtain information about binding preferences, affinities and 3D molecular features of such protein-ligand complexes. Comprehending these interactions is instrumental in elucidating their roles in immunology, disease pathology, and potential therapeutic applications.

3.2 Results and Discussion

The molecular binding between the glycosylated Siglec-7 and gangliosides (ligands GD3, DSGb3α3, and DSGb3α6) was investigated using a combination of biophysical and computational methods.

3.2.1 NMR Assignment of the Backbone ^{1}H and ^{15}N Resonances of Siglec-7 CRD

The N-terminal Ig-like C-type carbohydrate recognition domain (CRD) of Siglec-7 has been expressed in *E. coli* in M9 culture medium containing ^{15}N NH_4Cl and ^{13}C-glucose. The protein was recovered from inclusion bodies and after a refolding protocol, a highly pure isotopically enriched Siglec-7 CRD was obtained. ^{1}H, ^{15}N, and ^{13}C backbone resonances were assigned through solution-state NMR. 2D ^{1}H-^{15}N HSQC NMR experiments showed well-dispersed resonances, which were used as confirmation of the folding state of the protein (Figure 1). 3D triple-resonance experiments HNCA, HNCACB, HNCO, and CBCAcoNH were acquired at 900 MHz, while HNcaCO on a spectrometer operating at 500 MHz. 93% of the amino acid sequence from Y26 to T147 was assigned, excluding the 5 proline residues and the histidine-tag tail. The global folding of Siglec-7 CRD was not changed by the presence of glycosylated sites[25]. The isotopically enriched Siglec-7 CRD from *E. coli* was then used in protein-based NMR experiments for binding studies with gangliosides 1–3 (see below).

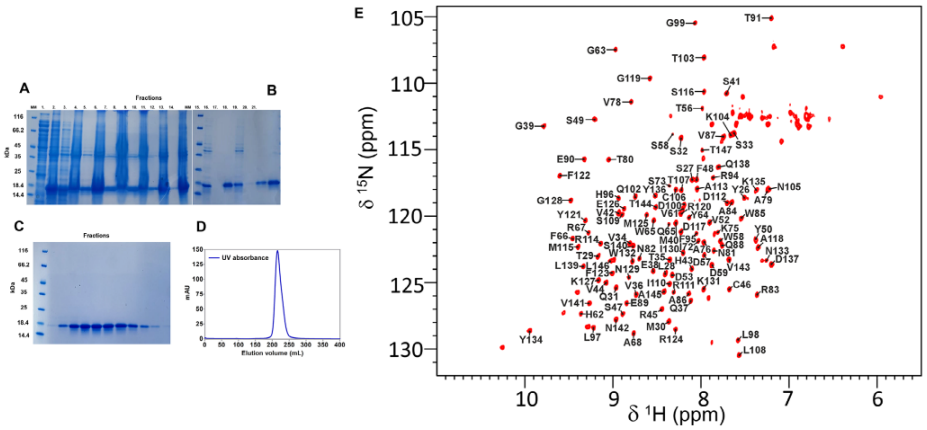


Figure III.1 Siglec-7 expression, purification, and backbone assignment by NMR. A) SDS-PAGE representation: Lysis step– (1) supernatant, (2) cell debris; Solubilization of Inclusion bodies I- (3) supernatant, (4) cell debris; Solubilization of Inclusion bodies II– (5) supernatant, (6) cell debris; Solubilization of Inclusion bodies III– (7) supernatant, (8) cell debris; Solubilization of Inclusion bodies IV– (9) supernatant, (10) cell debris; Solubilization of Inclusion bodies V– (11) supernatant, (12) cell debris; Solubilization of Inclusion bodies VI– (13) supernatant, (14) cell debris. B) SDS-PAGE representation: HisTrap purification, 1st cycle– Flow-through (15), Column wash (16), Elution (17); HisTrap purification, 2nd cycle– Flow-through (18), Column wash (19), Elution (20), 1st Elution and 2nd Elution (21). C) SDS-PAGE

representation: Size-exclusion chromatography– collected fraction of Siglec-7 CRD. D) Size-exclusion chromatogram representation of Siglec-7 CRD as single peak. E) 2D 1H-15N HSQC NMR spectrum of the apo Siglec-7 CRD in 20 mm KPi, 50 mm NaCl, pH 7.4 acquired on a spectrometer operating at 900 MHz at 298 K. NH amino acid resonances obtained by the protein assignment are reported in the spectrum.

3.2.2 Ligand-Binding Thermodynamics

The binding affinities of ligands GD3, DSGb3α3, and DSGb3α6 for the protein Siglec-7 CRD were determined by fluorescence titration experiments (Figure 2 and Table 1). Briefly, a protein solution (at fixed concentration of 4 μm) was titrated with a solution of each ligand. The fluorescence intensity of the protein was then monitored at increasing ligand concentration. We observed a concentration dependent reduction in fluorescence intensity of the protein upon ganglioside binding, i.e., the binding of sialoglycans caused a quenching of the emission of the aromatic residues of the protein. The strength of complex formation, quantitatively described by the binding constant (K_b), were determined by non-linear regression analysis of the experimental data using a 1:1 binding model equation (see Experimental Section)[26]. In Figure 2, the obtained binding isotherms are shown and the values of the binding constants at the temperature of 25 °C are collected in Table 1. An inspection of Table 1 revealed that, at 25 °C, ligand GD3 has the highest affinity for Siglec-7 CRD, with a K_b of (6.3 ± 0.1)·104 M^{-1}. Instead, for both ligands DSGb3α3 and DSGb3α6, a significantly lower K_b value was observed (Kb ≈ 103 M^{-1}), indicating a reduced affinity of both. To understand the possible reasons behind such observations, we determined the thermodynamics parameters of the complex formation by evaluating the temperature dependence of the binding constant and analyzing the data by means of the van't Hoff equation (see Experimental Section).

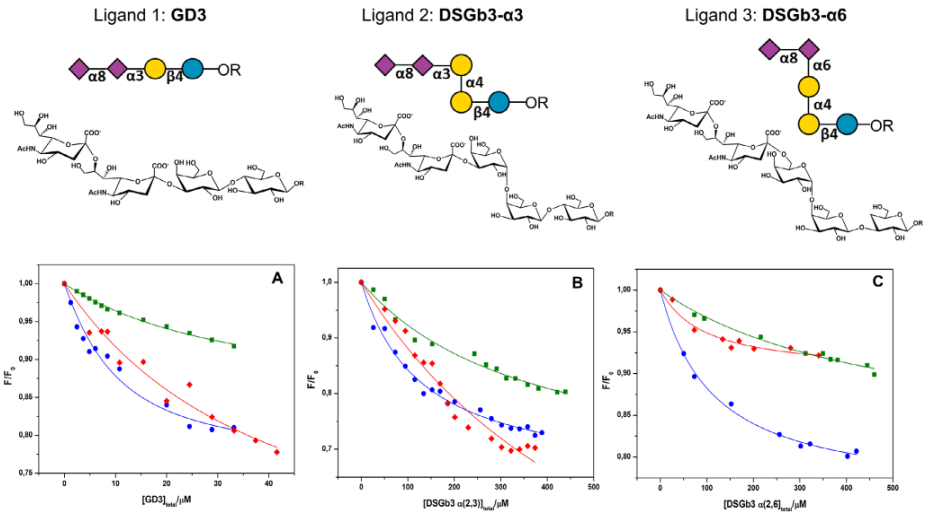


Figure III.2 Binding isotherms obtained by means of fluorescence spectroscopy for the complex formation between Siglec-7 CRD and (A) ligand 1 (GD3), (B) ligand 2 (DSGb3α3), and (C) ligand 3 (DSGb3α6) at the temperatures of 10 °C(bluecircles),25°C (green squares) and 35 °C (red diamonds). The solid lines represent the best fit of the experimental data according to a 1:1 binding model equation. All the experiments were performed in PBS buffer, pH7.4.

Table III.1 Binding constants (Kb) and corresponding equilibrium constants (K_D) for the complex formation between Siglec-7 CRD and ligands GD3, DSGb3α3 and DSGb3α6 obtained by means of fluorescence spectroscopy titrations at the temperature of 10, 25, and 35 °C.

Siglec-7 + ligand 1 (GD3)			Siglec-7+ ligand 2 (DSGb3α3)			Siglec-7 + ligand 3 (DSGb3α6)		
T [°C]	K_b [M^{-1}]	K_D [M]	T [°C]	K_b [M^{-1}]	K_D [M]	T [°C]	K_b [M^{-1}]	K_D [M]
10	$(1.9\pm0.7)\cdot10^5$	$5.2\cdot10^{-6}$	10	$(8.1\pm0.2)\cdot10^3$	$1.2\cdot10^{-4}$	10	$(1.5\pm0.6)\cdot10^4$	$6.7\cdot10^{-5}$
25	$(6.3\pm0.1)\cdot10^4$	$1.6\cdot10^{-5}$	25	$(2.7\pm0.2)\cdot10^3$	$3.7\cdot10^{-4}$	25	$(1.6\pm0.4)\cdot10^3$	$6.2\cdot10^{-4}$
35	$(4.1\pm0.1)\cdot10^4$	$2.4\cdot10^{-5}$	35	$(1.4\pm0.3)\cdot10^3$	$7.1\cdot10^{-4}$	35	$(1.3\pm0.4)\cdot10^3$	$7.7\cdot10^{-4}$

As observed from the thermodynamics parameters (Table 1), for ligand GD3, an enthalpy change of binding ($\Delta_bH°$) of −43.6 ± 8.1 kJ mol^{-1} was calculated. To test the reliability of this estimated $\Delta_bH°$ value, we performed a complementary ITC experiment by titrating a solution of the ligand with a solution of Siglec-7 CRD (see Experimental Section). Since the binding constant was quite low, the ITC experiment couldn't be performed in the usual way, reaching the saturation[27]. Instead, the experiment was carried out by injecting a small amount of protein into the calorimetric cell containing a large excess of the ligand, ensuring that all the injected protein was bound to the ligand. In these conditions, a series of similar heat peaks were obtained. The recorded heats, normalized by the moles of injected protein, provided the enthalpy change of binding ($\Delta_bH°$) whose estimated value is −48.9 ± 3.9 kJ

mol^{-1}, in excellent agreement with the one obtained by means of the van't Hoff analysis. By means of van't Hoff analysis, the $\Delta_bH°$ values were also determined for ligands DSGb3α3 and DSGb3α6. We found that the $\Delta_bH°$ are −51.1 ±0.4kJ mol^{-1} and−63.5 ±24.4kJ mol^{-1} for ligand DSGb3α3 and DSGb3α6, respectively. These results clearly indicated that $\Delta_bH°$ was very similar for the three ligands, suggesting that similar interactions were involved (see NMR section below). Of note, it was found that the entropy change of binding ($\Delta_bS°$) was negative for all the ligands, highlighting that the com plex formation was enthalpically driven but entropically disfavored. However, significant differences in the $\Delta_bS°$ values among the ligands were observed, being the lowest in the absolute value for ligand 1 compared to the $\Delta_bS°$ observed for ligand DSGb3α3 and DSGb3α6, revealing that the cause of the reduction of the binding affinity arises from the more negative entropic contribution to the binding leading to a lower Gibbs energy change of binding ($\Delta_bG°$).

3.2.3 Molecular Binding Between Siglec-7 and Ligand GD3

The interaction between Siglec-7 and GD3 was used to assess the functionality of the recombinant protein, to describe the binding mode, and dissect the 3D complex. The molecular details of the interaction were unveiled by both ligand- and protein-based NMR spectroscopy, as well as computational approaches, including MM and MD simulations (Figure 3)[28]. Saturation Transfer Difference (STD) NMR experiments were performed on the mixture to map the recognized epitope of GD3, and to determine the protons closest to the receptor surface (Figure 3A). The different profiles between the reference and the STD NMR spectra (in black and red respectively, Figure 3A) clearly indicated a selective binding between Siglec-7 and ligand GD3. Most STD NMR responses derived from the two sialic acid residues, the main belonging to the terminal sialic acid N, indicating its primary involvement in the interaction with Siglec-7. The highest STD signal originated from the acetyl group of Sia N (AcN) and was set to 100%; the other STD effects were derived accordingly. In turn, residue K

contributed mainly with H7, H4, and its acetyl group, and at lower extent with H5-H6. The interaction of Siglec-7 with GD3 was further investigated by protein-based NMR experiments (Figure 3B). Aliquots of GD3 were sequentially added to [U-^{15}N] labeled Siglec-7 CRD and ^{1}H-^{15}N-HSQC NMR spectra were acquired. As expected, during the NMR titration, some protein signals experienced a decrease in signal intensity and/or chemical shift perturbations (CSP) (Figures 3B and 4A,B). These changes were significant for cross-peaks assigned to residues located near Arg124, the canonical residue in the Siglec-7 binding site. In particular, the signals affected by the largest CSP were assigned to Lys127, Lys131, and Trp132, followed by Glu126, Gly128, Ile130, as well as Tyr134 and Lys135 of GG' loop, supporting the anticipated interaction involving Arg124. Other CSP effects were observed for Tyr26, Ser27, Arg111, and for His62 of BC loop and Ile72, Ser73 of the CC' loop. On the other hand, the signals affected by the largest decreases in intensity were assigned to Asn133, which even disappeared with the addition of the ligand, and Lys127 and Lys131, the same residues which further experienced CSP. To a lesser extent, other intensity decreases were also observed for Thr56, Gly63, Phe95, and Met125.

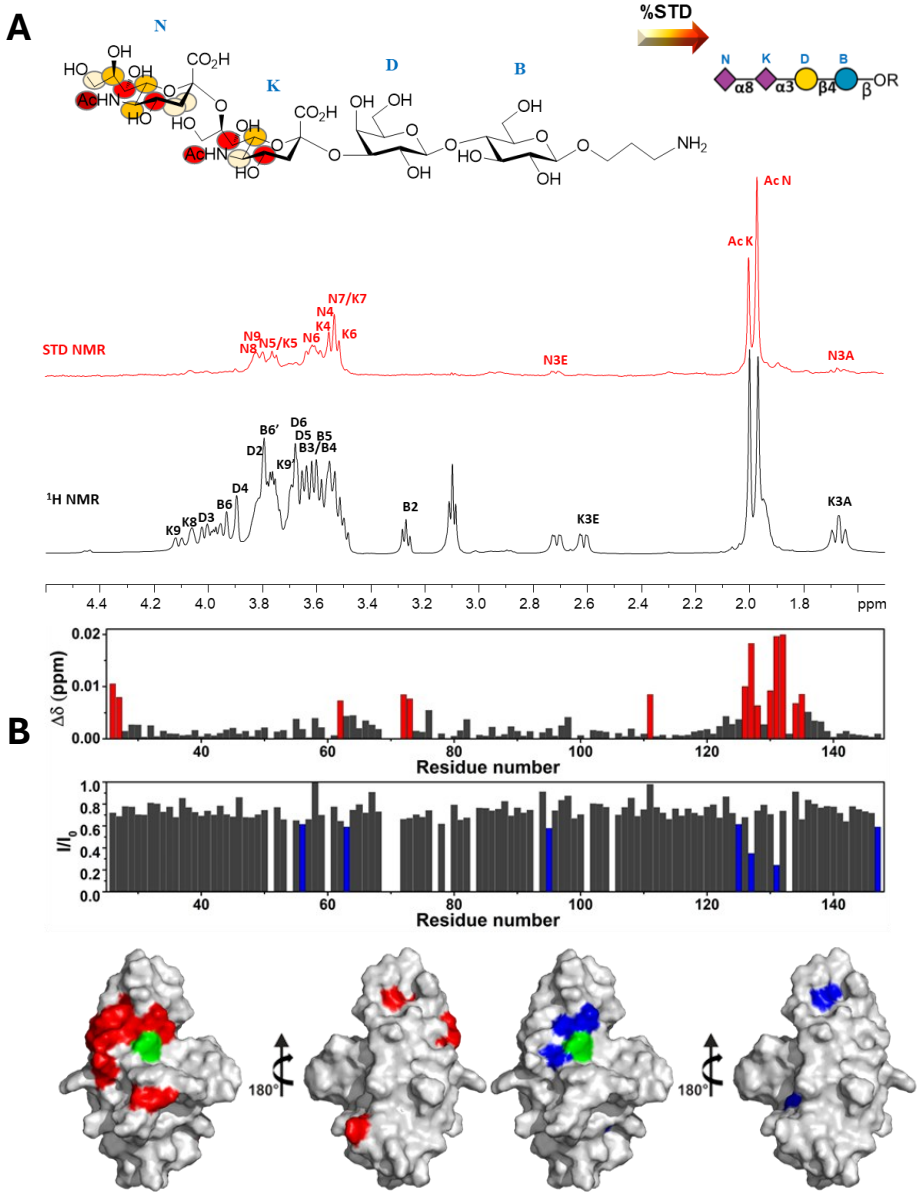


Figure III.3 Ligand- and Protein-based NMR analysis of ligand GD3 in interaction with Siglec-7. A) Epitope mapping of ligand GD3 highlights the protons most involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMR spectrum (red) and I0 is the peak intensity of the off-resonance spectrum (black). Despite some isochronous signals between N and K, the isolated signals from K, such as K8, as well as the diastereotopic K9 and K3 (axial and equatorial) protons were not recognized by the protein. However, an STD enhancement referring to the acetyl group (AcK) was observed, meaning a partial contribution of the internal sialic acid (K) occurred. Bioactive conformation of ligand GD3, the surface was colored according to the STD effects. B) Plots representing chemical shift perturbation (CSP) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensityinblue and CSP in red in the presence of 200 μm ligand GD3. Arg124 is colored green.

The findings from both NMR titration and STD experiments were essential in

constructing the binding site (Figure 4A,B). The α2-8 linkage connecting the two sialic acids of GD3 shows high levels of conformational flexibility due to the five torsion angles by which it is characterized: ϕ (C1-C2-O-C8′), ψ (C2-O C8′-C7′), ω9 (O9′-C9′-C8′-O), ω8 (O8′-C8′-C7′-O7′), and ω7 (O7′ C7′-C6′-O6′). Once the conformational behavior of GD3 in the free state, previously evaluated[29], was further studied via MD simulations, the ligand in its representative pose was modeled into Siglec-7. Depending on the φ (C1-C2-O-C'3) torsion angle around the Neu5Ac-α-(2,3) Gal glycosidic linkage[30], ligand 1 could populate two different conformations, namely –g and t, corresponding to φ values of 60° and 180°, respectively. In the bound state, a preference for φ=−60°wasobserved from MD simulations also supported by tr-NOESY experiments. Overall, the spectral shifts observed in Siglec-7 upon the addition of ligand 1 were in agreement with a protein-ligand binding affinity at a low micromolar level. The combination of ligand- and protein-based NMR experiments allowed the modeling of 1 into the binding site. The complex was subjected to MD simulation and the 3D models were evaluated and validated with RedMat[24] (Figure 4E), a program that calculates the predicted STD-NMR intensities from MD trajectories and compares them to the measured STD effects. Notably, the calculated STDs matched well with those experimentally observed, confirming the reliability of the MD simulation on Siglec-7 and GD3 (ligand 1) and thus of the proposed 3D model in Figure 4. As shown in Figures 4C and 5, the carboxylate group of the terminal sialic acid established the key salt bridge with the Arg124. Also, a significant H-bond was detected between the H5N of terminal Sia N and Lys131, one of the amino acids most perturbed by both CSP and intensity decrease (Figures 3B and 4A,B). To a lesser ex tent, H-bonds between O8 of the glycerol chain of N with Asn133 were observed, which was the amino acid mostly perturbed by a decrease in signal intensity. Regarding the internal Sia residue K, an H-bond between O4 of sialic acid and Trp74 was recurrent during the simulation. The analysis also revealed the presence of hydrophobic aminoacids in the binding pocket involving the aromatic components of Trp132 and Trp74, the latter situated within the CC'

loop of the protein, in proximity to the residues N and K, respectively (Figure 4C). In particular, in our 3D model, Trp132 established a π–π interaction with Tyr26, both amino acids affected by CSP variations during the protein NMR titration (Figures 3B and 4B). Overall, the 3D model of Siglec-7 and GD3 (ligand 1) proposed here matched NMR binding studies, indicating the terminal sialic acid of the ganglioside as the primary residue involved in the interaction with Siglec-7 and the internal sialic acid partially contributing to the recognition, while the other residues were solvent exposed.

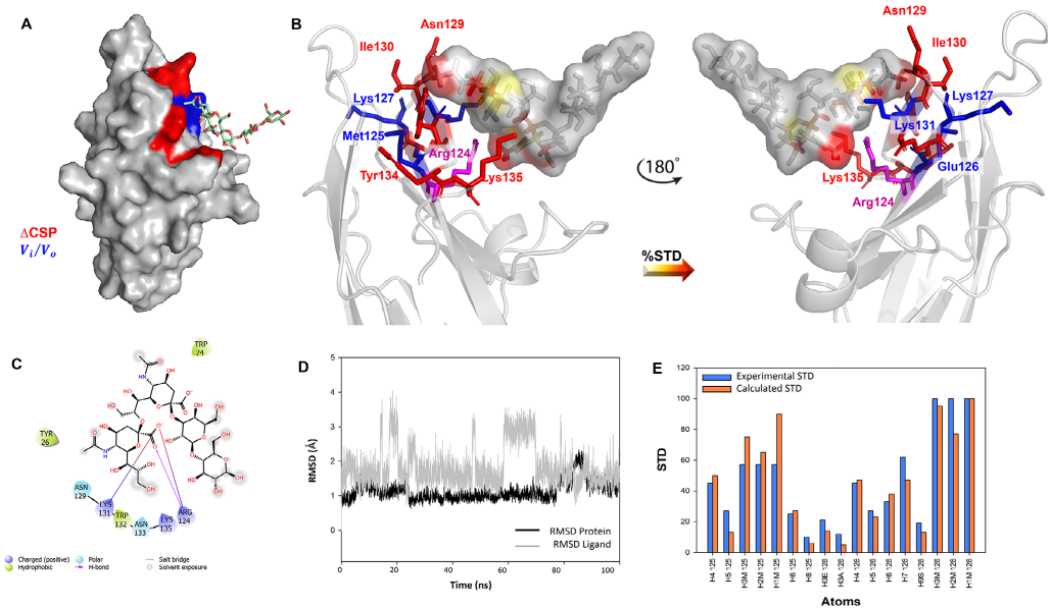


Figure III.4 3D model of Siglec-7CRD-ligand GD3 complex. A) 3D complex with the protein surface colored according to the chemical shift perturbation (in red) and intensity decreased (in blue) detected by protein-based NMR titration. B) 3D views of the Siglec-7−GD3 complex: the amino acids involved in the interactions as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code. C) 2D plot of the interactions occurring at the protein-ligand interface. D) RMSD along MD simulation. E) Comparison between experimental (blue) and calculated %STD for protons of ligand 1 by RedMat analysis. A NOE R-factor of 0.2 was calculated.

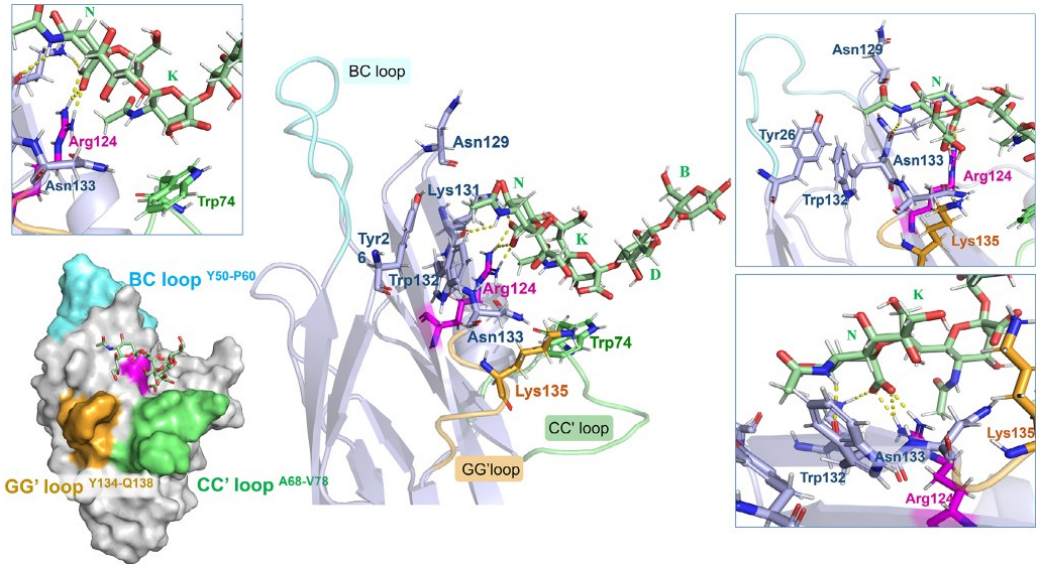


Figure III.5 Representation of the Siglec-7−GD3 complex with the BC, CC' and GG' loops highlighted in cyan, green, and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops.

3.2.4 Molecular Binding between Siglec-7 and Ligand DSGb3α3

The molecular recognition of DSGb3α3 by Siglec-7 was analogously investigated by NMR and computational studies. STD NMR experiments indicated that Siglec-7 mainly accommodates sialic acid residues, primarily the terminal Sia (N) (Figure 6A). An inspection of the STD spectra highlighted strong involvement of the acetyl group, H7 and H4 of Sia N and Sia K, that contributed to a lesser extent. Furthermore, the other sugar residues were mostly excluded from recognition by Siglec-7 (Figure6A). Siglec-7 titration with ligand DSGb3α3 was then performed to have deeper insights into the 3D complex; both CSP and decreases in signal intensity were observed and allowed us to map the region of the protein involved in the interaction with DSGb3α3, especially surrounding the key residue Arg124 (Figures 6 Band 7).

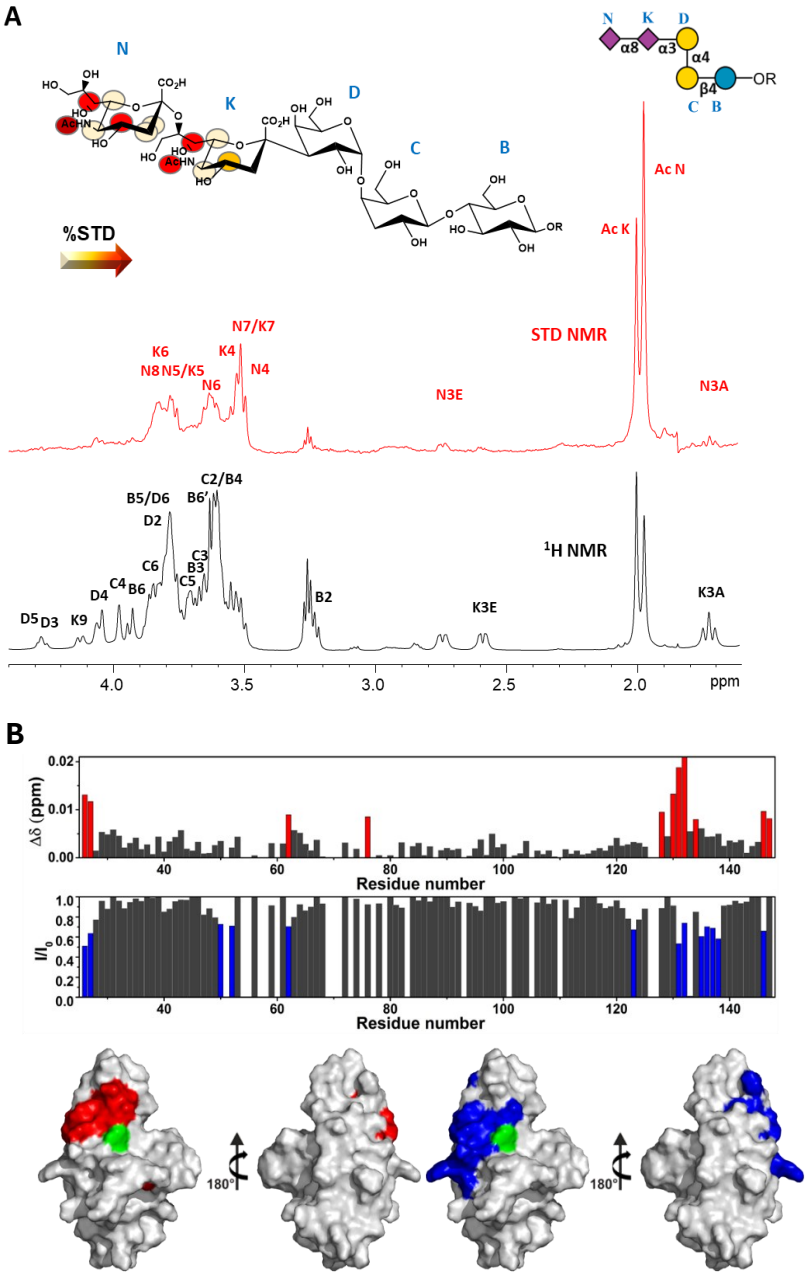


Figure III.6 Ligand- and Protein-based NMR analysis of ligand DSGb3α3 in interaction with Siglec-7. A) Epitope mapping of ligand DSGb3α3 highlights the protons more involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMRspectrum(red)andI0 isthepeakintensity of the off-resonance spectrum (black). Bioactive conformation of ligand DSGb3α3, the surface was colored according to the STD effects. B) Plots representing chemical shift perturbation (CSP) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensity in blue and CSP in red in the presence of 200 µM of ligand DSGb3α3. Arg124 is colored green.

In this region, the most significant CSPs were attributed to Gly128, Ile130,

Lys131, Trp132, and Tyr134, while Lys131, Trp132, Asn133, and all the amino acids from Lys135 to Gln138 of the GG' loop experienced a decrease in signal intensity. The residues of GG' loop (Tyr134 by CSP and Lys135, Tyr136, Asp137, and Glu138 by intensity decreases) were indeed particularly influenced by the presence of ligand DSGb3α3; we also observed a significant CSP for the signal of Ala76 of the BC loop and decreases in signal intensity for residues belonging to the BC loop (Tyr50 and Val52). Besides Lys131, and Trp132, which were affected by both kinds of changes, we also found Tyr26, Ser27, and His62, close in space to the binding site experiencing CSPs (Figure 7). The bioactive conformation adopted by ligand 2 upon binding to Siglec-7 was then explored by tr-NOESY experiments coupled to MD simulation. In the free state, ligand DSGb3α3 could populate –g and t conformations according to the φ (C1-C2-O-C'3) torsion angle around the Neu5Ac-α-(2,3)-Gal glycosidic linkage (−60° and 180° respectively). Therefore, an MD simulation in the bound state was first carried out with ligand DSGb3α3 in the t conformation, maintained for approximately 20 ns after which the φ torsion angle turned to -60° (change from t to-g), maintained until the end of the MD simulation. In addition, a second MD simulation in the bound state was undertaken considering ligand DSGb3α3 in-g conformation as the starting pose: for the entire simulation time, the φ torsion angle populated the value of −60°. Therefore, in both MD simulations of the bound states, the preferred conformation of the φ torsion angle around Neu5Ac-α-(2,3)-Gal of ligand DSGb3α3 appeared to be −60°. Confirmation came from a detailed analysis of key tr-NOE contacts, that indicated a propensity of DSGb3α3 for adopting the -g conformation, supported by a decrease in NOE intensity corresponding to $K3_{ax}$-D3, and by the appearance of a weak NOE corresponding to K9-D3. Regarding the torsion angles around Neu5Ac-α-(2,8)-Neu5Ac linkage, φ (C1-C2-O8-C8) can populate different conformational states between −60° and −180° in the free state, while φ at −60° was favored in the bound state. In addition, tr-NOE contacts between the terminal sialic acid (N) and the glucose residue (B) further suggested that ligand DSGb3α3 adopted a bent

conformation upon binding to Siglec-7, matching with a φ torsion angle around Neu5Ac-α-(2,8)-Neu5Ac at −60° (Figure 7).

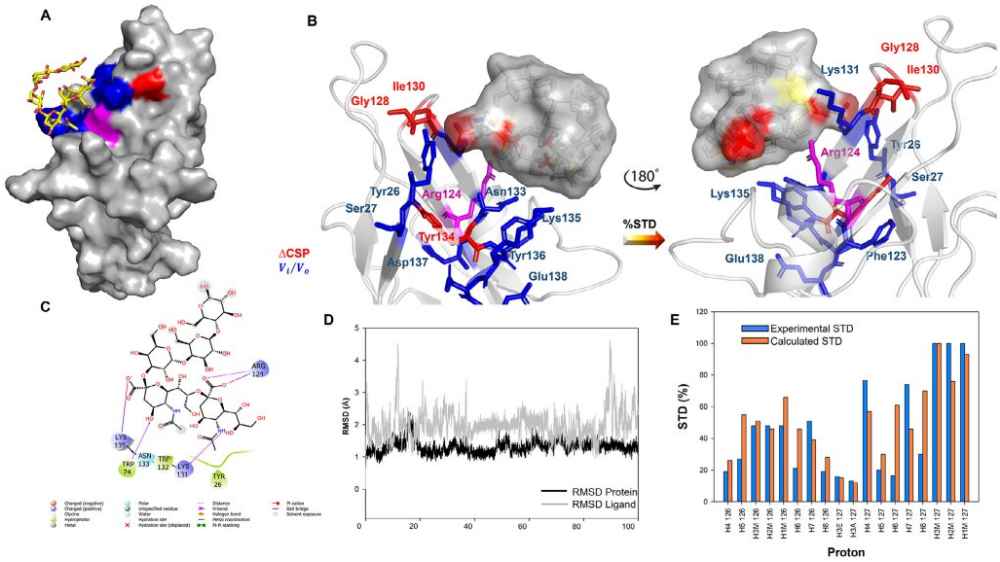


Figure III.7 3D model of Siglec-7CRD-ligand DSGb3α3 complex. A) 3D complex with the protein surface colored according to the chemical shift perturbation (in red) and intensity decreases (in blue) detected by protein-based NMR titration. B) 3D views of the Siglec-7−DSGb3α3 complex: the amino acids involved in the interactions as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code. C) 2D plot of the interactions establishing at the protein-ligand interface. D) RMSD along MD simulation. E) Comparison between experimental (blue) and calculated %STD for protons of ligand DSGb3α3 by RedMat analysis. A NOE R-factor of 0.3 was calculated.

Notably, these torsions remained stable along the MD simulation. Moreover, the MD trajectories were analyzed through RedMat algorithm (Figure 7E) and a 3D model of Siglec-7 with DSGb3α3 was proposed (Figure 8). The results illustrate the critical role of the carboxylate group of the terminal sialic acid N in establishing a salt bridge with Arg124. Additionally, hydrogen bonds were identified again between the H5N of the terminal sialic acid N and Lys131, affected by both CSP and decrease of signal intensity, as well the previous interactions with Asn133, which showed a substantial reduction in signal intensity during the protein NMR titration (Figure 6B). As for the internal Sia residue K, this residue was facing toward the aromatic Trp74 belonging to the CC' loop by which a hydrogen bond was also established (Figures 7C and 8). Other aromatic amino acids located in the binding site included Tyr26 and

Trp132, which were affected by both CSP and intensity decreases and established π-π interactions with each other.

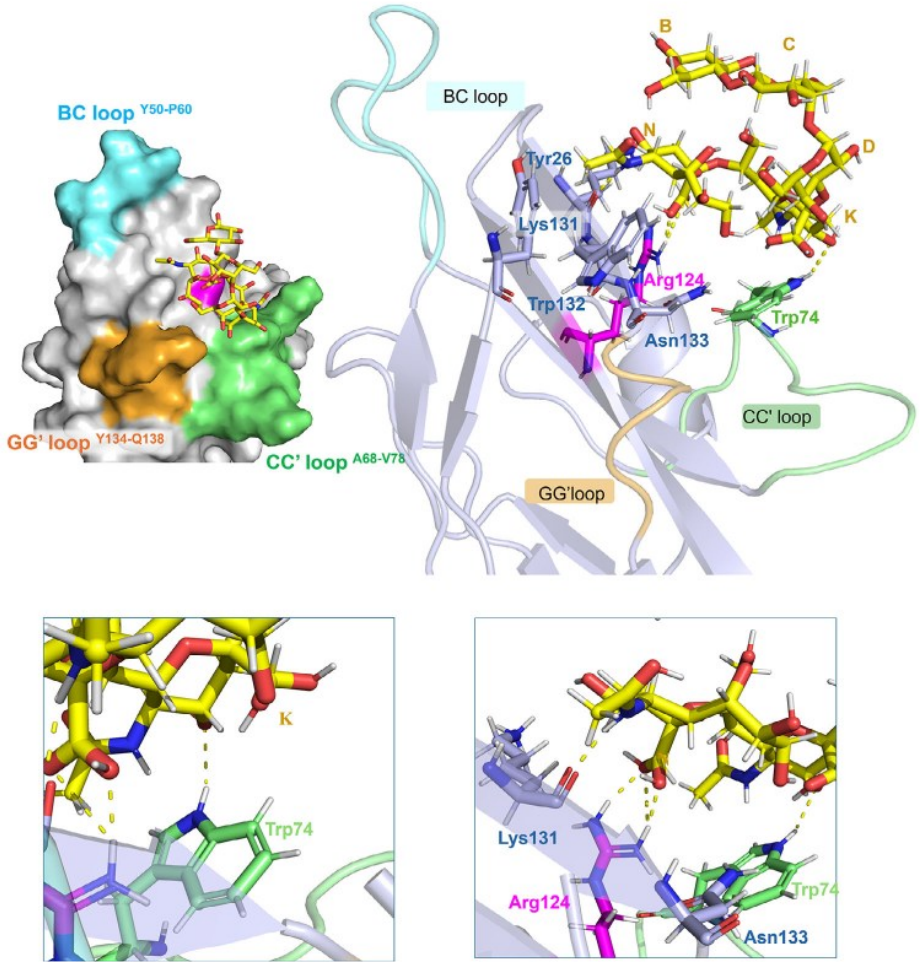


Figure III.8 Representation of the Siglec-7− DSGb3α3 complex with the BC, CC' and GG' loops highlighted in cyan, green, and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend).

3.2.5 Molecular Binding Between Siglec-7 and Ligand DSGb3α6

NMR and computational studies have also been performed to study the molecular recognition of ligand DSGb3α6 by Siglec-7. Similarly to ligand DSGb3α3, STD NMR experiments of ligand DSGb3α6 showed that the

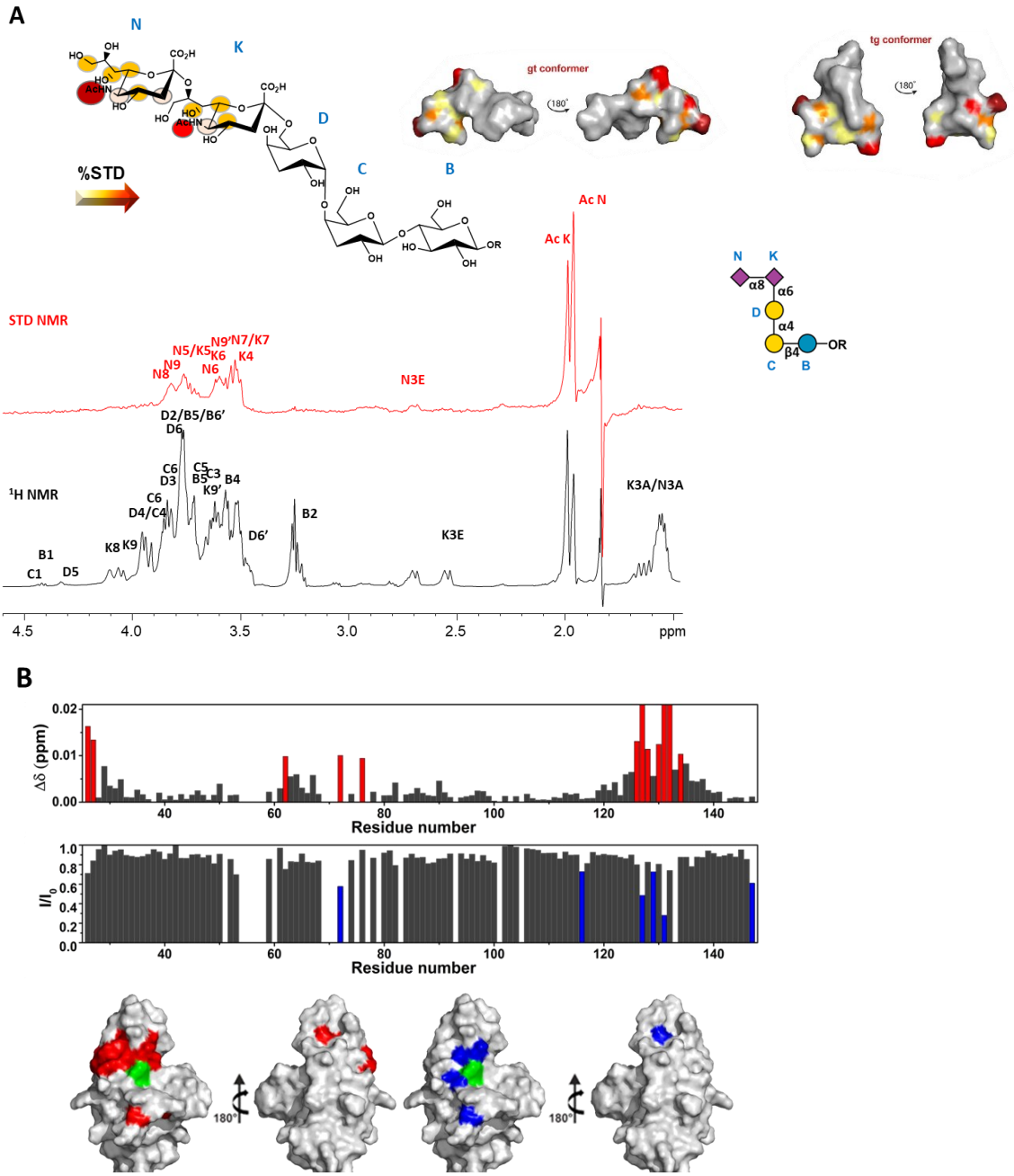


Figure III.9 Ligand- and Protein-based NMR analysis of ligand DSGb3α6 in interaction with Siglec-7. A) Epitope mapping of ligand 3 highlights the protons more involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMRspectrum(red)andI0 isthepeakintensity of the off-resonance spectrum (black). Bioactive conformation of ligand DSGb3α6, the surface was colored according to the STD effects. B) Plots representing chemical shift perturbation (CSP) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensity in blue and CSP in red in the presence of 200 µm ligand DSGb3α6. Arg124 is colored green.

terminal Sia residue (N) was crucial for the interaction (Figure 9A) and

contributed from N4 to N9, including the acetyl group that gave the 100% STD effect. Interestingly, as for the diastereotopic methylene signals of Sia, a low STD enhancement was only noticed for N3E. Regarding Sia K, medium to low STD responses were attributed to K4, K7, K6, and the acetyl group. Similarly to ligands GD3 and DSGb3α3, the other residues were excluded from interaction with Siglec-7, as indicated by the absence of corresponding STD NMR signals. The molecular binding between Siglec-7 and ligand DSGb3α6 was also investigated by protein-based NMR binding experiments (Figure 9B). Both CSP and decreases in signal intensity were observed during the titrations, which allowed the definition of the binding site of Siglec-7 (Figure 10).

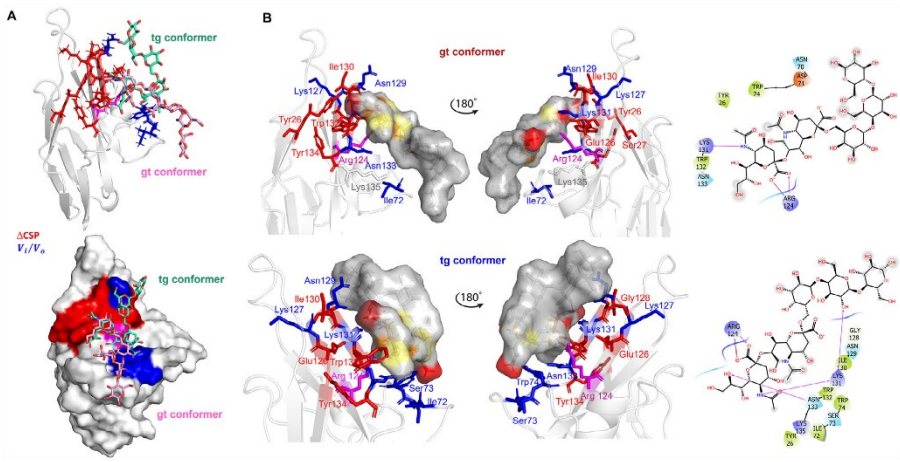


Figure III.10 A) Superimposition of 3D models of Siglec-7CRD bound to ligand DSGb3α6intg and gt. B) 3D views of the Siglec-7− DSGb3α6 complex: the amino acids involved in the interactions as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was col ored according to the STD edit code. 2D diagram of the interactions established at the protein-ligand interface, with DSGb3α6intg (A) and gt (B) conformations.

Therefore, residues from Glu126 to Tyr134, in the vicinity of Arg124, exhibited consistent perturbations; in particular, Lys127, Lys131 and Trp132 experienced the largest CSP, followed by Glu126, Gly128, Ile130, and Tyr134; the remaining amino acids were instead affected by intensity decrease (Lys127, Asn129, Lys131 and Asn133). Notably, Lys127 and Lys131 were affected by both CSP and intensity decrease. Moreover, the amino acids in proximity to this latter region included Tyr26, and Ser27, which only gave CSP effects. On the other hand, lower perturbations were evidenced for residues His62 of the BC

loop and Ala76 of the CC loop influenced by CSP and for Ile72 of the CC loop affected by both CSP and intensity decrease (Figures 9B and 10). Tr-NOESY experiments and computational studies allowed us to depict the 3D model. Thus, Neu5Acα2-6-Gal linkage was affected not only by φ (C1-C2-OC6′) andψ (C2-O-C6′-C5′), but by an additional ω torsion angle around the C5′-C6′ (O-C6′-C5′ O5′). Previous computational studies in the free state showed the ψ angle mainly acquired values around 180°[31]also maintained in the bound state. As for the φ torsion angle, the absence of NOE contacts between the diastereotopic protons at position 3 of sialic acid (K3A and K3E) and protons at position 6 of galactose (D residue) is an indication of a preference for the -*g* conformation, in turn confirmed by MD simulations. Additionally, the ω torsion angle could populate three different values in the free state, 60°/180°/−60°,cor responding to *gt/tg/gg* conformations, respectively[26,32]. Overall, both MD simulations and tr-NOESY NMR experiments revealed no conformational differences between free and bound state, and that both *tg* and *gt* conformations were likely populated (Figure 10). Indeed, ω at −60° was less energetically favored and the different multiplicity of the diastereotopic protons at position 6 of the galactose residue D observed on the ^{1}H ^{13}C-HSQC spectrum confirmed the exclusion of the *gg* conformation. As for the other two torsions, no significant NOE contacts were found for ligand DSGb3α6 to discriminate in the bound state between the two conformers. The coexistence of both *gt* and *tg* conformers was further supported by protein based NMR experiments (Figure 11), which showed that amino acids located in the CC' loop of the protein, such as Ile72 and Trp74, were affected in both cases. This also occurred with amino acids nearby or part of the GG' loop such as Lys131, Trp132, Asn133, Tyr134, or Lys135. The interaction of the key Arg124 with the terminal sialic acid K of the ligand occurred in both *gt* and *tg* conformations, hydrogen bonds could also be identified between Lys131 and the terminal sialic acid, in particular the NH of the acetyl group. Particularly, in the *tg* conformation the H-bonds between the terminal sialic acid (N) O8 and (at lesser extent) O9 of the glycerol chain and Asn133 were also observed. Within the internal Sia K residue, a hydrogen

bond with Trp74 was observed in both conformers (Figure 11). Therefore, the combination of ligand-based and protein-based NMR with MD simulations and Redmat calculations allowed us to obtain 3D models of the Siglec-7-ganglioside complexes, highlighting the interactions occurring at the interface.

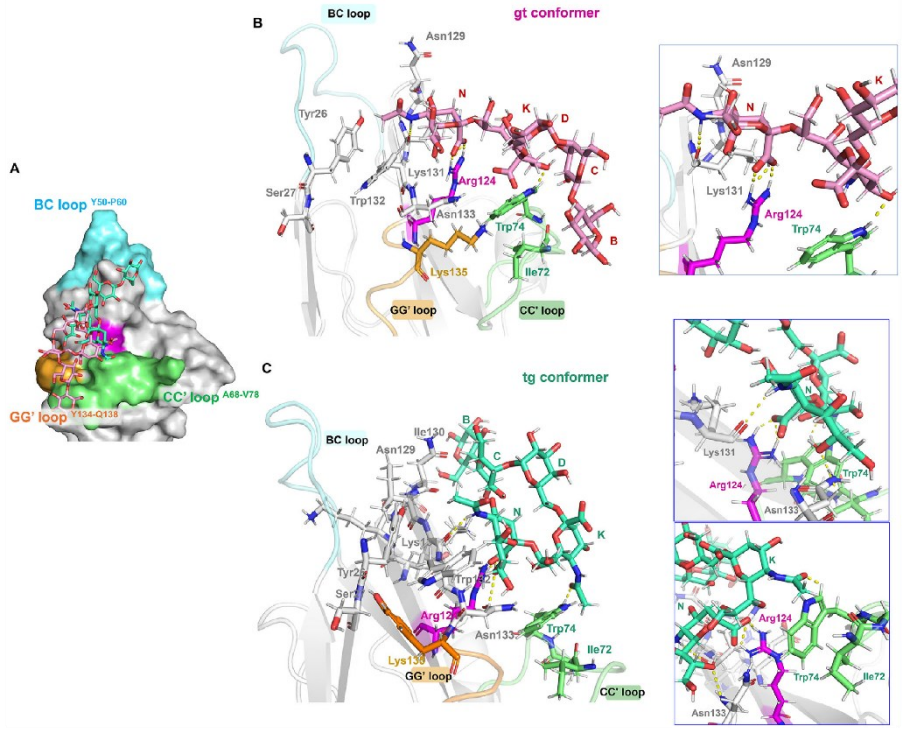


Figure III.11 Representation of the Siglec-7− DSGb3α6 complex in both gt (pink) and tg (cyan) conformations with the BC, CC', and GG' loops highlighted in cyan, green and orange, respectively. A) Different views of the Siglec-7− DSGb3α6 complex in the gt conformation highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend). B) Different views of the Siglec-7− DSGb3α6complex in the *tg* conformation highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the aminoacids were colored according to the loops legend).

To investigate whether the glycosidic linkages of gangliosides could affect their binding accommodation, apart from the main ganglioside GD3 we further evaluated the molecular binding with modified GD3 structures. Therefore, we used two Gb3 derivatives, DSGb3α3 and DSGb3α6, which maintained the disialylated Neu5Ac-α-(2 8)-Neu5Ac portion and varied in the linkage of the internal sialic acid to galactose and in length of the glycan chain We found that the K_b values follow the order GD3>DSGb3α3≈DSGb3α6, thus GD3 has the

highest affinity for the protein (Table 1). To dis entangle the intricate observed behavior, we evaluated the thermodynamics parameters (namely $\Delta_bH°$ and $\Delta_bS°$) of the interaction by determining the temperature-dependence of K_b. We found that the $\Delta_bH°$ is negative for all the ligands highlighting the formation of non-covalent interactions (as H-bonds, salt bridges) between the residues in the binding pocket of the protein and the sugars. Interestingly, the $\Delta_bH°$ values are very similar for all the ligands, indicating the enthalpy contribution to the binding Gibbs energy is not responsible for the observed alteration in the binding constant. Indeed, the sialic acid moieties of GD3, DSGb3α3, and DSGb3α6 are the only residues involved in the interaction with Siglec-7, with a bias for the terminal Sia of the Neu5Ac-α-(2-8)-Neu5Ac glycosidic linkage, while the other sugars result solvent exposed and, consequently, represent the more flexible portions of the ligands. On the other hand, highlights significant differences among the three ligands in the entropy change of binding ($\Delta_bS°$). Of note, the $\Delta_bS°$ is more negative for ligands DSGb3α3 and DSGb3α6 with paid for complex formation is higher for ligands DSGb3α3 and DSGb3α6. Several reasons can be invoked for such observations. However, our NMR and computational results pointed out how ligands DSGb3α3 and DSGb3α6 possess higher rotational degrees of freedom in the unbound state compared to ligand GD3. Consequentially, upon complex formation, the reduction of such degrees of freedom is more prominent for ligands DSGb3α3 and DSGb3α6, leading to more negative $\Delta_bS°$, which in turn, causes an increase of $\Delta_bG°$, disfavoring the complex formation. Notably, the most negative entropy was observed for the complex with ligand DSGb3α6, which indeed shows a conformational equilibrium between *tg* and *gt* rotamers in the bound state. Furthermore, protein-based NMR experiments could profile the binding site of Siglec-7. In all complexes the amino acids mostly affected by the gangliosides corresponded to those located in the proximity of Arg124, crucial for sialoglycan binding since it establishes the key salt bridge with the carboxylate group of the terminal Sia. For example, Gly128, as well as the region close to and partially including the GG' loop, the residues from Ile130 to Tyr134 GG',

were affected by the largest CSP (Gly128, Ile130, Lys131, Trp132 GG', and Tyr134 GG') and decrease of signal intensity (Lys131 and Asn133GG') in all the three complexes. Additionally, other residues (Tyr26, Ser27, and His62 BC) structurally close to the binding region were found to be affected by significant CSPs in the presence of all the three ligands. These results support the importance of the binding site neighboring Arg124, which always establishes a key salt bridge with the carboxylate group of the terminal sialic acid. In particular, the terminal Sia of all gangliosides establishes hydrogen bonds mainly through its carboxylate group, NH of the acetyl group and the glycerol chain with the arginine sidechain. Interestingly, in the structural models of the complexes, we also observed this residue in close contact with the aromatic residue Trp132GG', further stabilized by π–π interactions with Tyr26. This model is backed by the NMR data, since both aromatic residues experience significant CSPs in the presence of all the three ligands. Moreover, two different lysine residues in the binding site of Siglec-7 are also perturbed in the NMR spectra: a sizable CSP is observed for Lys135GG' in the presence of both GD3 and DSGb3α3 and decrease in signal intensity is detected for Lys127 in the presence of both GD3 and DSGb3α6. Overall, the numerous chemical shift perturbations and the lower effects in signal intensity decrease of the residues in the binding site of Siglec-7 agree with the binding affinities in the micromolar range calculated by fluorescence analyses. This is also consistent with the binding affinities calculated with other gangliosides; indeed, Siglecs–carbohydrate interactions typically exhibit low affinity, relying on high avidity conferred by multivalency and ligand clustering[33,34].The combination of the results of protein-based NMR assay and ligand-based STD experiments is instrumental in delineating binding site characteristics of Siglec-7, which is found to be influenced by interactions with ligands GD3, DSGb3α3, and DSGb3α6.

3.3 Conclusion

Gangliosides are sialylated glycosphingolipids ubiquitous in all human cells, very abundant in the brain and nervous system. Based on their location, they play several biological roles under healthy conditions, such as cell–cell and cell-matrix interactions, modulation of signal transduction, and coordination of signal events with other cells. However, ganglioside composition can be altered during malignant transformation and high levels of ganglioside expression contribute to pathological conditions, such as GD3 and GD2 in melanoma and neuroblastoma respectively, leading to tumor formation and progression. Moreover, the presence of gangliosides was detected in the blood of certain cancer patients, making them potential biomarker for tumor diagnosis and/or prognosis or targets for antibody therapies[35,36]. In vitro studies have shown that most gangliosides containing one or two sialic acids have suppressive effects against the activation of immune cells, such as B cells, NK cells, dendritic cells, and monocytes, promoting tumor progression. Inhibition of NK cells activity and regulation of NKs cytotoxicity by GM3, GM2, and GD3, also occur through interaction with Siglecs. Indeed, GD3 is strongly engaged by Siglec-7 on NK cells, leading to immunosuppressive responses in vitro and in vivo[37]. Furthermore, dehydroxylated ceramide-containing GD3 reduced the sensitivity to NK cells, indicating that modifications in ceramide-containing glycosphingolipids not only influence Siglec-7 binding properties but also impact the biological effects triggered by their interactions[38]. Notably, cancers typically show significantly lower levels of FA2H (fatty acid 2-hydroxylase) than corresponding normal tissues, such as skin and colon. Consequently, loss of hydroxylation of ceramides in cancers may enhance the ability of cancer cells to evade NK cell attacks, leading to the escape of cancer cells from immune surveillance system[38]. Siglec-7 is an inhibitory immune receptor that is emerging as a significant target of interest for cancer immunotherapy. Siglec-7 ligands offer valuable insights into motifs and key elements essential for its engagement and the consequent evasion of immune recognition. This interaction establishes Siglec-7 and corresponding ligands as innovative glyco-

immune checkpoints in the realm of cancer treatment. In this frame, the structure of the N-terminal V-set domain of Siglec-7 interacting with a synthetic analog of GT1b was solved[39]. The presence of the branch and the axial position of the Gal C4-O make GT1b highly restricted, influencing the ligand's conformation and recognition. Despite Neu5Acα(2,8)Neu5Acα glycosidic linkage in GT1b adopts a stereochemical arrangement that varies between antiperiplanar and anticlinal with respect to the synclinal conformation observed in GD3andGb3derivatives studied in this work, the terminal sialic acid remains the major determinant of ligands' binding in all cases. Additionally, hydrogen bonds involving terminal Neu5Ac N5 with Lys131-CO and terminal Neu5Ac-O8 with Asn133-N were consistently observed. However, in GT1b, Neu5Ac-O9 establishes H-bonds with Lys135-NZ and Asn133-CO, whereas in GD3, the diastereotopic protons of the glycerol chain are more distant from the protein. Sato et al. proposed a model of Siglec-7 in which two binding sites contribute to glycan recognition: the primary site, where R124 is located, and a secondary site containing other arginine residues, R67 and especially R94, both situated in a cavity placed under the CC' loop. In our complexes, the salt bridge between the carboxyl group of the non-reducing terminal Sia and Arg124 remains stable for the entire simulations, confirming it as the primary site, excluding Arg67 and Arg94 from the interaction[40,41]. In the context of the Siglecs-gangliosides axis of modulation of immune response and tumor microenvironment, antibody therapies have been investigated in pre-clinical and clinical studies[42] and synthetic mimetics of carbohydrate portions of GD2 and GD3 gangliosides have been tested as potential vaccines,[43] although they only partially succeed. Therefore, the effectiveness of ganglioside mimetic vaccines needs to be improved to be effective. Consequently, molecular-level understanding of ganglio sides interaction with Siglec-7 can not only shed light on the mechanisms underlying immune evasion in cancer, but also help the development of targeted strategies to modulate these interactions. With this aim, we combined structural biology methodologies, NMR techniques, biophysical studies, and computational approaches to

provide 3D models and binding affinities of Siglec-7 interacting with the carbohydrate moieties of its preferred ganglioside, GD3, and structurally related Gb3 derivatives. Notably, the lipid portion of ceramide can enhance the interaction of gangliosides with Siglec-7 and may influence the orientation of the C-C′ loop; however, the underlying molecular mechanisms remain to be investigated[40]. By incorporating the unique effects of each ligand on the protein structure and dynamics, this deep understanding of Siglec-7 interaction with ganglioside and their impact on binding site architecture can be achieved for further optimization and drug development. In this work, we highlighted the crucial role of thermodynamics in explaining the binding affinities of protein ligand interactions. Indeed, despite GD3 and Gb3 derivatives show comparable binding epitopes and are recognized by a similar set of amino acids in the primary site of Siglec-7, the ligands' conformational behavior influences the formation of the complexes, and thus the affinity of the binding. Indeed, the higher flexibility of Gb3 derivatives, especially DSGb3α6, reflects in a loss of entropy upon binding to Siglec-7, increasing $\Delta_b G°$ and reducing affinity. In addition, Siglec-7′s reported binding to $\alpha(2\rightarrow3)$ and $\alpha(2\rightarrow6)$ linked sialosides underscores the need for comparative analyses of its interactions with these glycans, following our focus on $\alpha(2\rightarrow8)$-linked gangliosides. A direct comparison of Siglec-7 binding to all three linkages ($\alpha(2\rightarrow3)$, $\alpha(2\rightarrow6)$, and $\alpha(2\rightarrow8)$) will be essential for a comprehensive understanding of its ligand specificity. This is particularly important because these link ages exhibit distinct structural presentations and biological roles. Therefore, exploring the structural diversity of gangliosides and their dynamic interplay with Siglec-7 provides an understanding of the intricate cellular communication processes, offering po tential avenues for therapeutic interventions to restore immune recognition and enhance anti-cancer immune responses.

3.4 References

[1] R.Schauer,J.P.Kamerling,Adv. Carbohydr. Chem. Biochem. 2018, 75, 1.

[2] C. Di Carluccio, R. E. Forgione, A. Molinaro, P. R. Crocker, R. Marchetti, A. Silipo, Carbohydr. Chem. 2021, 44, 31.

[3] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, Essentials of Glycobiology, 4th ed., Cold Spring Harbor Laboratory Press, New York 2017.

[4] S. Groux-Degroote, Y. Guérardel, S. Julien, P. Delannoy, Biochemistry (Moscow) 2015, 80, 808.

[5] R.L.Schnaar,Glycoconjugate J. 2023, 40, 159.

[6] R.L.Schnaar,J. Mol. Biol. 2016, 428, 3325.

[7] R. L. Schnaar, R. Gerardy-Schahn, H. Hildebrandt, Physiol. Rev. 2014, 94, 461.

[8] Y. Ohmi, O. Tajima, Y. Ohkawa, Y. Yamauchi, Y. Sugiura, K. Furukawa, K. Furukawa, J. Neurochem. 2011, 116, 926.

[9] K. Furukawa, M. Kambe, M. Miyata, Y. Ohkawa, O. Tajima, K. Furukawa, Cancer Sci. 2014, 105, 52.

[10] E. Rodrigues, M. S. Macauley, Cancers 2018, 10, 207.

[11] D. Bordoloi, A. J. Kulkarni, O. S. Adeniji, M. Betina Pampena, P. S. Bhojnagarwala, S. Zhao, C. Ionescu, A. Perales-Puchalt, E. M. Parzych, X. Zhu, A. R. Ali, J. Cassel, R. Zhang, M. R. Betts, M. Abdel Mohsen, D. B. Weiner, Sci. Adv. 2023, 9, adh4379.

[12] S. van de Wall, K. C. M. Santegoets, E. J. H. van Houtum, C. Büll, G. J. Adema, Trends Immunol. 2020, 41, 274.

[13] U. Krengel, P. A. Bousquet, Front. Immunol. 2014, 5, 325.

[14] A. Ito,K.Handa,D.A.Withers,M.Satoh,S. itirohHakomori,FEBS Lett. 2001, 504, 82.

[15] T. Yamaji, T. Teranishi, M. S. Alphey, P. R. Crocker, Y. Hashimoto, J. Biol. Chem. 2002, 277, 6324.

[16] R. I. Ramos, M. A. Bustos, J. Wu, P. Jones, S. C. Chang, E. Kiyohara, K. Tran, X. Zhang, S. L. Stern, S. Izraely, O. Sagi-Assif, I. P. Witz, M. A. Davies, G. B. Mills, D. F. Kelly, R. F. Irie, D. S. B. Hoon, Mol. Oncol.

2020, 14, 1760.
[17] F. Malisan, R. Testi, BBA- Mol. Cell Biol. Lipids 2002, 1585, 179.
[18] T. J. Webb, X. Li, R. L. Giuntoli, P. H. H. Lopez, C. Heuser, R. L. Schnaar,M.Tsuji,C.Kurts,M.Oelke,J.P.Schneck,Cancer Res. 2012, 72, 3744.
[19] A. Bärenwaldt, H. Läubli, Expert Opin. Ther. Targets 2019, 23, 839.
[20] J. Daly, M. Carlsten, M. O'Dwyer, Front. Immunol. 2019, 10, 1047.
[21] J. Liu, X. Zheng, X. Pang, L. Li, J. Wang, C. Yang, G. Du, Acta Pharm. Sin. B 2018, 8, 713.
[22] P. Li, S. Huang, P. Chiang, C. Fan, L. Guo, D. Wu, T. Angata, C. Lin, Angew. Chem., Int. Ed. 2019, 131, 11395.
[23] C. H. Kim, Ganglioside Biochemistry, Springer, New York, NY 2020, pp. 1–13.
[24] R. Nepravishta, J. Ramírez-Cárdenas, G. Rocha, S. Walpole, T. Hicks, S. Monaco, J. C. Muñoz-García, J. Angulo, J. Med. Chem. 2024, 67, 10025.
[25] C. Abreu, C. Di Carluccio, T. Jeˇ cmen, O. Skoˇrepa, J. Bláha, R. Marchetti, A. Silipo, O. Vanˇ ek, Int. J. Biol. Macromol. 2025, 309, 142672.
[26] R. E. Forgione, C. Di Carluccio, J. Guzmán-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, iScience 2020, 23, 101401.
[27] R. Oliva, N. Jahmidi-Azizi, S. Mukherjee, R. Winter, J. Phys. Chem. B 2021, 125, 539.
[28] F. Nieto-Fabregat, M. P. Lenza, A. Marseglia, C. Di Carluccio, A. Molinaro, A. Silipo, R. Marchetti, Beilstein J. Org. Chem. 2024, 20, 2084.
[29] C. O. Soares, A. S. Grosso, J. Ereño-Orbea, H. Coelho, F. Marcelo, Front. Mol. Biosci. 2021, 8, 727847.
[30] C. Di Carluccio, R. E. Forgione, A. Bosso, S. Yokoyama, Y. Manabe, E. Pizzo, A. Molinaro, K. Fukase, M. Fragai, B. A. Bensing, R. Marchetti, A. Silipo, RSC Chem. Biol. 2021, 2, 1618.

[31] C. Di Carluccio, E. Crisman, Y. Manabe, R. E. Forgione, A. Lacetera, J. Amato, B. Pagano, A. Randazzo, A. Zampella, R. Lanzetta, K. Fukase, A. Molinaro, P. R. Crocker, S. Martín-Santamaría, R. Marchetti, A. Silipo, ChemBioChem 2020, 21, 129.
[32] R. E. Forgione, C. Di Carluccio, M. Kubota, Y. Manabe, K. Fukase, A. Molinaro, T. Hashiguchi, R. Marchetti, A. Silipo, Sci. Rep. 2020, 10, 1589.
[33] P. Li, S. Huang, P. Chiang, C. Fan, L. Guo, D. Wu, T. Angata, C. Lin, Angew. Chem., Int. Ed. 2019, 58, 11273.
[34] Y. Liu, M. Yan, M. Wang, S. Luo, S. Wang, Y. Luo, Z. Xu, W. Ma, L. Wen, T. Li, ACS Cent. Sci. 2024, 10, 417.
[35] S. Groux-Degroote, P. Delannoy, Int. J. Mol. Sci. 2021, 22, 6145.
[36] Q.Li,M.Sun,M.Yu,Q.Fu,H. Jiang,G.Yu,G.Li,Glycoconjugate J. 2019, 36, 419.
[37] C. Iltis, L. Seguin, L. Cervera, L. Duret, T. Hamidouche, S. Kunz, O. Croce, C. Delannoy, Y. Gue´rardel, F. Allain, L. Moudombi, P. Hofman, E. Cosson, J. Guglielmi, T. Pourcher, V. M. Braud, M. Shkreli, M. C. Michallet, C. C. Feral, E. Gilson, J. Cherfils-Vicini, bioRxiv 2021, 440408.
[38] N. Hashimoto, S. Ito, A. Tsuchida, R. H. Bhuiyan, T. Okajima, A. Yamamoto, K. Furukawa, Y. Ohmi, K. Furukawa, J. Biol. Chem. 2019, 294, 10833.
[39] H. Attrill, A. Imamura, R. S. Sharma, M. Kiso, P. R. Crocker, D. M. F. Van Aalten, J. Biol. Chem. 2006, 281, 32774.
[40] N. Yamakawa, Y. Yasuda, A. Yoshimura, A. Goshima, P. R. Crocker, G. Vergoten, Y. Nishiura, T. Takahashi, S. Hanashima, K. Matsumoto, Sci. Rep. 2020, 10, 8647.
[41] A. Yoshimura, R. Hatanaka, H. Tanaka, K. Kitajima, C. Sato, Biochem. Biophys. Res. Commun 2021, 534, 1069.
[42] S. Cavdarli, P. Delannoy, S. Groux-Degroote, Cells 2020, 9, 741.
[43] W.Tong,M.Maira,R.Roychoudhury,A.Galan,F.Brahimi,M.Gilbert, A. M. Cunningham, S. Josephy, I. Pirvulescu, S. Moffett, H. U. Saragivi, Cell Chem. Biol. 2019, 26, 1013.
[44] C. Di Carluccio, F. Milanesi, M. Civera, C. Abreu, S. Sattin, O.

Francesconi, A. Molinaro, O. Vanˇ ek, R. Marchetti, A. Silipo, Eur. J. Org. Chem. 2023, 26, 202300644.
[45] J. M. Pröpster, F. Yang, B. Ernst, F. H. T. Allain, M. Schubert, Protein Expression Purif. 2015, 109, 14.
[46] R. Keller, The Computer Aided Resonance Assignment,CANTINAVer lag, Germany 2004.
[47] M. P. Williamson, Prog. Nucl. Magn. Reson. Spectrosc. 2013, 73,1.
[48] S. Giuntini, E. Balducci, L. Cerofolini, E. Ravera, M. Fragai, F. Berti, C. Luchinat, Angew. Chem. 2017, 129, 15193.
[49] M. Mayer, B. Meyer, J. Am. Chem. Soc. 2001, 123, 6108.
[50] B. Meyer, T. Peters, Angew. Chem.– Int. Ed. 2003, 42, 864.
[51] M.M.B.Ribeiro, H. G. Franquelim, M. A. R. B. Castanho, A. S. Veiga, J. Pept. Sci. 2008, 14, 401.
[52] N.Jahmidi-azizi, S. Gault, C. S. Cockell, R. Oliva, R. Winter, Int. J. Mol. Sci. 2021, 22, 10861.
[53] GLYCAM-Web. Utilities for molecular modeling of carbohydrates WoodsGroup(2005-2024)(accessed:April2025),http://glycam.org.
[54] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M.K.Gilson,H.Gohlke,A.W.Goetz,D.Greene,R.Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M.Merz, Y. Miao, G. Monard, et al., AMBER, University of California, San Francisco 2018.
[55] D. R. Roe, T. E. Cheatham, J. Chem. Theory Comput. 2013, 9, 3084.

IV. Chapter IV: Mechanistic insights into Siglec-7 recognition of exogenous ligands

JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau
Article

Fusobacterium nucleatum Lipopolysaccharides O-Antigen Defines a Novel Siglec-7 Binding Epitope

Cristina Di Carluccio,[‡] Ferran Nieto-Fabregat,[‡] Linda Cerofolini, Celeste Abreu, Luis Padilla-Cortés, Giulia Roxana Gheorghita, Alessandro Antonio Masi, Lorena Buono, Manasik Gumah Adam Ali, Dimitra Lamprinaki, Antonio Molinaro, Nathalie Juge, Giovanni Smaldone, Ondřej Vaněk, Marco Fragai, Roberta Marchetti,* and Alba Silipo*

Cite This: https://doi.org/10.1021/jacsau.5c00810
Read Online

4.1 Introduction

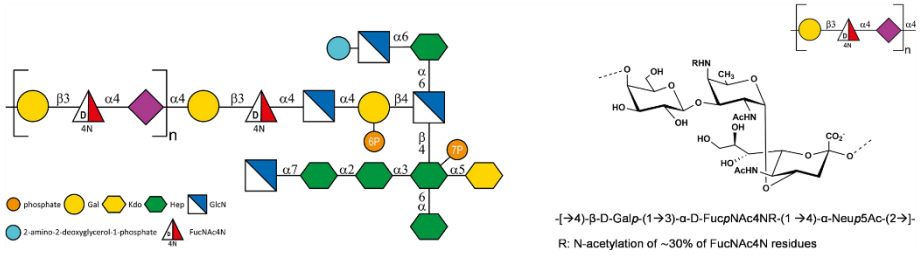
Fusobacterium nucleatum (Fn)[1,2], is a Gram-negative obligate anaerobe, among the most abundant species residing in the human oral cavity and rarely found in the lower gastrointestinal (GI) tract of healthy individuals[1]. *Fn* has been identified as one of the pathobionts that outgrow during dysbiosis preceding periodontal disease[2]. *Fusobacterium* can disseminate systemically in different body sites, such as the genitourinary tracts, and overcome placental and blood-brain barriers. A large number of studies suggest that *Fn* is closely related to the developments of systemic diseases, including rheumatoid arthritis, Alzheimer's disease, adverse pregnancy outcomes, and GI disorders such as IBD[3]. In addition, *Fn* has gained attention as emerging cancer-associated bacteria, overabundant in various types, including colorectal (CRC), pancreatic, esophageal, and breast cancers[4,5], and associated with shorter patients survival. In particular, CRC is associated with gut microbiota dysbiosis characterized by a reduction in protective or beneficial bacteria such as *Clostridium* and *Faecalibacterium* and an increase in cancer-associated polyketide synthase *Escherichia coli*, enterotoxigenic *Bacteroides fragilis* and *Fn*[2,6,7]. *Fn* promotes CRC progression via several mechanisms, including inhibition of host antitumor immunity, innate immune cell modulation, activation of cell proliferation, promotion of cellular invasion, induction of

chronic inflammation and immune evasion, and mobilization of immune cells in the tumor microenvironment.[7,8] A virulence factor in *Fn* is the lipopolysaccharide (LPS), a microbe-associated molecular pattern located on the outer membrane of Gram-negative bacteria. LPS is an amphiphilic molecule with a general structural architecture composed of (i) a glycolipid moiety named lipid A and (ii) a heteropolysaccharide portion containing the core oligosaccharide and the O antigen polysaccharide. Structural studies on the LPS of *Fn* have identified a wide array of carbohydrate components, including sialic acids and/or unusual sugars like fusaminic acid[3], highlighting the complexity of its glycome. These sugars contribute to the unique immunogenic properties of the bacterium, playing critical roles in its pathogenicity and immune evasion.

The sialic acid-binding immunoglobulin-type lectins (Siglecs) comprise a family of 15 cell surface proteins, mostly inhibitory receptors, which can be divided into two different classes according to their sequence and evolutionary similarity: the main subset of CD33-related Siglecs and the conserved Siglecs' subgroup (Scheme 1). Siglecs mostly recognize sialic acid (Sia) containing glycans through a key conserved arginine residue in the N-terminal domain that establishes a salt bridge with the carboxylate group of Sia. Therefore, engagement of sialylated glycoprotein and glycolipids on all mammalian cells by Siglecs induces tolerance to self-antigens and prevents unwanted autoimmune responses. However, hypersialylation, a frequent trait of tumor cells, increases sialoglycan−Siglec interactions to evade immune surveillance in the favorable immunosuppressive tumor microenvironments, making Siglecs a novel emerging immune checkpoint for cancer therapy[9]. Interestingly, several pathogens, including bacteria like *E. coli* K1, *Neisseria meningitidis*, *N. gonorrheae*, group B *Streptococcus* (GBS), and *Campylobacter jejuni*, and viruses, such as HIV-1, SARS-CoV-2, Ebola virus, have evolved cell-surface Sia mimicry as a mechanism for engaging Siglecs, thus attenuating host immune responses and promoting dissemination. This mimicry is achieved in

bacteria through envelope components (e.g., sialylated capsules and/or LPS) or, in some cases, modified flagellin[10,11].

Siglec-7 (CD328), a member of the CD33-related Siglecs subgroup, is primarily expressed on the surface of innate lymphoid human NK cells and is also found on dendritic cells[12]. It exhibits a binding preference for α2,8-linked disialylated ligands and internally branched α2,6-sialyl residues. Siglec-7 consists of an extracellular region, containing an N terminal V-set domain representing the carbohydrate binding region, two C2-type Ig-like domains, and an immunoreceptor tyrosine-based inhibitory (ITIM) motif in the cytosolic region (Scheme S1)[13].Siglec-7 *trans* interaction with cognate ligands drives the phosphorylation of its cytoplasmic ITIM domain that triggers the inhibition of NK cell pathways. Tumor cells and pathogens can exploit this mechanism to evade immune recognition and facilitate their migration through the circulatory system. For example, GBS surface protein β binds Siglec-7 on NK cells, inhibiting pyroptosis and suppressing NK cell sentinel activity[14].



Scheme IV-1. SNFG Schematic representation of *Fn10953* LPS structure. (A) OPS containing the O-antigen repeating unit and the core oligosaccharide. 40 (B) O-Antigen trisaccharide repeating unit as derived by the mild acid hydrolysis of Fn10953 O-polysaccharide chain: →4)-β-D-Galp-(1→3)-α-D-FucpNAc4NR-(1→4)-α-Neup5Ac-(2→; R = H (up to 75%), R = Ac (25-30%).

The selective recognition of sialylated glycans exposed on *C. jejuni* lipooligosaccharides by Siglec-7 has also been reported and potentially related to the clinical outcome and the development of secondary complications such as Guillain-Barré syndrome[15]. *Pseudomonas aeruginosa* sialylation contributes to bacterial pathogenicity and interaction with the host via reduction of complement deposition and with Siglec-dependent recognition. These interactions demonstrate that pathogens, including *Fusobacterium nucleatum*,

can exploit Siglec mediated immune modulation as a shared strategy across microbial and viral systems[16]. By leveraging sialic acid recognition, these pathogens suppress immune activation and evade host defenses, promoting successful colonization and persistence in diverse pathological contexts such as cancer and infectious diseases[11]. We previously showed that *F. nucleatum* ssp. *animalis* ATCC51191, the most predominant species in CRC, interacts with Siglec-7 expressed on innate immune cells, and that this interaction also occurred with purified bacterial outer membrane vesicles and LPS[17]. This interaction induced a pro-inflammatory profile in human monocyte-derived dendritic cells and a tumor associated profile in human monocyte derived macrophages, likely contributing to tumor progression, and unveiled LPS O-antigen as a potential ligand for Siglec-7. Specifically, we focused on *F. nucleatum* ssp. *polymorphum* 10953 (Fn10953), a strain frequently isolated from colorectal cancer tissues, whose LPS O-antigen possesses a unique repeating unit including peculiar sugars as the neuraminic acid residue and the AAT (FucpNAc4N) moiety (Scheme 1)[18]; the precise molecular mechanisms by which this O-antigen is recognized by human Siglec-7 have yet to be elucidated. To unravel this essential host−pathogen interaction, we employed a powerful combination of complementary approaches, including biochemical and biophysical techniques, comprehensive NMR spectroscopy from both protein and ligand perspectives, and advanced computational modeling. This integrated strategy allowed us to fully characterize the binding of Fn10953 LPS to Siglec-7, defining a previously unrecognized and novel molecular recognition mode.

4.2 Results

The full extracellular domain (FED) of Siglec-7 was expressed in HEK293 GnTI- (human embryonic kidney) cells[19]. The LPS from Fn10953 was extracted and purified following established methodology, confirming the composition of the O-antigen repeating unit (Scheme 1). Notably, the acetylation of the FucpNAc4N (2-acetamido-4-amino-2,4,6-trideoxy-D

galactose, AAT) unit at N4 was around 25−30%, as previously reported[18]. Furthermore, a series of oligomers containing an increasing number of O-antigen (OPS) repeating units were generated through mild acid hydrolysis (Scheme 1) exploiting the lability of the sialic acid glycosidic linkage. These oligomers were employed to elucidate the molecular basis of *Fn*10953 O antigen recognition and binding by Siglec-7.

4.2.1 Fn10953 LPS Binds Siglec-7 and Differentially Activates Blood Cells

To assess the recognition of Fn10953 LPS by Siglec-7, we first evaluated its binding to cells overexpressing Siglec-7. To this end, multiple hematopoietic cell lines were subjected to flow cytometric analysis using a PE-conjugated anti-Siglec-7 antibody (see methods for details). Among these cell lines, RPMI8226, a well-established model of multiple myeloma, demonstrated the highest surface expression of Siglec-7. To assess the interaction of Fn10953 LPS and its isolated O-antigen with Siglec-7, RPMI8226 cells were incubated with increasing concentrations of these ligands.

Subsequently, binding to Siglec-7 was evaluated by flow cytometry using a PE-conjugated anti-Siglec-7 antibody. As shown in Figure 1A, both Fn10953 LPS and its O-antigen bound Siglec-7-expressing RPMI8226 cells (IC50 of approximately 2 μg/mL for LPS and 18 μg/mL for OPS), while no binding was detected with Salmonella LPS, used as a negative control. Next, steady-state fluorescence analysis was conducted to determine the binding affinities of Siglec-7 for the Fn10953 O antigen. We observed a concentration-dependent reduction in fluorescence intensity of Siglec-7 upon glycan addition. The binding constants (K_b) were determined by nonlinear regression analysis (Figure 1B)[20]. No change in fluorescence intensity was observed when using the trisaccharide containing a single repeating unit (1-Mer), suggesting that it was insufficient for Siglec-7 recognition. In contrast, titration with longer oligomers, containing up to four repeating units (4-Mer), resulted in quenching

of Siglec-7 tryptophan residues, indicative of complex formation. These interactions exhibited comparable low-micromolar K_D similar to that observed for the entire LPS O-antigen (Figure 1B).

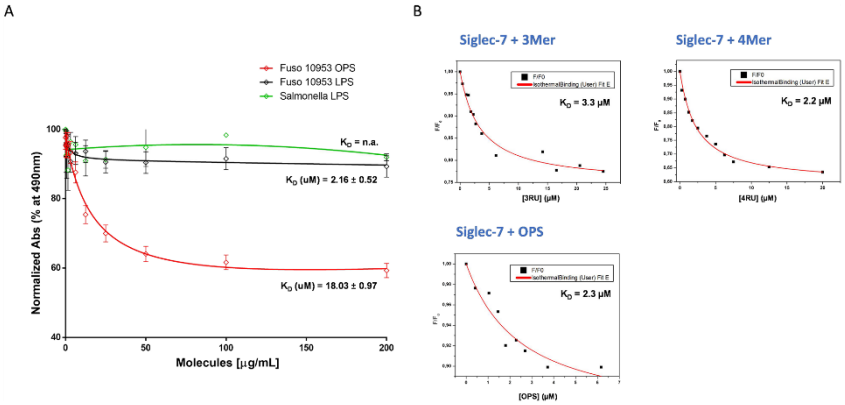


Figure IV.1 Recognition of Fn10953 LPS by Siglec-7 via ELISA and Fluorescence Analysis: (A) ELISA binding assay of Salmonella LPS (green line), Fn10953 LPS (black line), and Fn10953 full polysaccharide (OPS) (red line) with RPMI8226 cell line. The K D 2 (µM) and the corresponding R values are reported in the figure. The lower/poorer quality of the LPS binding data could be due to the steric cluster generated by this molecule, which could cover the binding sites generating a worse signal than OPS. (B) Fluorescence analysis of Siglec-7 with different oligomers from Fn10953. The quenching fluorescence titration of Siglec-7 with Fn10953 3-Mer, Fn10953 4-Mer, and Fn10953 full polysaccharide (OPS) were f itted using a nonlinear regression with the one site-specific binding model for the determination of the association constants K −1 b corresponding K_D (µM).

The effect of Fn10953 LPS and its O-antigen on the levels of Siglec-7 surface expression in blood cells was evaluated using flow cytometry. As expected, the results demonstrated that NK cells exhibited the highest expression of Siglec-7 (Figure 2A). The ability of Fn10953 LPS to activate individual blood components was first investigated using peripheral blood mononuclear cells (PBMCs) from 5 healthy donors by measuring HLA-DR expression levels. HLA-DR, a key class II MHC molecule involved in antigen presentation to CD4+, serves as a marker of immune activation. T lymphocytes, B-lymphocytes, and NK cells were identified, and HLA-DR expression levels on the gated cells were simultaneously assessed. As shown in Figure 2B, a significant increase in HLA-DR positive cells, compared to untreated cells, was observed for Fn10953 O-antigen on NK cells; the other combinations of treatments were not significant. T lymphocytes, B-lymphocytes, and NK cells were then purified and subsequently activated with both Fn10953 LPS and O

antigen. No significant changes were observed in the degree of activation of T-lymphocytes (Figure 2C) and B-lymphocytes (Figure 2D). Significant changes in NK cell activation were observed following treatment with Fn10953 O-antigen compared to untreated NK cells or those treated with Fn10953 LPS or Salmonella LPS, consistent with the findings observed in PBMCs (Figure 2E). These data, although preliminary, suggest the involvement of Siglec-7 in the activation process of NK cells in the presence of Fn10953 O-antigen.

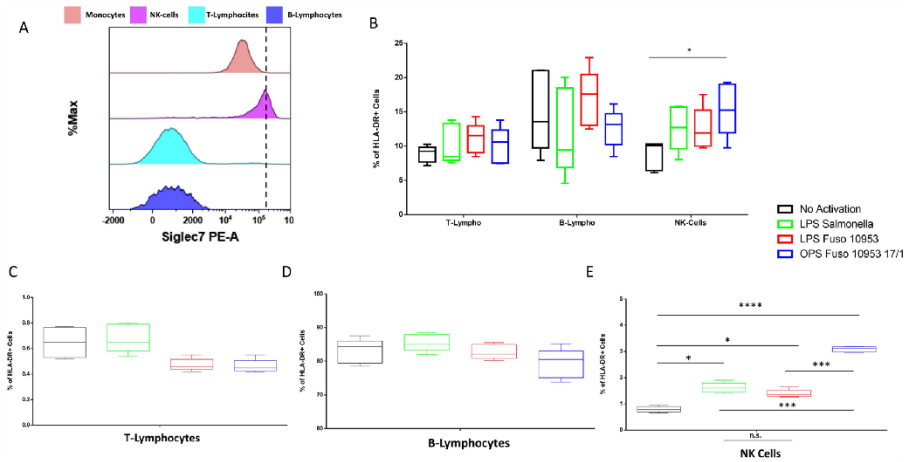


Figure IV.2 Siglec-7 protein expression on blood cells. (A) Cytofluorimetric analyses of Siglec-7 protein levels in monocytes (pink), NK cells (magenta), T-lymphocytes (light blue), and B-lymphocytes (blue). Percentage of HLA-DR+ cells in PBMC derived from 5 healthy donors (B) and in purified B-lymphocytes (C), T-lymphocytes (D), and NK cells (E) from the same subjects. Black bars represent untreated cells. Green bars represent Salmonella LPS-treated cells. Red bars represent Fn10953 LPS-treated cells. Blue bars represent Fn10953 O-antigen treated cells. *p-value <0.05,***p-value <0.001, ****p-value <0.00001; n.s., not significant. Paired *t* test.

4.2.2 Molecular Details of Siglec-7 and Fn10953 O-Antigen Oligomers

The molecular recognition features of Fn10953 LPS by Siglec-7 were elucidated through a combined approach of NMR spectros copy and MD (molecular dynamics) simulations. Ligand-based NMR techniques, including saturation transfer difference (STD), transferred NOESY (tr-NOESY), and relaxation experiments, were employed to investigate the molecular interactions between Siglec-7 and Fn10953 O-antigen oligomers. These

experiments elucidated the ligand binding epitopes and their conformational behavior upon binding[21]. In with fluorescent binding assays, no STD NMR effects were observed between Siglec-7 and 1-Mer (data not shown), indicating the absence of binding. This finding confirms that a single repeating unit of the Fn10953 O antigen is insufficient for recognition and binding by Siglec-7. Conversely, NMR binding experiments conducted on the core OPS fraction revealed selective binding of Siglec-7 to the O antigen moiety. Notably, no STD NMR effects were observed from protons belonging to the core region. To determine the minimal epitope required for O-antigen binding to Siglec-7, NMR experiments were performed with Fn10953 OPS oligomers. A first indication of complex formation was obtained by comparing the CPMG experiments of the 3-Mer alone and in the mixture with the protein. Indeed, a decrease of signals in the bound state and differences in the transverse relaxation time T 2 upon binding. values were detected STD NMR analysis carried out on the isolated 2-Mer (Figure 3A) revealed that the recognition primarily occurred through the internal sialic acid (N') and the AAT (FucpNAc4N) (A') residues (Figure 2A). Protons at positions 4, 5, and 7 of N' exhibited clear STD NMR enhancements, different from the reducing sialic acid unit N. Furthermore, a slight STD NMR enhancement observed for N'3 eq provided further evidence for Siglec-7 preference for the internal sialic acid residue; additionally, protons at positions 6 and 9 exhibited significant magnetization transfer, further corroborating the crucial role of the N' unit in the interaction with Siglec-7. Moreover, while STD signal intensities of the acetyl groups of N, A and A' residues were quite low, a robust STD response was detected for the acetyl group of N'. Regarding the recognition of the AAT moiety, a strong STD signal was observed for the anomeric proton of A' (the internal AAT), while residue A exhibited no binding signals. Protons at positions 2 and 6 of A' displayed moderate STD effects, whereas the remaining protons of A' showed STD signals below 20%. These findings collectively demonstrate the involvement of the internal AAT A' in binding to Siglec-7. In contrast, galactose units showed no STD NMR signals, demonstrating that they

are not involved in the binding process and away from the protein binding pocket (Figure 3A).

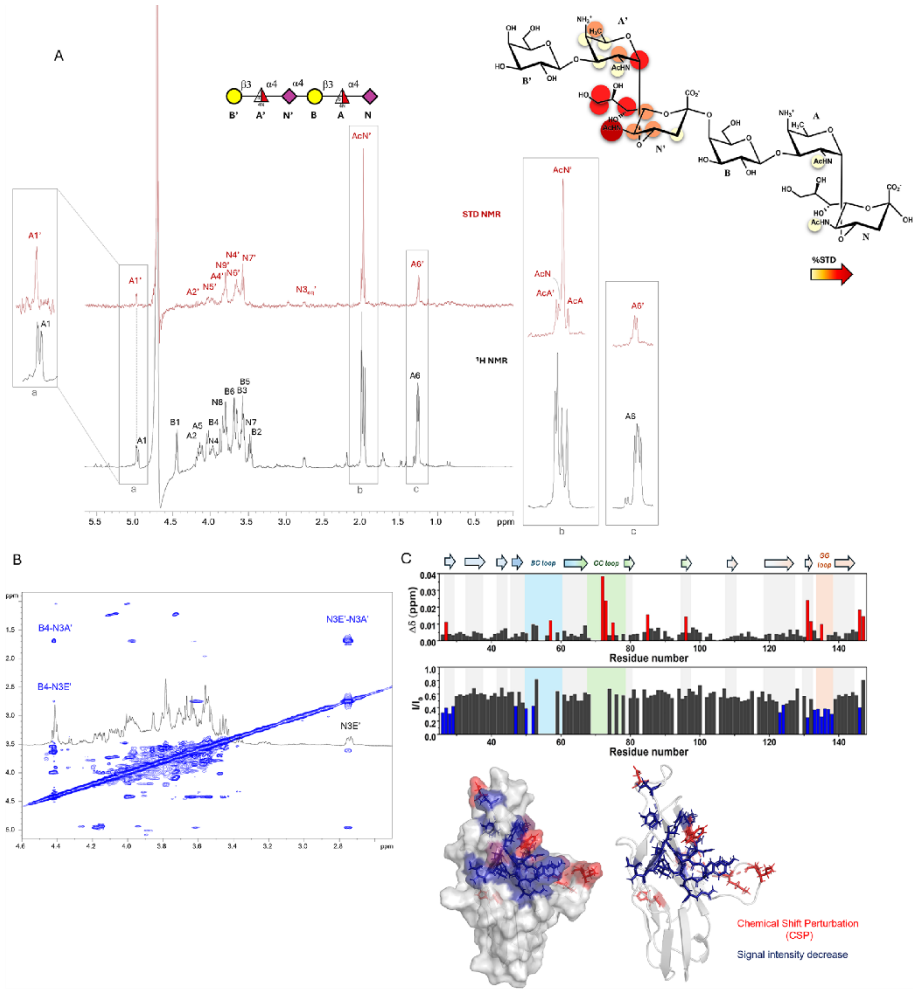


Figure IV.3 NMR studies of the interaction between Fn10953 OPS and Siglec-7. (A) Schematic structure of the 2-Mer ligand depicted following the Symbol Nomenclature for Glycans (SNFG). STD NMR spectrum composed of 1 H NMR reference spectrum (bottom) and 1D STD NMR spectrum (top), of the Siglec-7:2-Mer mixture at a 1:50 ratio. 2D representation illustrating the interacting epitope map resulted from STD NMR data, showcasing the interactions between 2-Mer and Siglec-7. (B) tr-NOESY spectrum of the Siglec-7:2-Mer mixture at a 1:50 ratio. The key NOE contacts shown in the spectrum were indicative of extended bioactive conformation of 2-Mer. (C) Diagrams of the CSP (top) and decrease in signal intensity (bottom) of the amino acids of 200 µM Siglec-7 in the presence of 4 mM OPS from Fn-10953. The CSP effects were evaluated with the formula $\boldsymbol{\Delta\delta} = \left(\frac{1}{2}\right)\sqrt{(\boldsymbol{\Delta\delta_H})^2 + \left(\frac{\boldsymbol{\Delta\delta_N}}{5}\right)^2}$. The residues experiencing the largest CSP (S27, D57, I72, S73, K75, W85, H96, K131, W132, K135, L146, and T147) have been highlighted in red; the residues experiencing the largest decrease in signal intensity (Y26, S27, L28, T29, S47, Y50, V52, F123, R124, K131, N133, Y134, K135, Y136, D137, Q138, and L146) have been highlighted in blue. On the right, the 3D surface of Siglec-7 with the amino acids

experiencing the largest CSP colored in red and those with the largest decrease in signal intensity in blue.

Moreover, no STD contribution was detected from the proton at position 4 of AAT when acetylated, likely suggesting that the low degree of OPS acetylation does not influence or participate in the general binding to Siglec-7. Together, these data showed that the *Fn*10953 2-Mer preferentially accommodated the internal sialic acid and not acetylated AAT residues into the binding site of Siglec-7. To further investigate the binding mode, tr-NOESY NMR and computational studies were combined to generate a 3D model of Siglec-7 and 2-Mer complex. Comparing NOESY and tr-NOESY spectra (Figure 3B), no substantial differences were observed in the NOE contacts or in the interproton distances, indicating that the hexasaccharide containing two repeating units underwent no substantial conformational changes upon binding. Notably, the presence of key NOEs, as the contact of B4−N'3 (axial and equatorial) and B1−A2 (Figure 3B), as well as the absence of the NOE between N'3 with A5′ and A6' identified a ligand conformational preference toward the t conformation (φ = 180°). To depict a 3D complex, the uncommon AAT sugar residue was parametrized using the AMBER18 package[22]; then, the 2-Mer was built according to the energetic minima calculated through the adiabatic maps by molecular mechanics calculation and subjected to MD simulation. The ligand in the free state exhibits conformational flexibility, primarily ascribable to φ (C1−C2−O-C4') torsion angle around Neu5Ac-α-(2,4)-Gal glycosidic linkage, that could populate two different values, corresponding to 180° (conformer *t*) and −60° (conformer-*g*), which lead to an extended and folded shape, respectively. In contrast, the ψ (C1−O−C4'−H4') torsion angle remains relatively stable, maintaining a value around 0°. In detail, the MD simulations corroborated the NOE-based findings. Cluster analysis of the free ligand trajectories revealed that it predominantly resides in the extended t conformation (φ = 180°) for approximately 70% of the simulation time. This computational evidence further supports the notion that the 2 Mer preferentially adopts an extended conformation in its free state.

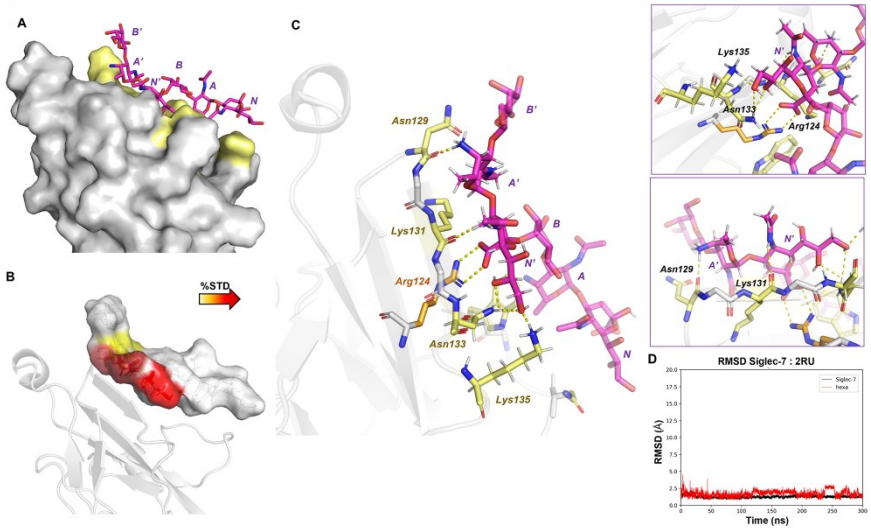


Figure IV.4 Analysis of MD simulation of the Siglec-7:2-Mer complex: (A) 3D model of the Siglec-7:2-Mer complex, showing the protein surface with the interacting amino acids highlighted in yellow. (B) 3D depiction of the interaction, with the ligand surface colored based on STD NMR and computational findings. (C) 3D representation of a dominant pose from the most prevalent MD simulation cluster. The primary amino acids involved in binding are highlighted in yellow, with the key Arg124 in orange. The ligand is depicted in magenta, indicating a preference for an extended conformation in the MD simulations. Dashed yellow lines represent observed hydrogen bond interactions. A detailed view illustrates the interactions between Siglec-7 and the sialic acid (N') and AAT (A') units. (D) Protein (black) and ligand (red) RMSD of the Siglec-7:2Mer system. The ligand RMSD was calculated in reference to the protein.

Similarly and according to tr-NOESY experiments, the 2 Mer adopted an extended conformation upon binding. An almost comparable binding epitope and bioactive conformation were detected for 3-Mer. This suggests that Siglec-7 accommodates and recognizes the 3-Mer in a similar manner. To further elucidate the protein−ligand interaction, protein-based NMR experiments were conducted to identify the glycan binding region on the Siglec-7 surface (Figure 3C). To characterize the complex, a series of NMR titrations were performed and increasing concentrations of 3-Mer were added to a solution of [U-^{15}N]-labeled Siglec-7. The analysis of chemical shift perturbation (CSP) and/or decrease in signal intensity of cross-peaks in the 2D ^{1}H− ^{15}N HSQC spectra enabled the mapping of the Siglec-7 binding pocket. The addition of the ligand mainly induced a decrease in signal intensity, and the residues experiencing the largest decrease in signal intensity were the following: Tyr26, Ser27, Leu28, Thr29, Ser47, F123, R124, Lys131, and Asn133, as well as Tyr50 and Val52 of

BC loop, and the amino acids Tyr134, Lys135, Tyr136, Asp137, and Gln138 of the GG' loop. Otherwise, Ile72, Ser73, and Lys75 of the CC loop were affected by CSP, together with Ser27, Trp85, His96, Lys131, Trp132, Leu146, Thr147, and Asp57 of the BC loop as well as Lys135 of the GG' loop. These results indicate that the key Arg124 and the neighboring amino acids (from Lys131 to Gln138) form the main binding site of Siglec-7, with residues from BC and CC' loops partially involved in the binding with the O-antigen oligomers. The extended, bioactive conformation of 2-Mer was modeled in the V-set Siglec-7 binding site (PDB 2HRL[23]) and once optimized with Maestro Schrödinger software, the complex was subjected to minimizations and MD simulations (Figure 4)[24]. The results demonstrated the stability of the complex (Figure 4D)

and confirmed that 2-Mer preferentially accommodated in the bound state with

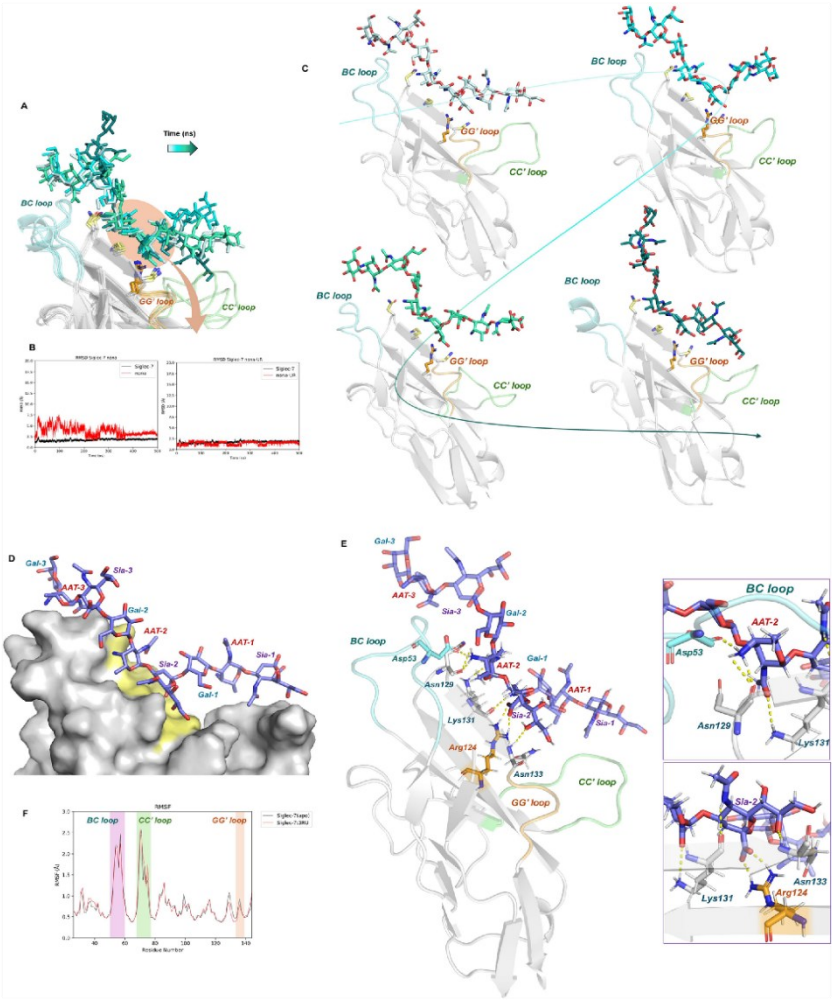


Figure IV.5 Dynamics and stability of the Siglec-7/3-Mer complex across MD simulations. (A) 3D visualization of the Siglec-7/3-Mer complex through the overlay of various representative poses (from light to dark blue) to depict the mobility of the complex during MD, with the 3-Mer ligand shaded in varying tones of blue over time. The BC, CC' and GG' loops are highlighted in cyan, green, and orange, respectively. Amino acids involved in 3-Mer recognition are marked in yellow, with the key Arg124 in orange. (B) Left: RMSD analysis of the Siglec-7/3-Mer system; right: RMSD of the inner trisaccharide (Gal2-AAT2-Sia2), illustrating the internal trisaccharide's stability during MD simulations, in contrast to the 3 Mer's variability due to the movements of terminal and reducing trisaccharides. (C) Temporal representation showcasing simultaneous movements of both terminal and reducing trisaccharides along with the BC and CC' loops, highlighting the anchored internal trisaccharide amidst 'wing-like' movements. (D) 3D model of the Siglec-7/3-Mer complex showing the protein surface with interacting amino acids in yellow. (E) 3D view of a representative pose from the most common MD simulation cluster, with the amino acids involved in the binding colored in gray, Arg124 in orange, and Asp53 from the BC loop in cyan. BC, CC', and GG' loops are again depicted in cyan, green, and orange, respectively, with dashed yellow lines indicating hydrogen bond interactions. A closer inspection reveals the interactions between Siglec-7 and the sialic acid (Sia-2, bottom) and AAT (AAT-2, top). (F) RMSF results showing the protein flexibility in the free (black) and bound (red) states. A slight decrease flexibility is observed only for the GG' loop.

an extended conformation, in full agreement with the tr-NOESY studies. MD results indicated that the 3D complex of Siglec-7 with Fn10953 2-Mer exhibited significant contacts primarily mediated by residues N' and A'. These findings closely aligned with the NMR results (Figure 3). The carboxylate moiety of N' played a crucial role in the interaction, forming a stable salt bridge with the guanidinium group of Arg124. This salt bridge emerged as the most prominent interaction throughout the MD simulation (Figure 4), effectively anchoring and orienting the glycan within the binding site. Further interactions were engaged by N' with Asn133 and Lys131 residues side chains, both located in the GG' loop, through the hydroxyl group at position 8 of the glycerol chain and the NAc moiety, respectively. Additionally, the hydroxyl group at position 9 interacted with both Asn133 and Lys135 residues. On the other hand, Lys131 simultaneously stabilized A' residue by interacting with its acetyl group. Notably, the interaction between Siglec-7 and the AAT residue (A') further involved Asn129, which formed polar contacts with the positively charged amino group at position 4. Furthermore, Asn70 and Trp74 of the CC' loop established transient hydrogen bonds interactions with the NAc moiety of the reducing N residue. Overall, these computational studies combined with STD and tr-NOESY NMR results, indicated that the 2-Mer adopted an extended conformation, with residues N' and A' playing crucial roles in the interaction with Siglec-7, engaged in polar, charged, and hydrophobic interactions. To construct a 3D model of the Siglec-7/3-Mer complex, the nonasaccharide was modeled into the receptor binding pocket. This positioning was guided by the experimental data, which indicated that the reducing sialic acid unit (Sia-1) did not interact with the protein. Consequently, the Sia-1 residue was oriented away from the protein surface during the model building process. Conversely, both Sia-2 and Sia-3 sialic acid units had the potential to interact with Arg124 within the binding site. On the one hand, when Sia-3 engages with Arg124, the ligand extends predominantly toward the CC' loop; conversely, when Sia-2 interacts with Arg124, the saccharide chain extends toward both the CC' and BC loops of the protein. This interaction mode maintained potential contacts with the CC'

loop while additionally allowing for possible interactions with the BC loop. Furthermore, MD simulations were conducted using both the g and t conformations of the ligand, exploring various orientations within the binding site. The MD results demonstrated that, as expected, the ligand preferentially adopted the *t* conformation around the Sia-Gal linkage. Other torsion angles adopted values consistent with the exo-syn anomeric effect. Furthermore, the most stable interactions throughout the MD simulations involved the internal trisaccharide epitope (Figure 5A−E). Indeed, the carboxylate group of Sia-2 served as a key anchor in the interaction with Arg124, forming the most stable interaction observed throughout the MD simulation. In addition to Arg124, Asn133, and Lys131 stabilized the hydroxyl group at position 8 and the NAc group of Sia-2, respectively. Furthermore, the Lys131 carbonyl oxygen of the peptide bone acted as a hydrogen bond acceptor in the interaction with the acetyl group of AAT-2.

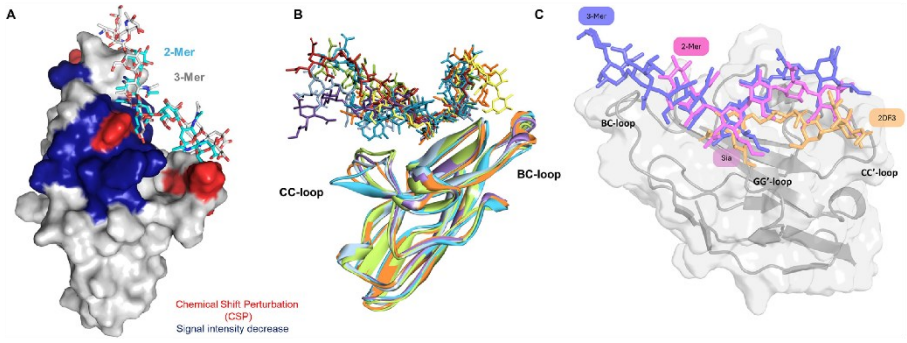


Figure IV.6 3D views of Siglec-7 in complex with 2-Mer and 3-Mer. (A) Superimposition of the complexes of Siglec-7 in interaction with 2-Mer (cyan) and 3-Mer (gray); the amino acids of Siglec-7 are colored according to protein-based NMR titration with 3-Mer. (B) Superimposition of the different poses of Siglec-7 in interaction with 3-Mer obtained by MD simulation. (C) Comparison of ligand binding conformations within the BC loop and CC'-loop regions. The DSLc4 ligand (orange) is shown alongside the 2-Mer (pink) and 3-Mer (purple) ligands. Differences in their positioning on the protein surface highlight the distinct interactions each ligand forms within the binding site, especially in relation to the BC-loop and CC'-loop.

Similarly, the carbonyl oxygen of the Asn129 peptide backbone served as a hydrogen bond acceptor in its interaction with the positively charged amino group of AAT-2 mirroring interactions observed in the Siglec-7−2-Mer complex. Moreover, in the Siglec-7/3-Mer complex, the ammonium ion at N4 of AAT-2 also interacted with Asp53, part of the BC loop that acted as a

hydrogen bond acceptor. Finally, analysis of root mean square fluctuations (RMSF) during the MD simulation indicated that the BC and CC' loops exhibited a high degree of flexibility even in the bound state. In contrast, the GG' loop demonstrated a decrease in flexibility upon ligand binding (Figure 5F). This could be explained by the close proximity of GG' loop residues to the binding pocket cavity, where the critical interactions with the internal trisaccharide epitope occur. Thus, while the internal repeating unit remained stable during the MD, establishing polar interactions with Arg124, Asn129, Lys131, and Asn133, the BC and CC' loop acted as wings allowing the formation of polar and hydrophobic interactions with the terminal and reducing trisaccharide, respectively (Figure 5).

4.3 Discussion

Fusobacterium nucleatum, an oral bacterium strongly implicated in periodontitis, is also highly prevalent in various gastrointestinal diseases, including colorectal cancer (CRC). Emerging evidence strongly indicates that *Fn* plays a significant role in tumorigenesis, facilitating tumor growth, promoting metastasis, and modulating host immune responses to favor tumor progression. Recent findings suggest that *F. nucleatum* localization within tumors is primarily mediated by glycan lectin interactions[25]. Specifically, the surface-exposed Fap2 lectin of *Fn* recognizes the Gal-GalNAc epitope (the Thomsen−Friedenreich antigen), which is abundantly overexpressed on the surface of colorectal and breast cancer cells[4]. In addition, we previously reported that Fn10953 interacts with Siglec-7 on immune cells through the LPS. In this study, we employed a multidisciplinary approach, encompassing fluorescence spectroscopy, NMR spectroscopy, molecular modeling, and biological assays, to comprehensively investigate the molecular mechanisms underlying the interaction between Fn10953 LPS with Siglec-7. We showed that Fn10953 LPS, along with its isolated O-antigen, specifically binds to Siglec-7 on RPMI8226 cells, a multiple myeloma model. Salmonella LPS, as a negative control, did not show binding, confirming specificity. Among blood

cells, Siglec-7 is highly expressed on NK cells, aligning with its known role in modulating NK cell functions. Exposure of PBMCs from healthy donors to Fn10953 LPS O-antigen significantly upregulated HLA-DR expression on NK cells. In contrast, T-lymphocytes and B lymphocytes did not exhibit significant activation. Purified NK cells exhibited significant activation responses to both Fn10953 LPS and its O-antigen, in contrast to T-lymphocytes and B lymphocytes, which showed minimal activation. The higher level of NK cell activation induced by Fn10953 LPS O-antigen suggests a potent immunostimulatory role mediated by specific Siglec-7 engagement. The ability of Fusobacterium nucleatum to engage Siglec-7 exemplifies a shared strategy among pathogens to evade immune surveillance through sialic acid-dependent interac tions. This mirrors the findings in reproductive infections and viral pathogenesis, where Siglecs are co-opted to modulate immune responses and promote pathogen survival. 11 16 The 'double-edged sword' nature of Siglec' interactions are particularly evident with Fusobacterium nucleatum, as its immune modulation via Siglec-7 engagement parallels strategies observed in viruses such as HIV and SARS-CoV 2. This conserved mechanism of immune evasion highlights the potential of Fn to profoundly impact the tumor microenvironment by modulating immune responses. At the molecular level, our data demonstrated that the core region of the LPS did not contribute to Siglec-7 binding. Moreover, equilibrium affinity constants for the entire O antigen and for oligosaccharides containing varying numbers of repeating units were comparable, all ranging within the micromolar range (Figures 3C and 6). Furthermore, the single trisaccharide repeating unit (1-Mer) was found to be insufficient for interaction with Siglec-7. This observation suggests that a longer glycan chain is necessary to induce recognition by Siglec-7. We identified the internal sialic acid N' and the AAT residue A' as key contributors to the binding process, while the galactose unit did not significantly interact with Siglec-7. Notably, the 2-Mer, which preferentially adopted an extended conformation in the free state (Figures 3B and 4), maintained this conformation upon binding to Siglec-7. Extensive MD simulations confirmed the ligand's

preference for the t conformation around the Sia glycosidic linkage (φ = 180°). The internal Sia residue N' was accommodated in the binding pocket establishing the crucial salt bridge between its carboxylate group and the key Arg124 residue, as well as hydrogen bonds with its NHAc and the glycerol chain and Lys131, Asn133, and Lys135 residues. Additionally, significant interactions were observed between the positively charged amino group at N4 of AAT A' and the carbonyl group of the Asn129 backbone. The above lead to the overall stabilization of the protein GG' and CC' loops upon binding. The ligand containing three repeating units (3-Mer) demonstrated a preferential interaction with Siglec-7 primarily through Sia-2, favoring its engagement with Arg124 within the Siglec-7 binding site compared to the other two Sias residues. Both Sia-2 and AAT-2 emerged as the key determinants of ligand recognition, and induced a chemical shift perturbation and an intensity decrease in the neighboring amino acids, as evidenced by protein-based NMR experiments (Figures 3B and 6) and corroborated by their stability in the RMSD analysis (Figure 5B). Thus, signals of Arg124 and the contiguous/adjacent amino acids, comprising Lys131, Asn133, and Lys135, were affected by the intensity decrease and were also identified as the main residues in contact with Sia-2 and AAT-2 of the 3-Mer oligosaccharide. Notably, Lys135, together with Tyr134, Tyr136, Asp137, and Glu138 belong to the GG' loop, which is slightly stabilized in the bound state, as monitored by the RMSF (Figure 5F). Notably, protein-based NMR experiments performed on the full Fn10953 O-antigen (OPS) identified the key interaction sites of Siglec-7 that corroborated findings with the 3-Mer. Computational analyses suggested a dynamic binding interface, with the BC and CC' loops exhibiting a "wing-like" movement, potentially interacting with different regions of the O-antigen. However, the BC loop appeared to form more stable interactions with the ligand compared to the CC' loop, which exhibited more transient contacts. Indeed, from protein-based NMR results, in the BC loop, Tyr50 and Val52 were affected by an intensity decrease and Asp57 was subjected to a CSP; instead, the CC' loop was only affected by the CSP effects detected for Ile72 and Lys76, thus involved in a

faster exchange with the ligand. Therefore, the specific conformational features of the Fn10953 LPS O-antigen likely play a crucial role in Siglec-7 binding. Through presentation of an appropriate epitope within the Sia-AAT unit, the O-antigen promotes optimal accommodation within the Siglec-7 binding site. This binding mode was compared to the accommodation of the disialylated ganglioside DSLc4 in the binding site (crystal structure 2DF3), where the terminal sialic acid linked to GlcNAc interacts extensively with the GG' loop, whereas the terminal Sia linked to Gal interacts with the CC' loop. Conversely, Fn10953 LPS O-antigen accommodation inside the binding site is driven by internal sialic acid units that preferentially interact with the Siglec-7 BC loop (Figure 6C). We recently demonstrated that Neisseria meningitidis serogroup Y capsular polysaccharide targets the inhibitory receptor Siglec-7 through a binding mode crucially dependent on the specific arrangement of its sialic acids and the 3D structure, thereby revealing a key mechanism where bacterial glycans exploit host inhibitory pathways to facilitate immune evasion. 26 Here, our comprehensive investigation further provides crucial insights for understanding disease mechanisms and host−pathogen interactions. We clearly demonstrate a novel binding mode for the interaction between Fn10953 LPS and Siglec-7. This unique recognition mechanism is fundamentally driven by the intricate three-dimensional arrangement of the Fn10953 O-antigen. Specifically, this mode highlights the pivotal role of internal sialic acid residues, integral component of the O-antigen repeating unit and responsible for mediating stable complexes through key electrostatic and polar interactions. This discovery thus delineates a new paradigm for glycan-mediated bacterial exploitation of host inhibitory receptors. Finally, we provide critical insights into how *Fusobacterium nucleatum* LPS could modulate immunity via Siglec-7, contributing to immune suppression, evasion strategies, and altered immune cell recognition. Beyond this understanding, this study implications are profound, strongly suggesting the therapeutic potential of targeting Siglec-7 for both CRC and other related conditions.

4.4 Reference

(1) Brennan, C. A.; Garrett, W. S. Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. Nature Reviews Microbiology 2019, 17 (3), 156-166. DOI: 10.1038/s41579-018-0129-6.

(2) Alon-Maimon, T.; Mandelboim, O.; Bachrach, G. Fusobacterium nucleatum and cancer. Periodontology 2000 2022, 89 (1), 166-180. DOI: doi.org/10.1111/prd.12426 (acccessed 2024/05/06).

(3) Han, Y. W. Fusobacterium nucleatum: a commensal-turned pathogen. Current opinion in microbiology 2015, 23, 141-147. DOI: 10.1016/j.mib.2014.11.013 From NLM.

(4) Parhi, L.; Alon-Maimon, T.; Sol, A.; Nejman, D.; Shhadeh, A.; Fainsod-Levi, T.; Yajuk, O.; Isaacson, B.; Abed, J.; Maalouf, N.; et al. Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. Nature Communications 2020, 11 (1), 3259. DOI: 10.1038/s41467-020-16967-2.

(5) Wu, S.; Rhee, K.-J.; Albesiano, E.; Rabizadeh, S.; Wu, X.; Yen, H.-R.; Huso, D. L.; Brancati, F. L.; Wick, E.; McAllister, F. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nature medicine 2009, 15 (9), 1016-1022. Marchesi, J. R.; Dutilh, B. E.; Hall, N.; Peters, W. H. M.; Roelofs, R.; Boleij, A.; Tjalsma, H. Towards the Human Colorectal Cancer Microbiome. PLOS ONE 2011, 6 (5), e20447. DOI: 10.1371/journal.pone.0020447.

(6) Sepich-Poore, G. D.; Zitvogel, L.; Straussman, R.; Hasty, J.; Wargo, J. A.; Knight, R. The microbiome and human cancer. Science 2021, 371 (6536). DOI: 10.1126/science.abc4552 From NLM. Wang, S.; Liu, Y.; Li, J.; Zhao, L.; Yan, W.; Lin, B.; Guo, X.; Wei, Y. Fusobacterium nucleatum Acts as a Pro-carcinogenic Bacterium in Colorectal Cancer: From Association to Causality. Front. Cell Dev. Biol. 2021, 9. DOI: 10.3389/fcell.2021.710165. Zepeda-Rivera, M.; Minot, S. S.; Bouzek, H.; Wu, H.; Blanco-Míguez, A.; Manghi, P.; Jones, D. S.; LaCourse, K. D.; Wu, Y.; McMahon, E. F.; et al. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. Nature 2024, 628 (8007), 424-432. DOI: 10.1038/s41586-024-07182-w.

(7) Casasanta, M. A.; Yoo, C. C.; Udayasuryan, B.; Sanders, B. E.; Umaña, A.; Zhang, Y.; Peng, H.; Duncan, A. J.; Wang, Y.; Li, L.; et al. Fusobacterium nucleatum host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives colorectal cancer cell migration. Science signaling 2020, 13 (641). DOI: 10.1126/scisignal.aba9157 From NLM.

(8) Wong, S. H.; Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol 2019, 16 (11), 690-704. DOI: 10.1038/s41575-019-0209-8 From NLM. Wu, J.; Li, Q.; Fu, X. Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. Translational oncology 2019, 12 (6), 846-851. DOI: 10.1016/j.tranon.2019.03.003 From NLM.

(9) Stanczak, M. A.; Läubli, H. Siglec receptors as new immune checkpoints in cancer. Molecular Aspects of Medicine 2023, 90, 101112. DOI: https://doi.org/10.1016/j.mam.2022.101112.

(10) Angata, T.; Varki, A. SiglecSiglecsInteractions with PathogensPathogens. In Glycoscience: Biology and Medicine, Endo, T., Seeberger, P. H., Hart, G. W., Wong, C.-H., Taniguchi, N. Eds.; Springer Japan, 2021; pp 1-9. Kim, C.-H. Glycobiology of innate immunology; Springer, 2022. Hsieh, S.-L. Lectin in host defense against microbial infections; Springer, 2020. In Essentials of Glycobiology, Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., et al. Eds.; Cold Spring Harbor Laboratory Press, 2022.

(11) Chen, Y.; Chen, H.; Zheng, Q. Siglecs family used by pathogens for immune escape may engaged in immune tolerance in pregnancy. Journal of Reproductive Immunology 2023, 159, 104127. DOI: https://doi.org/10.1016/j.jri.2023.104127. Chang, Y.-C.; Nizet, V. The interplay between Siglecs and sialylated pathogens. Glycobiology 2014, 24 (9), 818-825. DOI: 10.1 093/glycob/cwu067 (acccessed 1/23/2025).

(12) Pillai, S.; Netravali, I. A.; Cariappa, A.; Mattoo, H. Siglecs and Immune Regulation. Annual review of immunology 2012, 30 (1), 357-392. DOI: 10.1146/annurev-immunol-020711-075018.

(13) Varki, A.; Angata, T. Siglecs-the major subfamily of I-type lectins. Glycobiology 2006, 16 (1), 1r-27r. DOI: 10.1093/glycob/cwj008 From NLM.

(14) Fong, J. J.; Tsai, C.-M.; Saha, S.; Nizet, V.; Varki, A.; Bui, J. D. Siglec-7 engagement by GBS β-protein suppresses pyroptotic cell death of natural killer cells. PNAS 2018, 115 (41), 10410-10415. DOI: doi:10.1073/pnas.1804108115.

(15) Heikema, A. P.; Jacobs, B. C.; Horst-Kreft, D.; Huizinga, R.; Kuijf, M. L.; Endtz, H. P.; Samsom, J. N.; van Wamel, W. J. Siglec-7 specifically recognizes Campylobacter jejuni strains associated with oculomotor weakness in Guillain–Barré syndrome and

Miller Fisher syndrome. Clinical Microbiology Infection 2013, 19 (2), E106-E112. Avril, T.; Wagner, E. R.; Willison, H. J.; Crocker, P. R. Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on Campylobacter jejuni lipooligosaccharides. Infect Immun 2006, 74 (7), 4133-4141. DOI: 10.1128/iai.02094-05 From NLM.

(16) Raïch-Regué, D.; Resa-Infante, P.; Gallemí, M.; Laguia, F.; Muñiz-Trabudua, X.; Muñoz-Basagoiti, J.; Perez-Zsolt, D.; Chojnacki, J.; Benet, S.; Clotet, B.; et al. Role of Siglecs in viral infections: A double-edged sword interaction. Molecular Aspects of Medicine 2023, 90, 101113. DOI: https://doi.org/10.1016/j.mam.2022.101113.

(17) 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. Frontiers in Immunology 2021, 12, DOI: 10.3389/fimmu.2021.744184.

(18) Vinogradov, E.; St. Michael, F.; Homma, K.; Sharma, A.; Cox, A. D. Structure of the LPS O-chain from Fusobacterium nucleatum strain 10953, containing sialic acid. Carbohydr. Res. 2017, 440-441, 38-42. DOI: https://doi.org/10.1016/j.carres.2017.01.009.

(19) Abreu, C.; Di Carluccio, C.; Ječmen, T.; Skořepa, O.; Bláha, J.; Marchetti, R.; Silipo, A.; Vaněk, O. Insights into stability, dimerisation, and ligand binding properties of Siglec-7: Isotope labelling in HEK293 cells for protein characterisation by NMR spectroscopy. International Journal of Biological Macromolecules 2025, 309, 142672. DOI: https://doi.org/10.1016/j.ijbiomac.2025.142672.

(20) Di Carluccio, C.; Forgione, R. E.; 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 Chemical Biology 2021, 2 (6), 1618-1630, 10.1039/D1CB00173F. DOI: 10.1039/D1CB00173F. Di Carluccio, C.; Forgione, R. E.; Montefiori, M.; Civera, M.; Sattin, S.; Smaldone, G.; Fukase, K.; Manabe, Y.; Crocker, P. R.; Molinaro, A.; et al. Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22. iScience 2021, 24 (1), 101998. DOI: https://doi.org/10.1016/j.isci.2020.101998.

(21) Di Carluccio, C.; Forgione, M. C.; Martini, S.; Berti, F.; Molinaro, A.; Marchetti, R.; Silipo, A. Investigation of protein-ligand complexes by ligand-based NMR

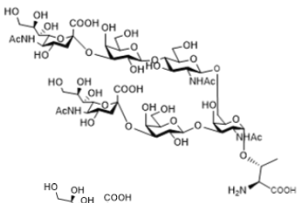
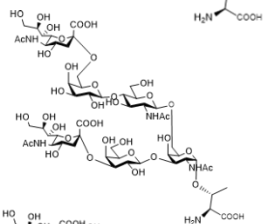
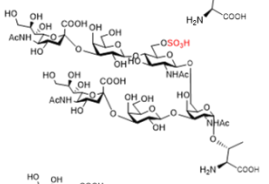
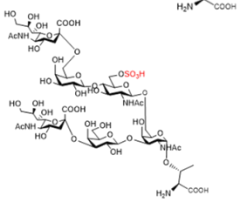
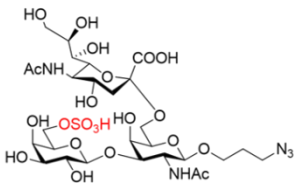
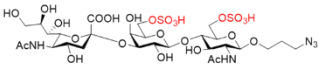
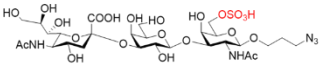
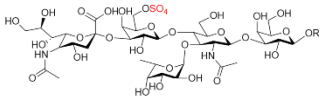
methods. Carbohydr Res 2021, 503, 108313. DOI: 10.1016/j.carres.2021.108313 From NLM.
(22) AMBER 2018; University of California, San Francisco, 2018. (accessed.
(23) Attrill, H.; Imamura, A.; Sharma, R. S.; Kiso, M.; Crocker, P. R.; van Aalten, D. M. F. Siglec-7 Undergoes a Major Conformational Change When Complexed with the α(2,8)-Disialylganglioside GT1b. Journal of Biological Chemistry 2006, 281 (43), 32774-32783. DOI: https://doi.org/10.1074/jbc.M601714200.
(24) Nieto-Fabregat, F.; Lenza, M. P.; Marseglia, A.; Di Carluccio, C.; Molinaro, A.; Silipo, A.; Marchetti, R. Computational toolbox for the analysis of protein–glycan interactions. Beilstein Journal of Organic Chemistry 2024, 20, 2084-2107. DOI: 10.3762/bjoc.20.180.
(25) Engevik, M. A.; Danhof, H. A.; Ruan, W.; Engevik, A. C.; Chang-Graham, A. L.; Engevik, K. A.; Shi, Z.; Zhao, Y.; Brand, C. K.; Krystofiak, E. S.; et al. Fusobacterium nucleatum Secretes Outer Membrane Vesicles and Promotes Intestinal Inflammation. mBio 2021, 12 (2). DOI: 10.1128/mBio.02706-20 From NLM.
(26) Di Carluccio, C.; Gerpe Amor, T.; Lenza, M. P.; Masi, A. A.; Abreu, C.; Longo, V.; Albano, F.; Nieto-Fabregat, F.; Salvatore, P.; Falco, G.; et al. Molecular Basis of Siglec-7 Recognition by Neisseria meningitidis Serogroup Y CPS: Implications for Immune Evasion. JACS Au 2025, 5(5), 2257-2269 DOI: 10.1021/jacsau.5c00214.
(27) Di Carluccio, C.; Padilla-Cortés, L.; Tiemblo-Martìn, M.; Gheorghita, G. R.; Oliva, R.; Cerofolini, L.; Masi, A. A.; Abreu, C.; Tseng, H.-K.; Molinaro, A.; et al. Insights into Siglec-7 Binding to Gangliosides: NMR Protein Assignment and the Impact of Ligand Flexibility. Advanced Science 2025, 12(21), e2415782.DOI: https://doi.org/10.1002/advs.202415782 (acccessed 2025/04/28).
(28) Di Carluccio, C.; Milanesi, F.; Civera, M.; Abreu, C.; Sattin, S.; Francesconi, O.; Molinaro, A.; Vaněk, O.; Marchetti, R.; Silipo, A. Tumor Carbohydrate Associated Antigen Analogs as Potential Binders for Siglec-7. European Journal of Organic Chemistry 2023, 26 (40), e202300644. DOI: https://doi.org/10.1002/ejoc.202300644 (acccessed 2024/09/10).
(29) Keller, R. The Computer Aided Resonance Assignment Tutorial. In CANTINA verlag, 2004.
(30) Williamson, M. P. Chemical Shift Perturbation. In Modern Magnetic Resonance, Webb, G. A. Ed.; Springer International Publishing, 2018; pp 995-1012.
(31) Frisch, M. J. gaussian09. http://www.gaussian.com/ 2009.

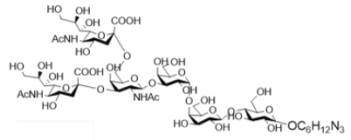
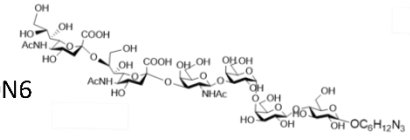
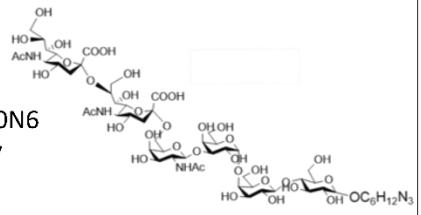
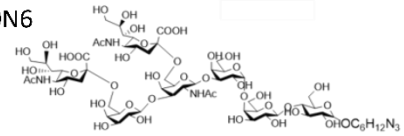
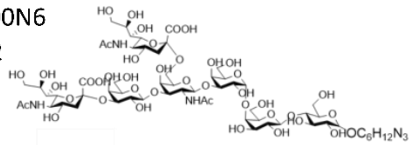
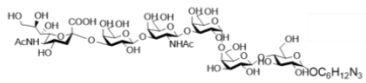
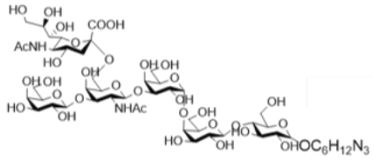
(32) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. J. Am. Chem. Soc 2001, 222 (1).
(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. J. Molec. Graphics 1996, 14, 33-38.
(34) Schrödinger Release 2022-3; Maestro, Schrödinger, LLC: New York, 2021.
(35) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of chemical theory and computation 2015, 11 (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255.
(36) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 2008, 29 (4), 622-655. DOI: 10.1002/jcc.20820 From NLM.
(37) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. 2013, 9 (7), 3084-3095.
(38) The PyMOL Molecular Graphics System; Version 2.3.4 Schrödinger,LLC,
(39) Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; et al. Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology 2019, 29 (9), 620-624. DOI: 10.1093/glycob/cwz045 (acccessed 9/10/2024).
(40) St Michael, F.; Fleming, P.; Cox, A. D.; Vinogradov, E. Structural analysis of the core oligosaccharides from Fusobacterium nucleatum lipopolysaccharides. Carbohydrate Research 2021, 499, 108198. DOI: https://doi.org/10.1016/j.carres.2020.108198.

V. Chapter V: Functional screening of Siglec-7 glycan ligands

5.1 Introduction

To deepen our understanding of the molecular determinants that regulate Siglec-7 glycan recognition, functional screening was performed with the aim of determining with high accuracy the kinetic constants and structural observations underlying the binding phenomenon. Although structural analyses and molecular modeling provide a static picture of the binding interface, it can often be difficult to grasp the energetics underlying molecular recognition. In this context, an integrative approach was adopted that combines quantitative biophysical measurements to interrogate the interaction between Siglec-7 and a diverse panel of glycans from different sources. The screening was designed to assess how subtle changes in glycan architecture, such as differences in sialylation linkage, sulfation pattern, or the presence of extended carbohydrate chains, influence both affinity and binding energy. This strategy reflects the broader goal of translating structural diversity into measurable energy parameters, thereby bridging the gap between molecular structure and biological function. The library of ligands analyzed in this work includes both endogenous and exogenous glycoconjugates (Figure1). These included tumor-associated carbohydrate antigens such as 6′-Sulfo-Sialyl Lewis^x (6′S-sLe^x), various ganglioside derivatives including DSGb5, DSGb4, as well as Core O-glycans in sulfated and non-sulfated forms. In addition, bacterial polysaccharides and lipooligosaccharides (CPS, LPS, and LOS) extracted from different pathogenic strains were included to study how Siglec-7 recognizes sialylated motifs of microbial origin. This comprehensive selection allowed for the comparison of ligands that differ not only in their chemical nature but also in their biological relevance, encompassing both endogenous and exogenous glycans.

Ligand ID	Glycan type	Exact Mass (Da)	Molecular Formula	Molecular structure
1	Core2-O α2,3	1431.51	C56H91N5 O40	
2	Core2-O α2,6	1431.51	C56H91N5 O40	
3	Core2-O 6'S α2,3	1511.47	C56H91N5 O44S	
4	Core2-O 6'S α2,6	1511.47	C56H91N5 O44S	
5	6′-sulfo-α2,6 Sialyl T-Antigen (6′Sα6-sTA)	837.24	C28H47N5 O22S	
6	di-6'-sulfo-3'Sln	917.20	C28H47N5 O25S2	
7	6′-sulfo-Sialyl T-Antigen (6′S-sTA)	837.24	C28H47N5 O22S	
8	6'Sulfo SLe^x	1166.40	C34H58N3 O26S	

9	α3α6DSGb4	1414.53	C54H90N6O37	
10	α3α8DSGb4	1414.53	C54H90N6O37	
11	α6α8DSGb4	1414.53	C54H90N6O37	
12	α6α6*DSGb5	1576.580	C60H100N6O42	
13	α3α6DSGb5	1576.55	C60H100N6O42	
14	α3MSGb5 (SSEA-4)	1285.49	C49H83N5O34	
15	α6MSGb5	1285.49	C49H83N5O34	

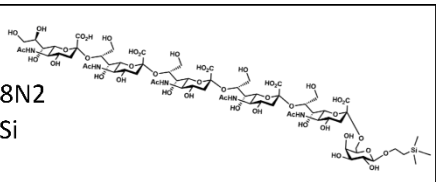
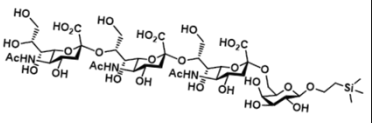
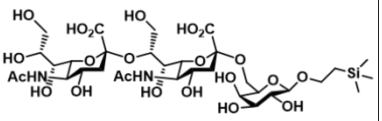
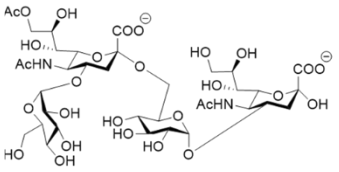
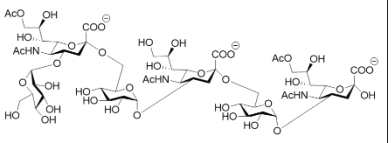
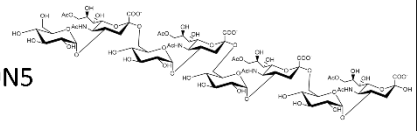
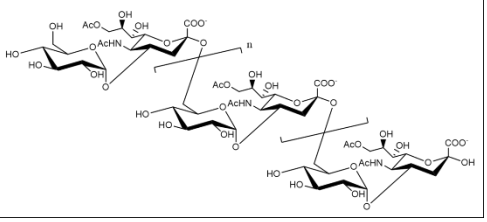
16	Polisialic Acid α2,8 (2RU)	862.33	C33H58N2O22Si	
17	Polisialic Acid α2,8 (3RU)	1153.42	C44H75N3O30Si	
18	Polisialic Acid α2,8 (5RU)	1735.61	C66H109N5O46Si	
19	Polisialic Acid α2,6 (2RU)	512.33	C33H58N2O22	
20	Polisialic Acid α2,6 (3RU)	1006.30	C44H75N3O30	
21	Polisialic Acid α2,6 (4RU)	1995.71	C66H109N5O46	
22	Polisialic Acid α2,6 (CPS)			

Figure V.1Overview of the glycans included in the Siglec-7 screening panel, reporting their structural class, exact mass, molecular formula, and representative molecular structures used for computational and experimental analyses

From an experimental standpoint, this screening combined steady-state fluorescence titration and native mass spectrometry (nMS) in a complementary workflow. The fluorescence experiments provided quantitative information on equilibrium binding constants (K_b), which were determined at different temperatures and further analyzed using the van't Hoff equation to derive the corresponding thermodynamic parameters ΔH, ΔS, and ΔG. The K_b values obtained were confirmed using nMS to obtain highly accurate values. The information obtained from this workflow reveals the relative contributions of enthalpy and entropy to the overall free energy of association, as well as binding stoichiometry and association constants, providing a comprehensive energetic interpretation of the binding event. In parallel, both techniques enabled the direct, label-free detection of intact protein-ligand complexes under physiological conditions. Through the precise measurement of mass shifts, this technique provided detailed information on the binding stoichiometry and independent estimates of the equilibrium constants. The combination of fluorescence spectroscopy and nMS thus offered a multidimensional perspective on Siglec–glycan interactions. While nMS defined the composition and stoichiometric ratios of the complexes formed, fluorescence titration revealed the underlying thermodynamic balance that governs their formation and stability. Simultaneous access to both equilibrium constants and thermodynamic parameters allowed a complete and internally consistent interpretation of the binding process, ensuring that the results from each technique could validate and reinforce the other. Together, these complementary datasets not only clarified the energetic basis of Siglec-7 recognition but also offered a kinetic and mechanistic framework capable of describing how structural variations in glycan ligands modulate the overall affinity landscape.

5.2 Results and discussion

5.2.1 Core O-GalNAc glycans

The 6′-sulfo-Sialyl T-Antigen (6′S-sT-Antigen) represents a terminally modified mucin-type O-glycan that plays a crucial role in cell adhesion and immune recognition. Structurally, it derives from the classical Core 1 T-Antigen (Galβ1-3GalNAcα1-Ser/Thr) through α2,6-sialylation of the terminal galactose and additional sulfation at the 6′-position. These terminal modifications introduce both conformational and electrostatic variability, generating a highly anionic recognition surface particularly relevant for lectin-mediated interactions. The biological relevance of this motif is especially pronounced in epithelial tissues and tumor microenvironments, where aberrant expression of sulfated sialyl-T antigens has been correlated with altered immune recognition and metastatic potential. To gain a broader view of how subtle structural variations influence Siglec-7 binding, the screening panel was extended to include an additional analogue of the 6-sulfo-Sialyl T-Antigen bearing an α2,6 linkage on the GalNAc residue and a sulfate group positioned on the GlcNAc moiety, as well as the 6′-Sulfo-Sialyl Lewis^x (6′S-sLe^x) — a tetrasaccharide commonly expressed on leukocytes and endothelial cells that represents a key ligand for selectins and other sialic-acid-binding lectins. Furthermore, the dataset incorporated 3′-Sialyl lactosamine (3′-SLN), a representative linear disaccharide sharing the same terminal sialylated motif but lacking the mucin-type backbone. These related structures provide an ideal comparative framework for dissecting the influence of sulfation site, sialyl linkage, and core architecture on Siglec-7 recognition. Native mass spectrometry spectra recorded for the interaction between Siglec-7, and the Core O-glycan derivatives revealed well-resolved signals corresponding to both the free protein and the 1:1 protein–ligand complex. Across all four ligands, the non-covalent interactions were preserved under the soft ionization conditions employed. The presence of discrete mass peaks at *m/z* =2003.6560; 2013.6491; 2003.7802 and 2044.8177 (ligand 5, ligand 6, ligand 7 and ligand 8

respectively) indicates a specific and stoichiometric association rather than nonspecific clustering (Figure 2A).

Among the ligands tested, clear differences in bound-state intensity were observed. The **ligand 8** exhibited the most pronounced complex formation, characterized by a strong and well-defined (1:1) signal envelope, indicative of the highest binding affinity and enhanced stability of the Siglec-7/ligand complex. In contrast, **ligand 7** displayed a markedly weaker bound-state signal under identical conditions, reflecting lower affinity and less favorable complex stabilization. **Ligand 5** and **ligand 6** showed an intermediate response, with

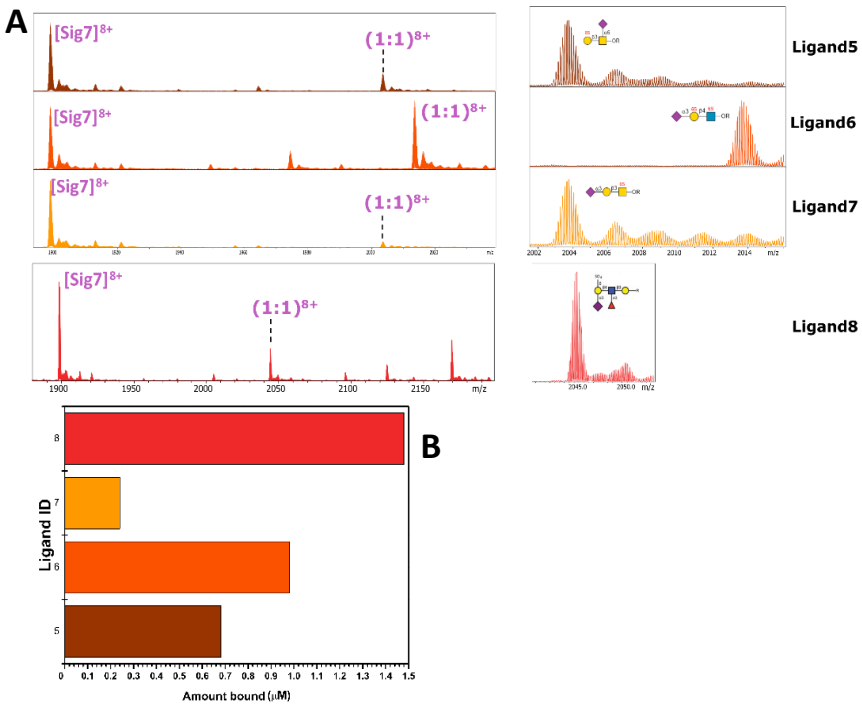


Figure V.2(A) Native mass spectrometry analysis of Siglec-7 interactions with Core O-glycan derivatives. Spectra show the free protein and the corresponding 1:1 Siglec-7/ligand complexes under non-denaturing conditions. All spectra confirm a 1:1 stoichiometry consistent with a single carbohydrate-recognition domain. (B) Histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7d1 for these sialoglycans.

detectable levels of complex formation. Quantitative integration of the peak intensities confirmed that all complexes form with a 1:1 stoichiometry, consistent with the single carbohydrate recognition domain of Siglec-7. The relative binding ranking trend—**ligand 8 >>> ligand 6 >> ligand 5 > ligand**

7—correlates with the expected hierarchy of affinity derived from their chemical composition and substitution pattern (Figure 2B).

Table V.1 Summary of the equilibrium binding constants (K_b and K_D) determined for ligand 5,6,7 and 8 tested using native mass spectrometry (nMS).

	nMS	
Ligand	**K_b /[M^{-1}]**	**K_D [M]**
5	$1.15 \cdot 10^4$	$6.63 \cdot 10^{-5}$
6	$5.97 \cdot 10^4$	$1.67 \cdot 10^{-5}$
7	$5.51 \cdot 10^3$	$1.82 \cdot 10^{-4}$
8	$4.36 \cdot 10^4$	$2.29 \cdot 10^{-5}$

The Core-2 O-glycan family represents one of the most widespread and structurally versatile branches of mucin-type glycosylation in higher eukaryotes. In the context of Siglec recognition, Core-2 O-glycans offer an ideal model to explore how branching, linkage geometry, and the presence of

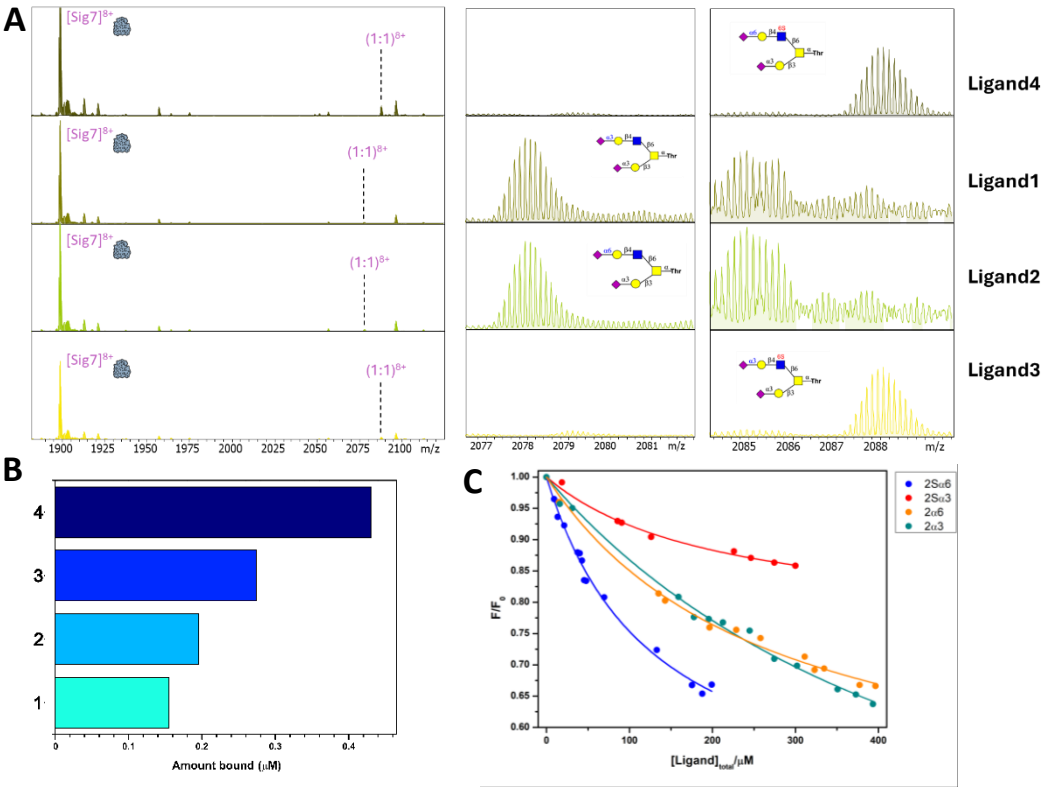


***Figure V.3** (A)*Electrospray mass spectra of protein Siglec-7d1 with ligand 1 (Core 2 α2,3) ligand 2 (Core 2 α2,6), ligand 3 (Core 2 6'S α2,3) and ligand4 (Core 2 6'S α2,6). (B) A histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7d1 for these sialoglycans. (C) Binding isotherms measured by fluorescence spectroscopy following the formation of the complex between Siglec-7 and lignad1(green circle), lignad2 (orange circle), lignad3 (red circle) and ligand4 (blue circle); The solid lines represent the best fit of the experimental data according to a 1:1 binding model.

charged substituents influence binding specificity and energetic stability. This dataset compares four variants of Core-2 O-glycans: two sulfated analogues and their non-sulfated counterparts, each presented with distinct terminal sialic acid linkages (**ligand 1**: Disialyl Core2 α2,3α2,3; **ligand 2**: Disialyl Core2 α2,6α2,3; **ligand 3**: Disialyl Core2 6-Sulfo α2,3α2,3; **ligand 4**: Disialyl Core2 6-Sulfo α2,6α2,3). Native ESI conditions were intentionally "*soft*" (source energy set to 0eV), with low micromolar Siglec-7 and each ligand in excess in single-point mixtures; the bound state region is displayed at 2077.95 *m/z* (for ligand 1 and ligand 2), and 2088.07 *m/z* (for ligand 3 and ligand 4), where the 8+ charge state of the monomeric ectodomain is observed (Figure3A). In this spectral window, all four ligands generate a clear mass peak consistent with the formation of 1:1 complexes, with no evidence of higher stoichiometries under these conditions. This behavior corresponds to the lectin architecture of Siglec-7 and indicates that, at least in the tested regime, binding remains monovalent and specific. Steady-state fluorescence titrations conducted in parallel provided complementary information on the binding equilibria. Progressive addition of Disialyl Core-2 O-glycan to the protein solution resulted in a reproducible quenching of the intrinsic tryptophan emission signal, consistent with complex formation. Curve fitting using a 1:1 binding model (Figure 2C) yielded equilibrium constants (K_b) in good agreement with those determined by nMS. The combined results suggest that sulfation enhances the strength of the Siglec-7–ligand interaction, whereas glycosidic linkage selectivity favors the α2,6-sialylated form over the α2,3 analogue.

Table V.2 Equilibrium binding constants of Siglec-7 with Core-2 O-glycan derivatives obtained from native mass spectrometry (nMS) and steady-state fluorescence titration. Both K_b(association constant) and K_D(dissociation constant) are reported. The values determined by nMS are related to 8+ charge states, whereas fluorescence-based constants derive from non-linear fitting curves using a single-site binding model.

	nMS		Fluorescence	
Ligand	**K_b /[M^{-1}]**	**K_D [M]**	**K_b /[M^{-1}]**	**K_D [M]**
1	$3.2 \cdot 10^{3}$	$3.1 \cdot 10^{-4}$	$1.85 \cdot 10^{3}$	$5.41 \cdot 10^{-4}$
2	$4.2 \cdot 10^{3}$	$2.4 \cdot 10^{-4}$	$3.76 \cdot 10^{3}$	$2.66 \cdot 10^{-4}$
3	$4.8 \cdot 10^{3}$	$2.1 \cdot 10^{-4}$	$4.87 \cdot 10^{3}$	$2.05 \cdot 10^{-4}$
4	$8.2 \cdot 10^{3}$	$1.2 \cdot 10^{-4}$	$8.12 \cdot 10^{3}$	$1.23 \cdot 10^{-4}$

5.2.2 Gangliosides Globo-series

Given their structural diversity and strong anionic character, gangliosides represent key endogenous ligands for Siglecs, mediating finely tuned interactions that regulate immune homeostasis and cell communication. The inclusion of DSGb4 and DSGb5 derivatives in the functional screening aimed to dissect how variations in sialylation topology—specifically α2,3-, α2,6-, and α2,8-linkages—affect Siglec-7 recognition, binding stoichiometry, and overall affinity.

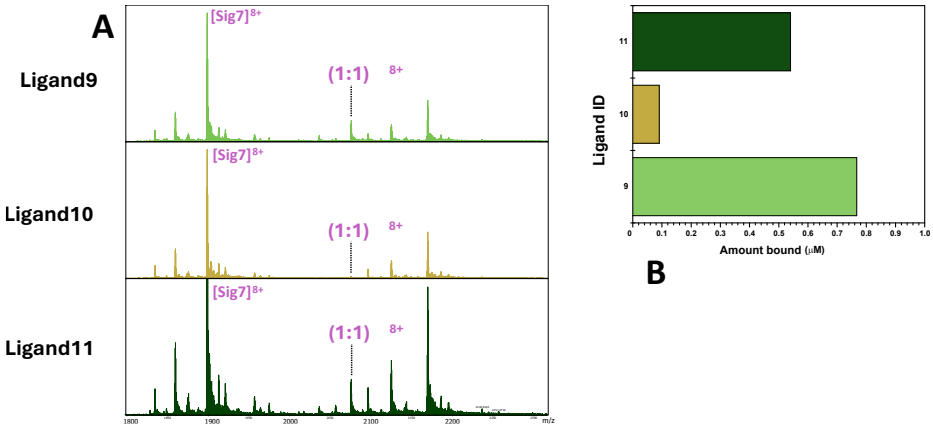


Figure V.4 *(A)*Native mass spectra of protein Siglec-7d1 with ligand 9 (α3α6-DSGb4) ligand 10 (α3α8-DSGb4) and ligand 11 (α6α8-DSGb4). (B) A histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7d1 for these sialoglycans.

Native mass spectrometry experiments performed under soft ionization conditions clearly revealed the formation of 1:1 Siglec-7/DSGb4 complexes for all three derivatives analyzed — α3α6-DSGb4 (**ligand 9**), α3α8-DSGb4 (**ligand 10**), and α6α8-DSGb4 (**ligand 11**). The spectra displayed distinct charge-state, corresponding to the free protein and to the bound complex, confirming the stability of the interaction in the gas phase (Figure 4A/B).

Although the overall stoichiometry remained identical across all variants, clear differences were observed in the relative intensity of the complex peaks. The ligand9 showed the strongest bound-state signal, indicating the highest affinity among the tested derivatives. **Ligand 11** produced an intermediate response, while **ligand 10** exhibited the weakest binding under the same experimental conditions. This affinity ranking — ligand9 > ligand11>> ligand10 — suggests that the α2,6 sialyl linkage plays a decisive role in stabilizing the Siglec-7/ganglioside interaction, while α2,8 linkages alone provide suboptimal geometry or reduced electrostatic complementarity within the binding pocket.

Table V.3 Summary of the equilibrium binding constants (K_b and K_D) determined for ligand 9,10 and 11 tested using native mass spectrometry (nMS).

	nMS	
Ligand	**K_b /[M^{-1}]**	**K_D [M]**
9	$1.96 \cdot 10^{4}$	$5.09 \cdot 10^{-5}$
10	$1.90 \cdot 10^{3}$	$5.27 \cdot 10^{-4}$
11	$1.28 \cdot 10^{4}$	$7.81 \cdot 10^{-5}$

To elucidate how the sialylation pattern and linkage topology affect Siglec-7 recognition, a comparative analysis was carried out on two DSGb5 isomers—α6α6DSGb5 (**ligand 12**) and α3α6-DSGb5 (**ligand 13**) forms—together with their monosialylated derivatives, α2,3 MSGb5, also named SSEA-4 (**ligand 14**) and α2,6 MSGb5 (**ligand 15**). These gangliosides represent physiologically relevant ligands that differ only in the position and number of sialic acid residues, thus providing an ideal framework to dissect the specific energetic contributions of α2,3 and α2,6 linkages to Siglec-7 binding.

Native electrospray ionization mass spectrometry (nMS) was employed to evaluate the binding of Siglec-7 to DSGb5-derived glycans under non-denaturing conditions. Analysis of the charge state distribution shown that the Siglec-7 retained a non-degraded and folded structure in the gas phase, enabling

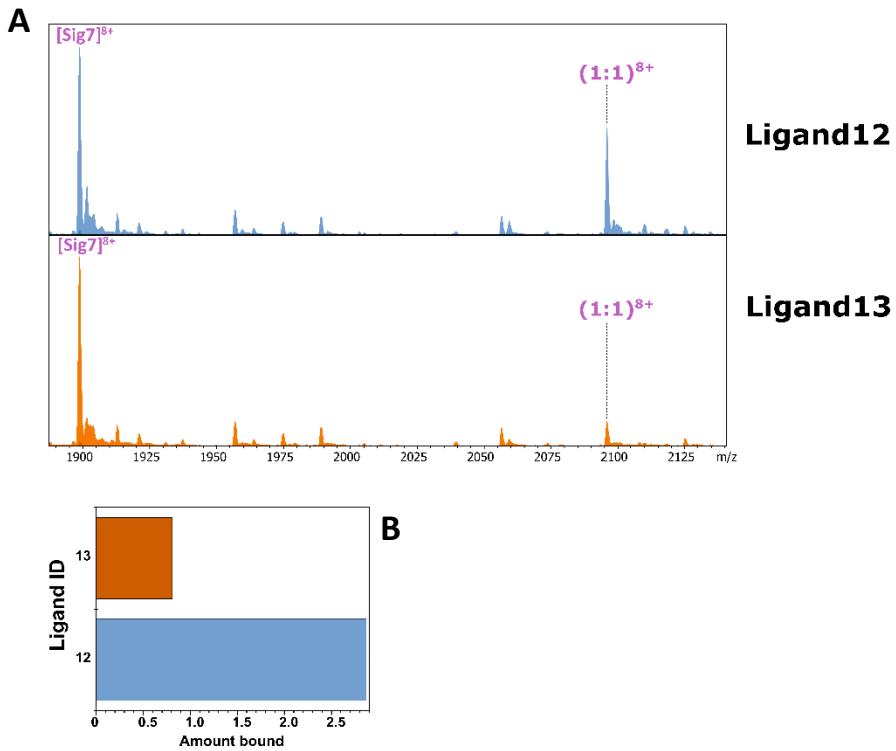
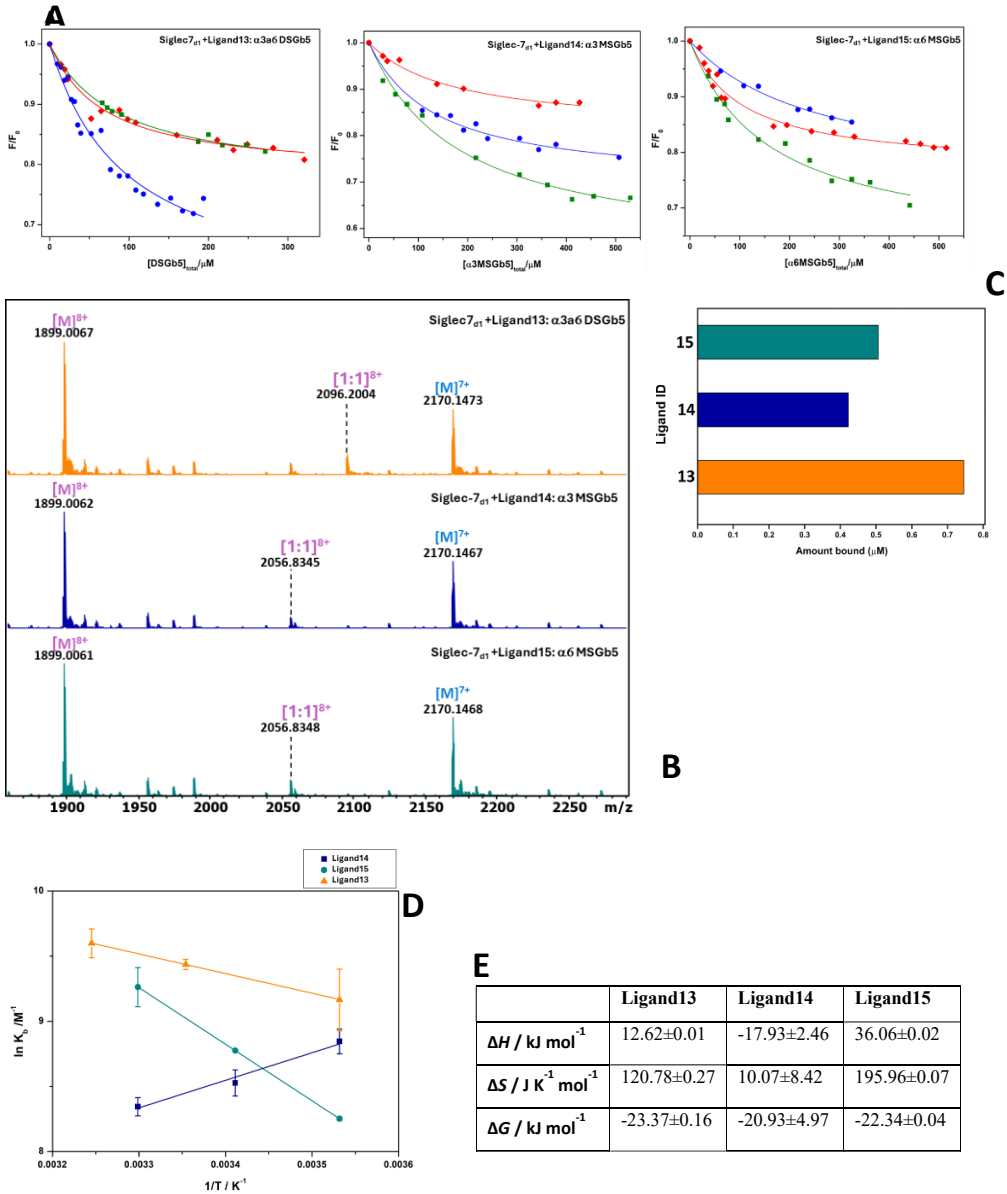


Figure V.5 **(A)**Native mass spectra of protein Siglec-7d1 with ligand12 (α6α6-DSGb5) and ligan13 (α3α6-DSGb5). (**B**) A histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7d1 for these sialoglycans.

reliable quantification of protein–ligand complexes. Mass spectra recorded in the presence of fixed ligand concentrations revealed well-resolved 1:1 protein–ligand complexes for all four systems. Quantitative evaluation of ion populations yielded association constants in line with those obtained by fluorescence spectroscopy. **Ligand 12** exhibited the strongest affinity, reflecting the cooperative contribution of its double α2,6 branches. Ligand13 shows a lower affinity than disialylate α2,6- (Figure5A/B), while monosialylated ligands displayed weaker binding (Figure 6B/C).

To better understand these systems, the interaction between Siglec-7 and ligands derived from DSGb5 was monitored with spectrofluorimetric titration recordered at different temperatures, to determine the thermodynamic parameters underlying the interaction. Upon incremental addition of ligands, the emission signal at ~353 nm decreased in a concentration-dependent manner, indicating stable complex formation. The resulting binding parameters revealed clear differences between the three systems. DSGb5 (ligand 13) displayed the strongest binding, with a binding constant K_b of $1.25 \cdot 10^4$ M^{-1}, in line with the equilibrium constant calculated with nMS. Among the monosialylated ligands,



	Ligand13	Ligand14	Ligand15
ΔH / kJ mol^{-1}	12.62±0.01	-17.93±2.46	36.06±0.02
ΔS / J K^{-1} mol^{-1}	120.78±0.27	10.07±8.42	195.96±0.07
ΔG / kJ mol^{-1}	-23.37±0.16	-20.93±4.97	-22.34±0.04

Figure V.6 (**A**) Binding isotherms measured by fluorescence spectroscopy following the formation of the complex between Siglec-7 CRD and α3α6 DSGb5 (ligand13), α2,3 MSGb5 (ligand14), and α2,6 MSGb5 (ligand15) in Phosphate Buffer Saline (PBS) buffer, pH 7.4, at temperature of 10°C (blue circles), 25°C (green squares) and 35°C (red rhombuses) for ligand 13, while for ligand 14 and ligand 15, 10°C (blue circles), 20°C (green squares) and 30°C (red rhombuses). The solid lines represent the best fit of the experimental data according to a 1:1 binding model. (**B**) Electrospray mass spectra of protein Siglec-7_{d1} with ligand13, ligand14 and ligand15. Analysis focused on the 8+ charge states and 1:1 protein–ligand complexes; however, additional low-intensity signals corresponding to 1:2 protein–ligand stoichiometries were also detected for ligands 14 and 15, likely indicative of nonspecific binding events. (**C**) histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7_{d1} for these sialoglycans. (**D**) Plots of ln(K_b) vs. 1/T for the Siglec-7 complex formation between (blue square) α2,3 MSGb5, (green circle) α2,6 MSGb5 and (orange triangle) α3α6 DSGb5. The solid lines represent the best fits to the experimental data. The slope of ln(K_b) vs 1/T is equal to - $\Delta_b H°/R$, allowing estimation of the enthalpy change of binding, $\Delta_b H°$. (**E**) Table reported the $\Delta_b G°$, $\Delta_b H°$ and $\Delta_b S°$ value obtained from plotting ln(K_b) vs 1/T.

ligand 15 exhibited slightly higher affinity (K_b = 6.47·10^3 M^{-1}) compared to

ligand 14 (K_b = 5.05·10^3 M-1). These relative affinities reproduce the same trend observed by nMS analysis: **ligand 13 >> ligand 15 > ligand 14** (Figure6A/C). Temperature-dependent analysis enabled van't Hoff plots, from which enthalpy and entropy contributions were extracted (Figure 6D/E). The disialylated **ligand 13** exhibited intermediate thermodynamic behavior, with a moderate entropic gain and mild endothermicity, indicating a combined contribution of enthalpic and entropic terms. **Ligand 15**, which contains a single α2,6-linked sialic acid, displayed the most entropically favored profile, characterized by a large positive $\Delta_bS°$ and a significantly endothermic $\Delta_bH°$, indicating that binding is strongly driven by entropy. This result contrasts sharply with **ligand 14**, which showed a mildly favorable entropy change coupled with a highly exothermic enthalpy, indicative of a classical enthalpy-driven recognition process.

Table V.4 Equilibrium binding constants of Siglec-7 with ligand12,13,14 and liang15 obtained from native mass spectrometry (nMS) and steady-state fluorescence titration (ligand12 still under analysis). Both $\boldsymbol{K_b}$(association constant) and $\boldsymbol{K_D}$(dissociation constant) are reported. The values determined by nMS are related to 8+ charge states, whereas fluorescence-based constants derive from non-linear fitting curves using a single-site binding model.

	nMS		**Fluorescence**	
Ligand	**K_b /[M^{-1}]**	**K_D [M]**	**K_b /[M^{-1}]**	**K_D [M]**
12	1.0910^5	9.1710^{-6}	//	//
13	2.11·10^4	4.75·10^{-5}	1.25(±0.5)·10^4	7.97·10^{-5}
14	7.18·10^3	1.39·10^{-4}	7.62(±2.9)·10^3	1.31·10^{-4}
15	8.74·10^3	1.18·10^{-4}	3.890(±0.9)·10^3	2.57·10^{-4}

This model suggests that solvent rearrangement at the Siglec-7 interface is the primary determinant of binding thermodynamics for the different ligands analyzed. In particular, the large positive $\Delta_bS°$ observed for **ligand 15** is consistent with the displacement of ordered water molecules from the binding site upon complex formation. Structural evidence directly supports this interpretation: in the Siglec-7/GT1b complex, a water molecule hydrogen-bonded to the backbone nitrogen of Asn133 is displaced by the C8-OH group

of sialic acid (Attrill et al., 2006). This observation illustrates how the specific topology of the sialic acid moiety can physically expel bound water and reorganize the hydration network at the binding interface, providing a structural explanation for the large entropy gain measured for the α2,6-linked ligand.

Literature provides quantitative demonstrations that the expulsion of conserved waters generates favorable entropy but may incur an enthalpic penalty. In the well-studied case of Concanavalin A (ConA), modifying a trimannoside ligand to occupy a position typically filled by structured water resulted in increased $\Delta_bS°$ (water released to bulk) but rendered $\Delta_bH°$ less favorable (J.R. Woods et al., 2001). These studies offer a clear and quantitative precedent for interpreting the thermodynamics observed for Siglec-7, where α2,6 topologies appear to evacuate high-energy bound waters more effectively than α2,3 topologies, yielding greater entropy gains at the expense of enthalpy.

Conversely, the strongly favorable $\Delta_bH°$ and modest $\Delta_bS°$ observed for ligand14 (α2,3-linked ganglioside) are consistent with the retention of water-mediated hydrogen bonds rather than displacement of structured water. Additional studies have corroborated this *'entropy-from-desolvation'* principle in lectin-glycan interactions. For example, E-selectin–sialyl-Lewis^x binding has been described in terms of a *"pre-organized water oligomer*", where ligand binding is favored by releasing interfacial water clusters along an extended interface, resulting in an entropy-driven recognition event (Binder et al., 2012). Although selectins and Siglecs differ structurally, both protein families illustrate how linkage and structural presentation (e.g. α2,6 vs α2,3) tune hydration patterns and modulate $\Delta_bS°$.

In the case of the disialylated **ligand 13**, the thermodynamic and binding data supports a cooperative mechanism involving complementary engagement of both sialic acid arms. DSGb5 binds Siglec-7 with the highest affinity observed among the three ligands, with measured association constants of approximately $1.25 \cdot 10^4$ M^{-1} (spectrofluorimetric titration) and $1.89 \cdot 10^4$ M^{-1} (native MS). The observed moderate endothermicity and positive entropy binding, likely reflect a combination of enthalpic stabilization from directed polar or ionic contacts

mediated by the α2,3- arm, and entropic gain arising from solvent displacement driven by the α2,6- arm. Such a dynamic and complementary mode of interaction allows ligand13 to achieve the most favorable $\Delta_b G°$ and the highest overall affinity by balancing the enthalpic penalty associated with extensive solvent reorganization against significant entropic benefits. The dual-arm engagement proposed here is supported by broader studies indicating that ligand-induced displacement of structured waters at protein interfaces generally leads to enhanced entropy and less favorable enthalpy, exemplified in various lectin-ligand binding systems. Taken together, these observations and literature precedents establish a coherent thermodynamic framework to explain the differential binding profiles observed for DSGb5-derived ligands with Siglec-7. DSGb5 achieves its superior affinity by leveraging the complementary enthalpic and entropic contributions of its two sialylated arms, whereas monosialylated ligands demonstrate clearly divergent binding regimes dictated by their distinct topologies and consequent solvent interactions.

5.2.3 Polysialic Acids

Polysialic acids (polySia) are linear polymers of sialic acid residues linked through α2,8-, α2,6- and α2,3- glycosidic bonds. In pathogenic *Neisseria meningitidis*, the composition of the polysialic capsule defines the serogroup: serogroup B expresses α2,8-linked polysialic acid, whereas serogroup Y produces a capsule composed of α2,6-linked polysialic chains. These distinct linkage patterns strongly influence the polymer's three-dimensional conformation, charge distribution, and hydration properties, which in turn

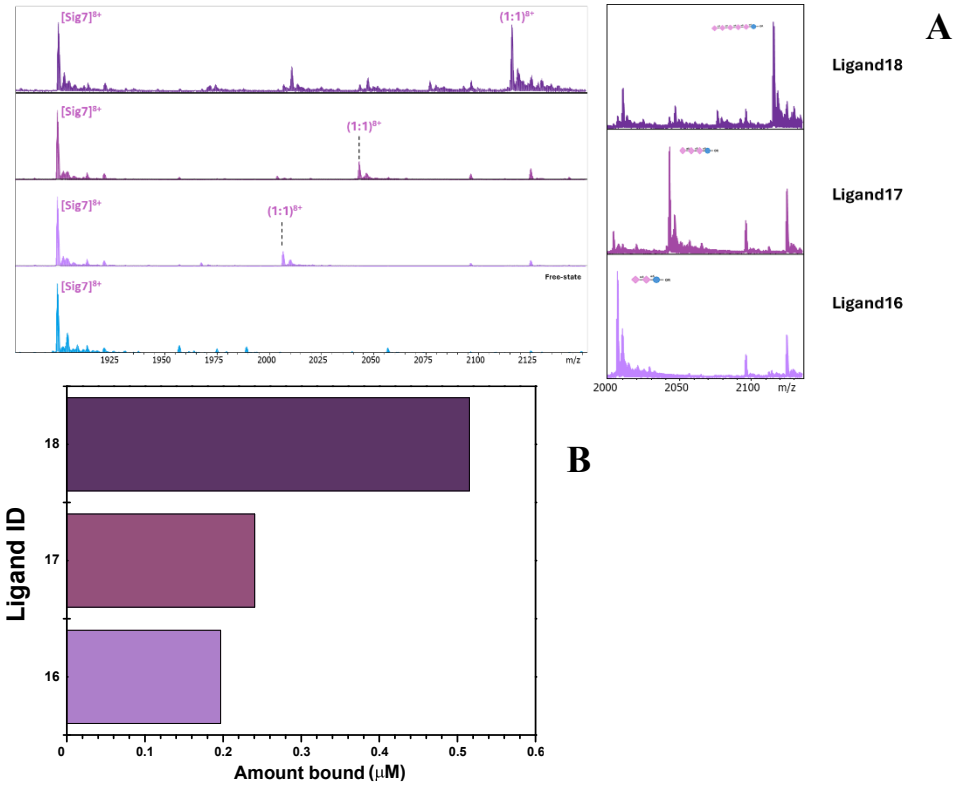


Figure V.7 (**A**)Electrospray mass spectra of protein Siglec-7_{d1} with ligand16, ligand17 and ligand18. Analysis focused on the 8+ charge states and 1:1 protein–ligand complexes as for the previously ligands. (**B**) histogram generated from MS-derived equilibrium association constants illustrates the differential binding affinities of Siglec-7_{d1} for these sialoglycans.

affect recognition by Siglecs. Under native ESI conditions, Siglec-7 formed well-defined 1:1 complexes. In all spectra, the mass peak of the free protein remained compact, indicating that the ionization conditions preserved the folded state, while a second peak at m/z = 2006.8055; 2043.3196; and 2116.0881 (ligand 16, ligand 17, and ligand 18, respectively) corresponded to

the [1:1] complex in the 8+ charge state. A clear dependence on chain length emerged from this analysis. Shorter oligomers produced weaker but detectable complex peaks; the intermediate-sized **ligand 17** produced a marked increase in bound state intensity; and the longest **ligand 18** showed the most prominent [1:1] signal, as reflected in the bound/free ratios summarized next to the spectra. This monotonic trend indicates that additional α2,8-linked residues enhance recognition, most likely by improving electrostatic complementarity and allowing extended polar contacts and water-mediated interactions around the sialic acid carboxylate. The peak shapes remained symmetric and free of pronounced addition, supporting the formation of specific complexes rather than nonspecific clustering of highly anionic ligands. The integration resolved based on the charge state allowed for a qualitative classification of apparent affinity following chain length. Taken together, these results establish that Siglec-7 effectively recognizes α2,8-polySia under native conditions and that the binding strength increases with polymer length within the tested window.
To further characterize the interaction between Siglec-7 and α2,8-linked polysialic acids, steady-state fluorescence titrations were performed for the ligands corresponding to the shorter (ligand 16) and longer (ligand 18) chain lengths. Upon incremental addition of each ligand to a fixed concentration of Siglec-7, a progressive quenching of the intrinsic tryptophan emission at approximately 353 nm was observed, consistent with specific complex formation in solution (Figure 8).
The titration curves were fitted using a 1:1 binding model, yielding binding constants (K_D) in close agreement with the affinity trend inferred from native mass spectrometry (Table 5). The longer 5-mer polysialic fragment (ligand 18) displayed the highest affinity, confirming the stabilizing effect of extended α2,8-linked chains on Siglec-7 recognition. In contrast, the shorter 2-mer oligomer (ligand 16) exhibited weaker, though still measurable, binding characterized by a lower association constant (K_b).

Fluorescence data for the intermediate 3-mer oligomer (ligand 17) are currently under analysis.

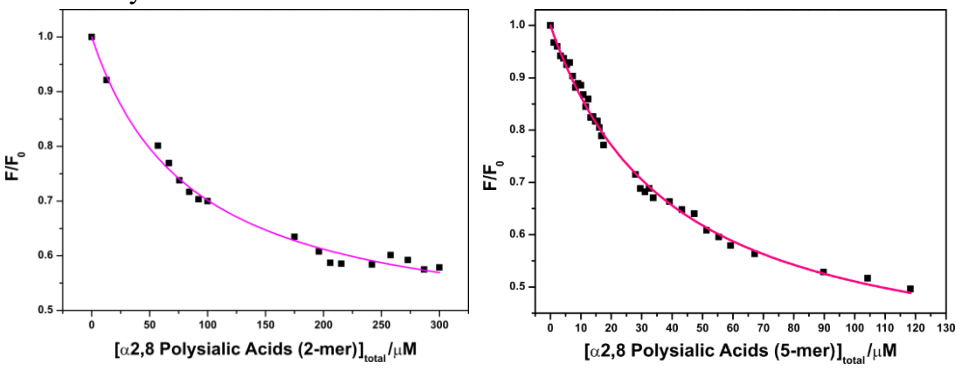


Figure V.8 Binding isotherms measured by fluorescence spectroscopy following the formation of the complex between Siglec-7 and lignad16(left panel) and lignad18 (right panel). The solid lines represent the best fit of the experimental data according to a 1:1 binding model.

Table V.5 Equilibrium binding constants of Siglec-7 and polysialilated ligand 16, 17 and 18, obtained from native mass spectrometry (nMS) and steady-state fluorescence titration (ligand 17 still under analysis). Both $\boldsymbol{K_b}$(association constant) and $\boldsymbol{K_D}$(dissociation constant) are reported. The values determined by nMS are related to 8+ charge states, whereas fluorescence-based constants derive from non-linear fitting curves using a single-site binding model.

	nMS		**Fluorescence**	
Ligand	**K_b /[M^{-1}]**	**K_D [M]**	**K_b /[M^{-1}]**	**K_D [M]**
16	$1.24 \cdot 10^{4}$	$8.08 \cdot 10^{-5}$	$1.22(\pm 0.1) \cdot 10^{4}$	$8.19 \cdot 10^{-5}$
17	$1.60 \cdot 10^{4}$	$6.24 \cdot 10^{-5}$	//	//
18	$5.45 \cdot 10^{4}$	$1.83 \cdot 10^{-5}$	$2.77(\pm 0.1) \cdot 10^{4}$	$3.61 \cdot 10^{-5}$

Fluorescence titration experiments were conducted to evaluate the interaction between Siglec-7 and α2,6-linked polysialic acid derivatives. The results, previously reported and discussed in our publication, are summarized below.

Analogously, concentration dependent reduction in fluorescence intensity upon Men-Y CPS (ligand 22), 4-mer (ligand 21) 3-mer (ligand 20) and 2-mer (ligand 19) binding was used to monitor the interaction and the equilibrium affinity constant (K_D) was determined by non-linear regression analysis. The interpolation of the fluorescence data provided similar K_D values all in the low micromolar range (Figure 9).

5.3 Conclusion

The systematic screening of 21 structurally diverse glycans against Siglec-7 revealed clear and biologically meaningful trends. First, linkage topology proved to be a dominant determinant of affinity: ligands bearing α2,6 linkages generally outperformed their α2,3 counterparts, and increasing inclusion of α2,8 motifs further enhanced binding—suggesting that spatial arrangement and conformational flexibility of sialic acid residues can modulate lectin–glycan

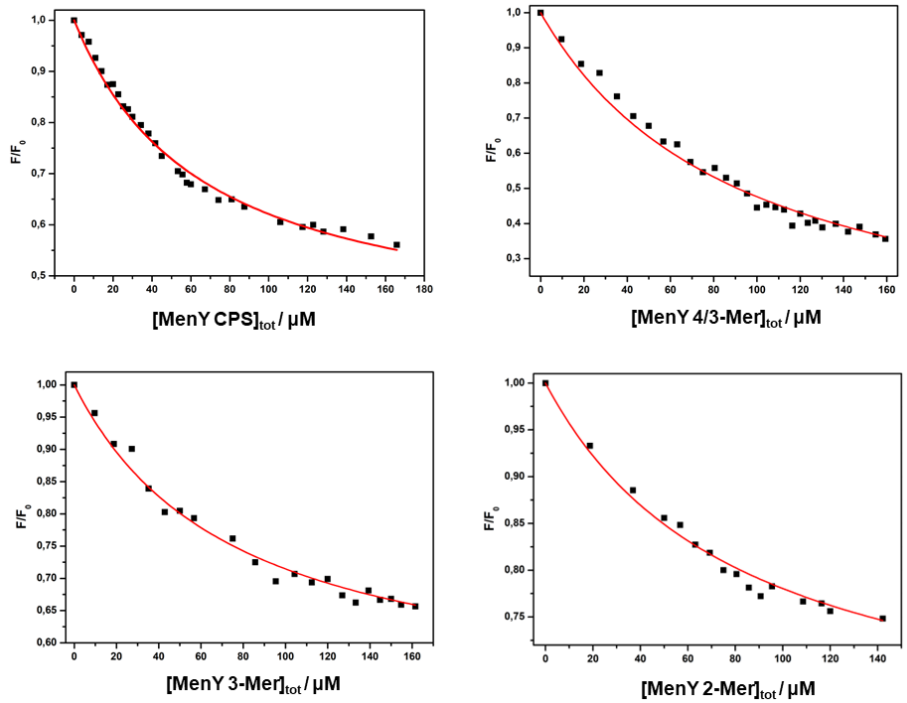


Figure V.9 Fluorescence titration of Siglec-7 upon the addition of increasing concentrations of Men-Y CPS and oligos (4-Mer, 3-Mer, and 2-Mer). The emission spectra were recorded by using an excitation wavelength of 295 nm and a temperature of 25 °C. Figure adapted from the reference article doi.org/10.1021/jacsau.5c00214.

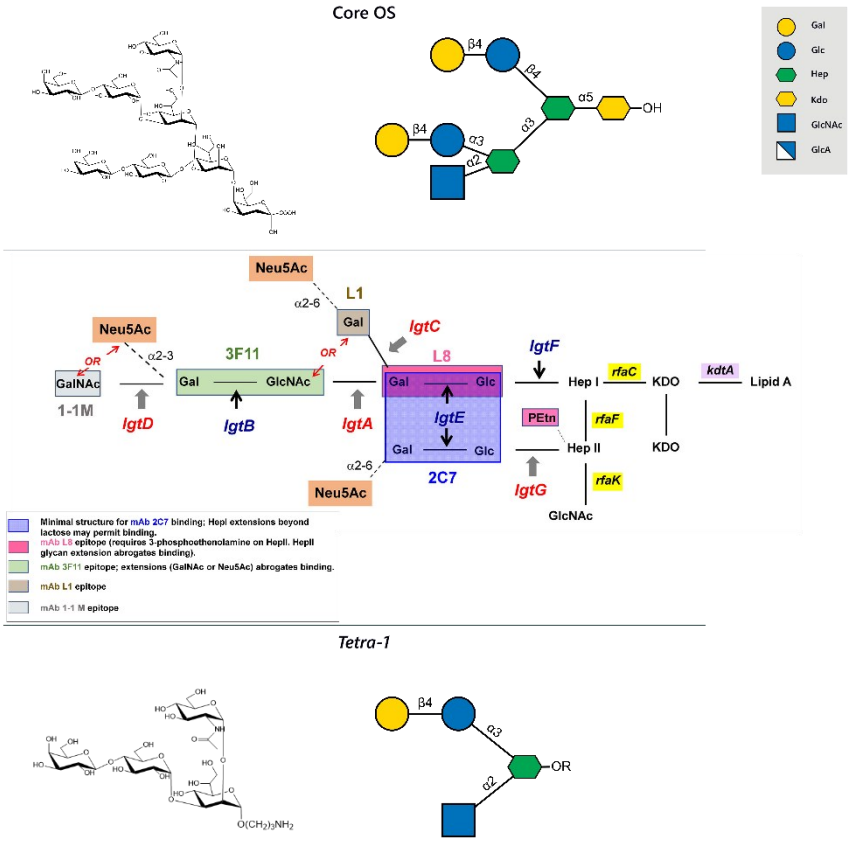
complementarity. Second, degree of sialylation contributed additively to binding strength: higher-order polysialylated ligands (tri-, tetra-, pentasialic units) consistently exhibited stronger interaction, likely due to extended electrostatic contacts and multivalent stabilization within the Siglec binding site. The absence of evident higher stoichiometries confirms that enhancement arises from refined single-site binding rather than multiple binding sites. Third, sulfation exerted a modulatory but nonuniform effect: while some sulfated ligands displayed modest increases in apparent affinity, others did not, indicating that the benefit of added negative charge may depend on local geometry and on whether the sulfate is favorably oriented toward basic residues in Siglec-7. Finally, the good concordance between nMS-derived and fluorescence-derived K_b / K_D values underscores the robustness of our label-free, complementary approach. By combining stoichiometric and thermodynamic perspectives, we have built a more granular energetic map of Siglec-7–glycan recognition. These insights not only refine our understanding of Siglec-7 specificity but also offer rational guidance for the design of synthetic ligands or inhibitors directed at modulating immune checkpoint functions mediated by sialic acid–lectin interactions.

VI. Neisseria gonorrhoeae LOS derivatives recognition by antibody 2C7

6.1 Introduction

In recent years, antibiotic resistance (AMR)[1] has emerged as a serious global health threat, mainly due to overuse of antibiotics in both human and veterinary medicine that cause considerable difficulties in the healing and treatment of infectious diseases, as well as causing an increase in hospitalizations and healthcare costs[2]. A concerning case is represented by *Neisseria gonorrhoeae*, an important cause of sexually transmitted infections worldwide, which has developed resistance to nearly all antibiotics, including penicillin and fluoroquinolones[3]. In addition to having acquired several resistance mechanisms to antibiotics, this bacterium is particularly difficult to treat because it has developed a variety of strategies to evade the host immune system[4,5]; these include antigenic and phase variation in surface-exposed factors, such as type IV pili, Opa proteins, and LOS[6,7]. A key component of the *N. gonorrhoeae* outer membrane is the prominent gonococcal Lipo-oligosaccharide (LOS), which play a seminal role in host cell colonization, adhesion and are thus critical for bacterial survival in the host. Structurally, gonococcal LOS consists of hydrophobic lipid A portion, an highly conserved inner core region and a structurally variable outer core oligosaccharides (OS) chains, build up by sequentially glycose additions on the conserved inner core portion; In detail, three OS chains (α,β,γ) extending from two heptose residues, which are connected to lipid A via the 2-keto-3-deoxy-mannooctulosonic acid (Kdo) residue[8,9], which form the trisaccharide core region. The α-chain branch out from the first heptose (HepI), while the β and γ chains protrude from the second heptose (HepII)[8,10]. The number of each chains and their composition is directly dependent on the differential expression of the lgtG, lgtA, lgtC and lgtD genes encoding the LOS glycosyltransferases[11]; These genes can be expressed differently between multiple gonococcal strains and even within the same strain during infection due to the effect of phase-varying genes expression[10].

However, the immunogenic potential of LOS has been identified as a key factor in the development of effective vaccines, particularly given the lack of success in generating a protective immune response with traditional vaccine antigens[12] based on pili or outern-membrane proteins[13]. Numerous *in vitro* and *in vivo* studies[14,15] have shown that specific monoclonal antibodies (mAbs), such as mAb 2C7, are able to target a conserved OS epitope, called 2C7 epitope[16], be made up of Lac-β-(1→4) that compose the α-chain, branch to αHepI and the β-chain Lac-α-(1→3) to αHepII residue (Scheme 1).



Scheme VI-1 *Neisseria gonorrhoeae* lipooligosaccharide (LOS) structure. (A) Schematic representations of three LOS glycoform illustrating the core oligosaccharide composition. Symbolic representation follows SNFG. (B) Biosynthetic pathway of LOS glycan core, highlighting the roles of key lgt glycosyltransferase genes. The minimal epitope (2C7) required for recognition by mAb 2C7 is shown in blue, with additional epitopes targeted by mAbs L8, L1, 3F11 and 1-M indicated in color-coded boxes. Image (B) adapted from ref.10. (C) SNFG representation of vaccine candidate used in this study.

Notably, 95% of the 68 clinical isolates from cervical secretions of women, exhibited reactivity to mAb 2C7[17]. From a clinical perspective, antibody-based treatment not only reduced the severity of the infection but also reduced its duration, suggesting a high prevalence of the targeted epitope in the clinical cases studied[14,18], as well as a significant reduction in the gonococcal fitness observed during treatment. Overall, the results of this study suggest that antibodies targeting the 2C7 OS epitope have the potential to be developed into a vaccine or therapeutic approach for gonococcal infections, by mediating bacterial recognition and opsonophagocytic complement-dependent effect[19].
Understanding the intricate relationship underlying the molecular recognition of OS by mAb 2C7 is crucial for enhancing the efficacy of antibodies targeting bacterial glycan epitopes. Therefore, we attempted to dissect this complex molecular system using the core oligosaccharide from *Neisseria gonorrhoeae* LOS and a specific synthetic OS region (here called Tetra-1), a potential vaccine candidate of the *Neisseria gonorrhoeae* LOS as well as to deepen the interaction with mAb 2C7. Notably, preliminary and promising results from Western Blot assays, which demonstrated the ability of mAb 2C7 to recognize the specific OS, prompted us to delve deeper into the molecular and structural details underlying this interaction. To shed light on this system, we employed a combination of NMR, computational, and biophysical approaches to evaluate the key factors governing this interaction. Our findings are promising not only for understanding the affinity and specificity of core LOS for mAb 2C7 but also for laying the groundwork for future improvements in mAb 2C7 as a therapeutic antibody directed against this bacterium.

6.2 Results

Extraction, purification and chemical characterization. LOS was extracted from bacterial cells using the hot phenol/water method[20]. The extracted LOS was enzymatically digested with DNase, RNase, and protease to remove cellular contaminants. An additional purification step was performed using

size-exclusion chromatography. Type and quality of the LOS were assessed by SDS-PAGE, followed by gel staining with silver nitrate[106] that revealed the rough nature of the extracted LPS (R-LPS or LOS), as indicated by the low-molecular-weight band observed at the bottom of the gel.

The LOS was then subjected to a mild acid hydrolysis, to selectively cleave the acid-labile glycosidic bond between the Kdo and the non-reducing GlcN of the lipid A region. The purified core oligosaccharide fraction was analyzed via negative-mode MALDI-TOF MS (Figure 1B) that revealed five species that confirmed the nature of the isolated OS.

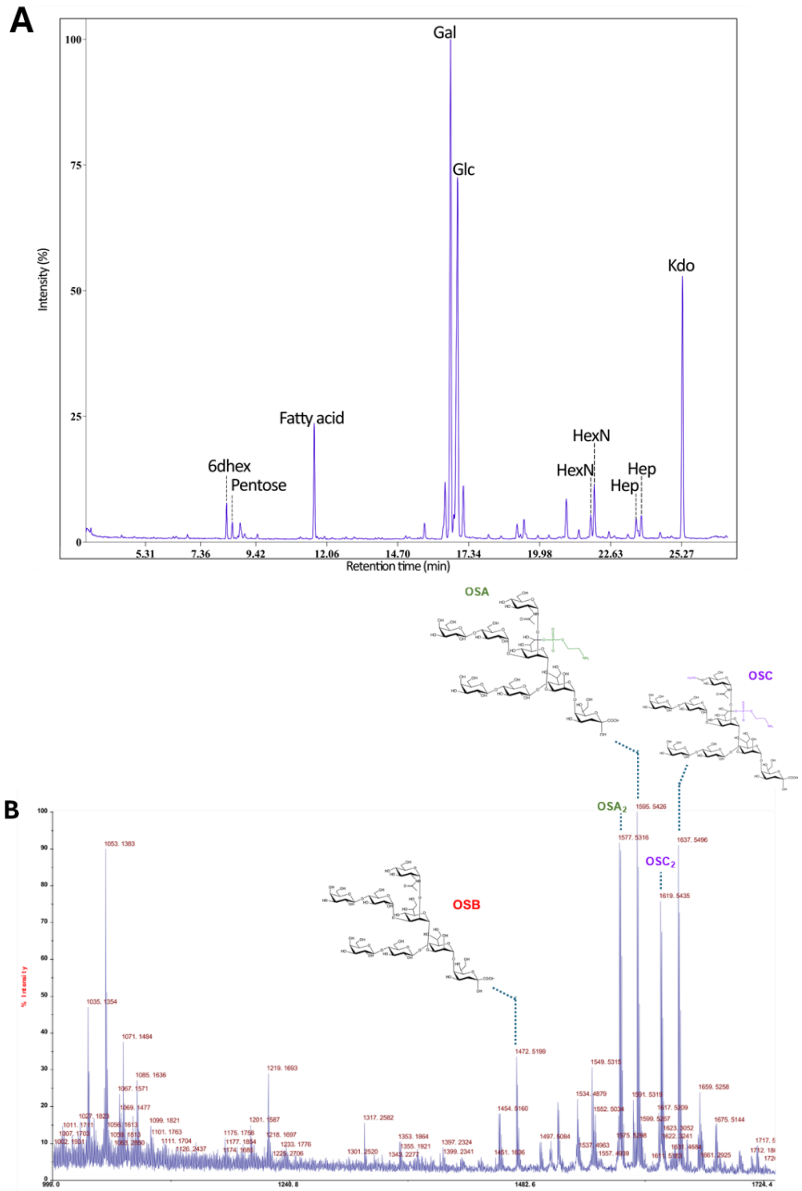


Figure VI.1 (A) Compositional analysis of the purified *Neisseria gonorrhoeae* LOS sample by Gas Chromatography–Mass Spectrometry (GC–MS). The chromatogram shows the derivatized monosaccharide components identified after acidic hydrolysis, reduction, and acetylation. Major peaks correspond to Galactose (Gal), Glucose (Glc), N-acetylglucosamine (GlcNAc), Heptose (Hep), and 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo), with minor components such as Hexosamines (HexN), Pentoses, 6-deoxyhexoses (6dHex), and fatty acid derivatives also detected. (B) MALDI-TOF mass spectrum of the oligosaccharide fraction obtained after mild acid hydrolysis (1% AcOH, 100 °C) of the LOS. The treatment selectively cleaves the acid-labile glycosidic bond between Kdo and GlcN, removing the lipid A moiety and allowing detection of the intact oligosaccharide region. Five distinct species were identified. Structural annotations are based on high-resolution accurate mass data and supported by monosaccharide composition.

Conjugation tests. To verify that the 2C7 epitope on the gonococcal oligosaccharide (OS) remains accessible after conjugation to the carrier protein Cross-Reactive Material (CRM), and to assess its potential use as a candidate antigen in glycoconjugate vaccine development, a Western blot analysis was performed by Dr. Sanjay Ram and his team at Department of Infectious Diseases and Immunology, University of Massachusetts Chan Medical School, Worcester, Massachusetts 01605, United States. Figure 3A shows the immunoblot results obtained by probing the OS–CRM conjugate with the monoclonal antibody mAb 2C7. Lane 1 and Lane 2 contain the OS–CRM glycoconjugate loaded at 0.2 µg and 0.4 µg, respectively, while Lane 3 contains unconjugated OS (derived from *N. gonorrhoeae*) as a positive control. This experimental design allows a direct comparison of antibody reactivity between the free OS and the OS covalently attached to the protein carrier. As expected, the unconjugated OS in Lane 4 showed a clear immunoreactive signal upon probing with mAb 2C7, confirming the presence of the recognized carbohydrate epitope in the native OS (positive control). Crucially, both conjugate samples (Lanes 2 and 3) also yielded distinct bands on the blot, indicating that mAb 2C7 successfully binds to the OS–CRM conjugate. The conjugate bands appear at a molecular weight corresponding to CRM (~60–70 kDa), slightly higher than the free OS, which is consistent with the OS attached to the ~62 kDa carrier protein. The immunostaining intensity in Lane 3 (0.4 µg conjugate) is notably stronger than that in Lane 2 (0.2 µg), reflecting the higher amount of antigen loaded and further demonstrating the specificity of the antibody binding. The presence of mAb 2C7-reactive bands in the conjugate lanes (2–3), comparable to the positive control OS in Lane 4, confirms that the epitope recognized by mAb 2C7 is preserved after conjugation. Therefore, the key carbohydrate antigenic determinant has not been destroyed or masked by the chemical coupling to CRM, and it remains accessible for antibody recognition in the glycoconjugate form.

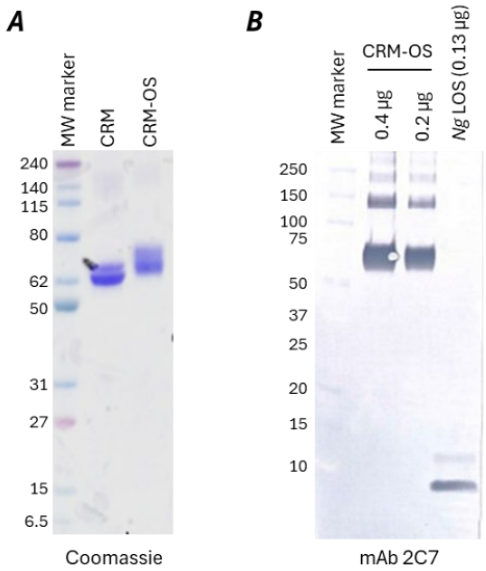


Figure VI.2 Characterization of CRM-OS conjugate. (A) Coomassie stain reveal the CRM-OS molecular weight. (B) Reactivity validation by using Western Blot on CRM-OS (Lane 2 and 3, at 0.2 and 0.4 µg, respectively) confirm the recognition for the conjugated system; OS of Neisseria gonorrhoeae are used as positive control (Lane 4).

NMR Structural Analysis. The isolated OS fraction was in turn analysed via NMR spectroscopy HSQC NMR spectrum (Figure 3) showed the presence of anomeric signals assigned to seven different spin systems: αHepI(F) (5.82 ppm); βGlc(E) (5.30 ppm); αGlcNac(D) (5.05 ppm); αHepII (H) (4.97 ppm);

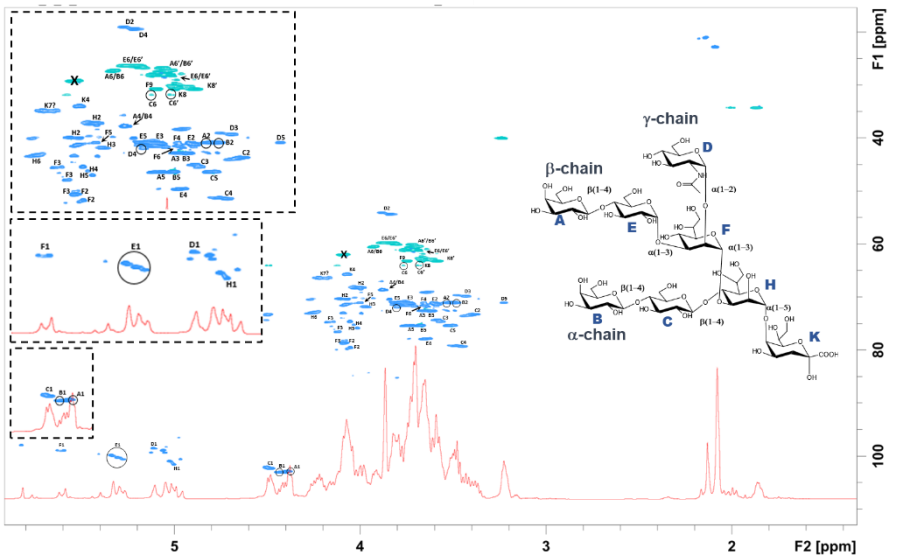


Figure VI.3 ^{1}H-^{13}C- HSQC spectrum of the OS extracted from *Neisseria gonorrhoeae* strain 15253; Zoom show the anomeric region (below and middle) and core region (above).

αGlc(C) (4.49 ppm); ᵗβGal(B) (4.43 ppm); ᵗβGal (A) (4.39 ppm). Given the antibody's preferential selectivity for the β- and γ-chain regions, the interaction of this OS segment with the antibody was further investigated using a synthetic epitope, designated Tetra-1 synthetized from Prof. Xuefei group (Departments of Chemistry and Biomedical Engineering, Institute for Quantitative Health Science and Engineering, Michigan State University), to provide further insight into the molecular features required for antibody recognition. The structural characterization of this oligosaccharide was carried out using 1D and 2D NMR experiments. The NMR analysis identified four distinct spin systems (Figure 4), corresponding to αHepI(F) (4.96 ppm), βGlc(E) (5.13 ppm), αGlcNAc(D) (5.08 ppm), and tβGal(A) (4.31 ppm).

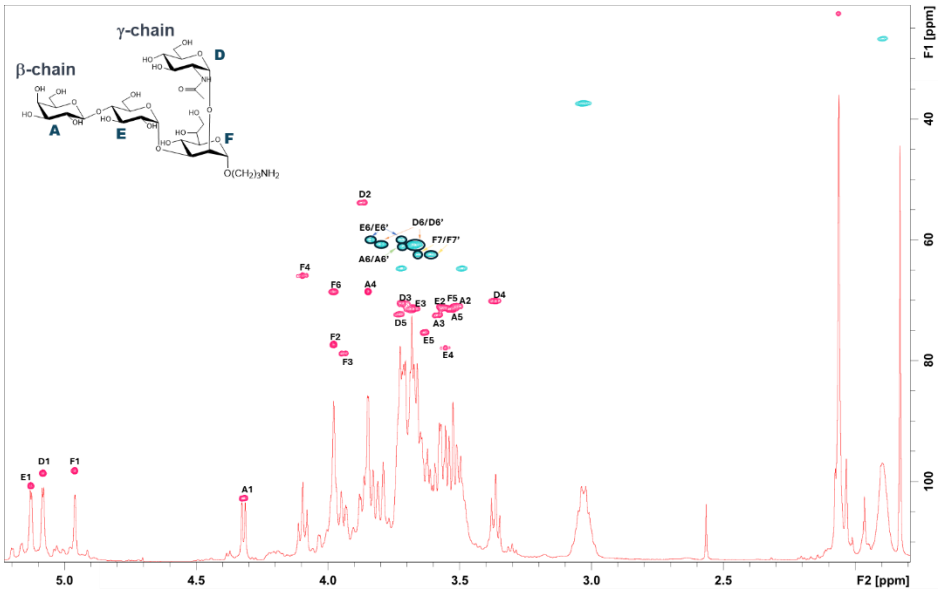


Figure VI.4 ^{1}H-^{13}C- HSQC spectrum of the synthetic vaccine candidate Tetrasaccharide (Tetra-1).

Core OS – mAb 2C7. To elucidate the key regions of the OS ligand predominantly involved in the binding events, ligand-based NMR experiments were used on the mixture mAb 2C7 and OS (Figure 5). STD NMR mapped the interacting epitope, localized in the terminal region of the β-chain, with the β-Gal residue showing the strongest STD enhancements. These results suggest that the α-chain region is solvent-exposed and does not significantly contribute to the binding interaction. The main intense STD signals were attributed to H2, H3, and H4 of β-Gal, followed by moderate contributions from H5. The terminal αGlc linked to βGal also showed STD signals, particularly at H2, H3,

and H4, indicating a partial contribution to the binding event. On the γ-chain, the αGlcNAc residue displayed the most substantial contributions at H2, with

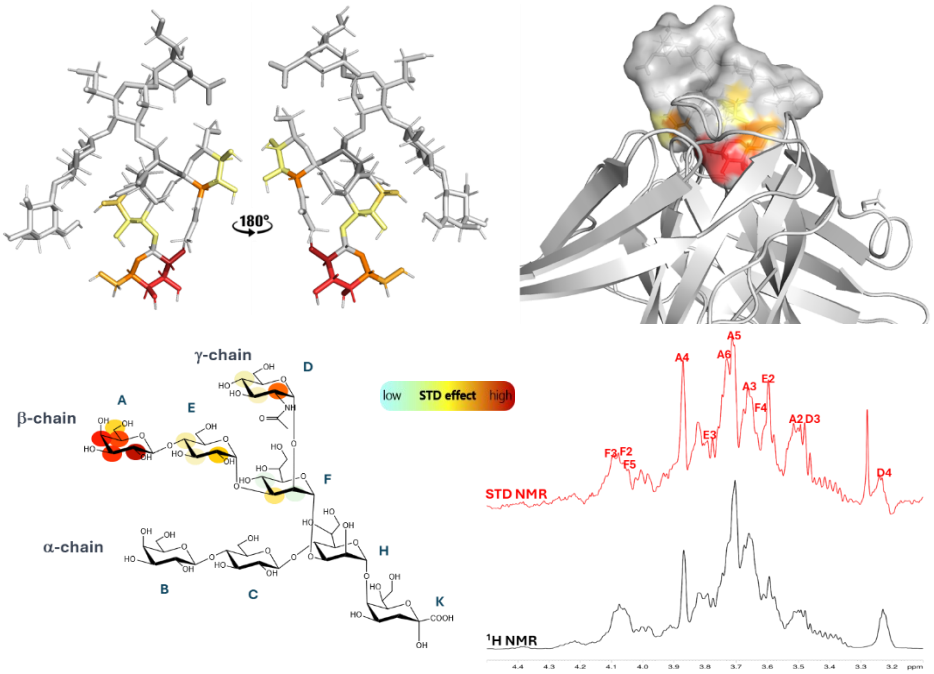


Figure VI.5 STD NMR analysis and Epitope mapping of *Neisseria gonorrhoeae* OS. (Top left) Representative 3D structures of the ligand in colored according to STD signal intensity. (Top right) Epitope mapping of OS (surface) highlights the protons most involved in interaction with mAb2C7(grey cartoon). (Bottom left) Schematic view of OS structure colored according to STD values. (Bottom right) ^{1}H and STD spectra; %STD are obtained by the ratio $(I0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ is the signal intensity in the STD-NMR spectrum (red) and I_0 is the peak intensity of the off-resonance spectrum (black).

H3 and H4. In contrast, very low STD effects were observed for the α-chain.

The bioactive conformation of OS was determined by Tr-ROESY NMR (see below and Figure 6B for detailed data), and computational studies brought further insights into the molecular recognition features. As for the computational analysis, we constructed and parametrized the uncommon sugars in the OS region (HepI, HepII, and Kdo) using the AMBER18 tool. Subsequently, the disaccharides forming the OS region were constructed to explore the energetically accessible conformational spaces through molecular mechanics (MM) calculations, following standard protocols (see chapter VIII for methods). The resulting adiabatic energy maps, which depict the ϕ and ψ glycosidic torsion angles (data not shown), confirmed that the global energy

minimum was aligned with the *exo-anomeric* effect, providing important information into the preferred conformations of the oligosaccharide.

On the basis of the adiabatic energetic maps obtained and Tr-ROESY data collected, the starting OS underwent extensive MD simulation. Using the crystallographic structure obtained by the Beernik P. group (Figure 7). the 3D complex was further optimized using the Maestro program suite and the obtained complex was subjected to extensive molecular dynamics (MD) simulations (500 ns), and the results were subsequently evaluated and compared based on NMR data. During the MD simulations, RMSD, inter-residue contact established and the glycosidic torsion angles ϕ/ψ were sampled to provide insights into the bioactive conformational behavior of the oligomer when interacts with the FAb 2C7. RMSD performed on both protein backbone

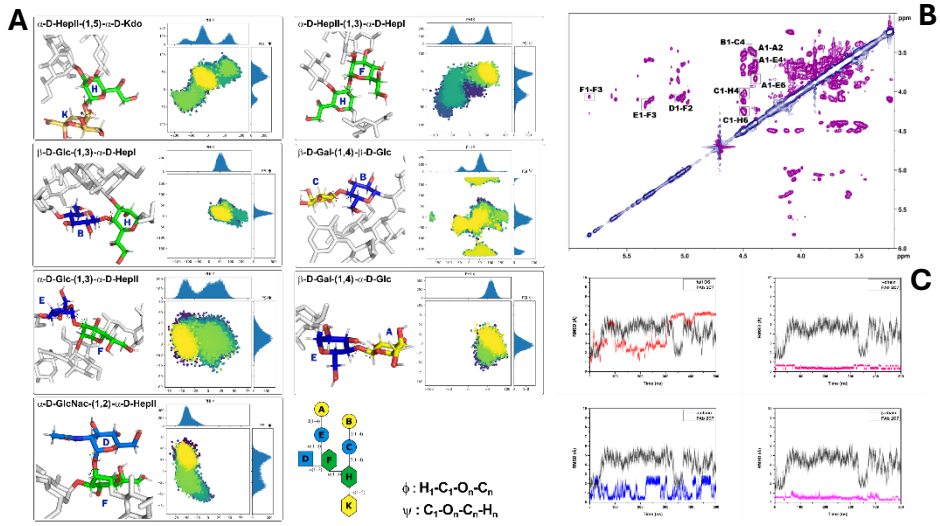
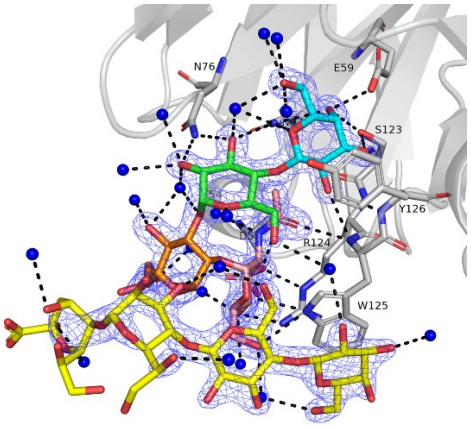


Figure VI.6 (A) Adiabatic energy maps illustrating the conformational landscape of the OS glycosidic linkages, with low-energy conformers represented in blue and highly populated regions highlighted in yellow. (B) Transferred-ROESY NMR spectrum confirming intermolecular NOE contacts between OS protons and mAb2C7. (C) Root mean square deviation (RMSD) plots from molecular dynamics simulations for each chain, depicting the stability and conformational fluctuations of both the ligand and the protein during the simulation timeframe (500 ns).

and oligo residues, using the first frame as a reference and evaluating both the entire OS and individual chains, indicates a clear stability of all the segments analyzed. The simulated average inter-residue distances, calculated using the $\langle r^{-6} \rangle$ values, showed a good correlation with the acquired experimental data.

The adiabatic map of the glycosidic ϕ/ ψ torsion obtained during the MD run, indicates that the ϕ angles predominantly adopted values consistent with the *exo-anomeric* conformation, as inferred from the inter-residue ROE contacts (Figure 6A and C). The findings reveal that the α-chain remains spatially distant from the binding pocket and does not form significant interactions with the



Ligand atom		Hydrogen Bonds	HB/atom
BGC 6	O2	Asn76(ND2):3.28	3
	O3	Asn76(ND2):3.14	
	O3	Asn57(ND2):3.00	
GAL 7	O2	Trp125(N):3.14 Tyr126(N):3.01	7
	O3	Ser123(OG):2.90	
	O4	Asn57(ND2):3.18 Glu59(OE2):2.78 Ser123(OG):3.24	
	O5	Asn57(ND2):3.05	
NDG 8	O3	Asp55(O):2.75 Arg124(NE):2.70	6
	O4	Arg124(NH2):3.06	
	O5	Trp125(NE1):3.27	
	O6	Trp125(NE1):3.15	
	O7	Trp125(N):2.95	
Total		16	

Figure VI.7 X-ray crystallographic structure of mAb2C7 in complex with OS. The electron density map (2Fo–Fc, contoured at 1.0 σ) highlights the OS ligand accommodated within the mAb2C7 binding pocket. Dashed lines indicate hydrogen bonds and polar interactions that stabilize the protein–ligand interface. The accompanying table summarizes the specific intermolecular contacts established between mAb2C7 residues and the OS ligand.

protein surface. In contrast, during the binding process, several hydrogen bond interactions were observed involving the β-chain; for istance, the hydroxyl group at position 3 of αGlc made contact (49%) with Asn56 , and further, predominant hydrogen bonds were mainly associated with the ¹βGal residue of the β-chain. Strong interactions sampled during MD simulation were observed between αGal-H4 and Glu59 (58%), αGal-O2 and Tyr126 (49%), αGal-O4 and Asn57 (45%) and αGal-H2 and Ser123 (31%). All interactions in the bound complex were found in the variable segment of the heavy chain (V_H) of Fab immunoglobulin. We observed strong correlation between experimental data and computed MD results, demonstrating the reliability of the simulation approach in reproducing experimental observations.

Synthetic Tetra1 – mAb 2C7 binding studies. Binding studies were analogously conducted on *Tetra*-1 synthetic oligosaccharide. STD NMR results confirmed

that the primary interaction occurred at the terminal region of β-chain, particularly at the Lac moiety linked to HepII. The STD enhancements observed for the H3 and H4 protons of the terminal βGal were the most significant, with contributions also from H2 and H5. Furthermore, the αGlc displayed favorable STD responses, particularly at H3, with a minor contribution from H2 and H4, suggesting a partial role in stabilizing the interaction. On the γ-chain, the αGlcNAc residue contributed moderately to binding, with its H2 proton showing detectable STD enhancement, while H3 and H4 were only weakly involved. αHepII moiety displayed significant contributions at H2 and H3. To further investigate the bioactive conformation of Tetra-1 in the bound state, Tr-ROESY experiments reveal no significant conformational changes between free and bound states, (Figure 9).

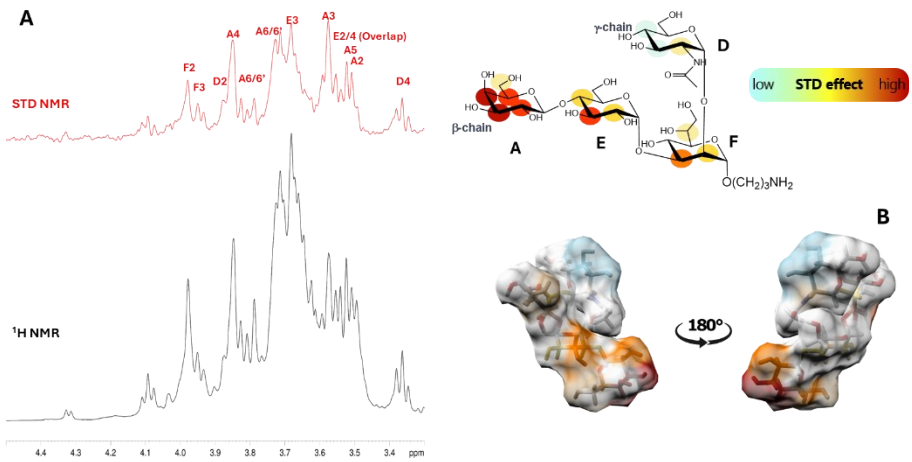


Figure VI.8 STD NMR and bioactive conformation of the Tetra-1 oligosaccharide. (A) STD NMR spectrum of Tetra-1 in complex with mAb2C7, showing the proton signals contributing to the binding epitope. Signal intensities correspond to the degree of ligand–antibody interaction, highlighting the main contact regions. (B) Surface representation of the bioactive conformation of Tetra-1 bound to mAb2C7. The ligand is shown in its lowest-energy pose within the binding interaction hotspots mapped according to the STD intensity distribution.

As above a 3D complex was built and subjected to 500 ns MD simulations. Notably, an excellent agreement was observed between the experimental and simulated data, highlighting the robustness of the simulation approach in reproducing experimental observations. Significant hydrogen bond interactions were sampled, primarily involving the hydroxyl groups at positions 2, 3, and 4

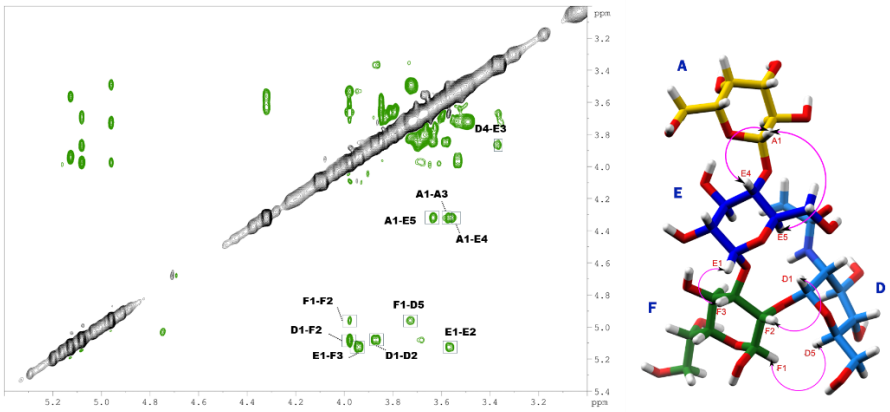


Figure VI.9 Tr-ROESY of synthetic Tetra-1.

of the terminal βGal. Specifically, ᵗβGal-O2 formed hydrogen bonds with Ser123 as an acceptor (17%) and with Tyr126 as a donor (47%). For ᵗβGal-O3, interactions were observed with Glu59 and Ser123 as acceptors (54% and 16%, respectively), while Ser123 also acted as a donor (12%). At the ᵗβGal-O4 position, Glu59 was identified as a predominant acceptor (88%), and Asn57 contributed as a donor (44%).

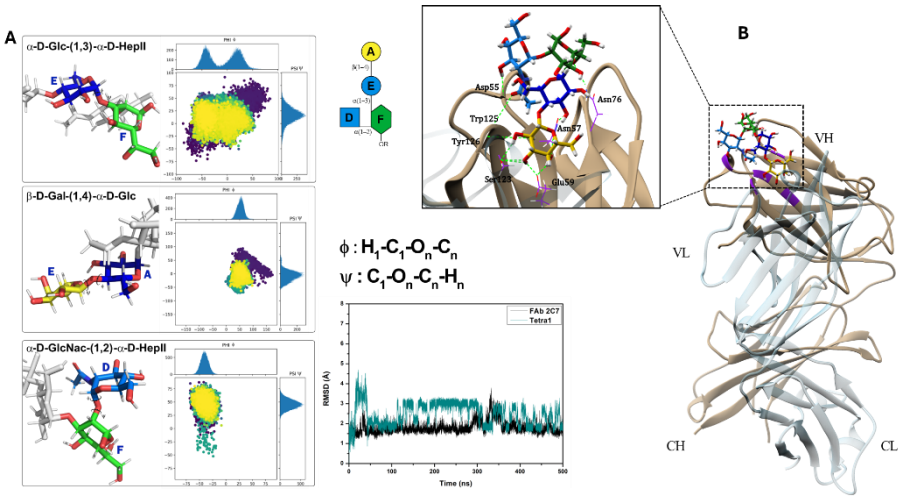


Figure VI.10 Integrated NMR and molecular dynamics analysis of the Tetra-1 oligosaccharide. (A) Adiabatic energy maps and RMSD illustrating the conformational landscape of OS glycosidic linkages, where highly populated regions are represented in yellow; RMSD plots from 500 ns molecular dynamics simulations showing the structural stability of both the ligand and the antibody. (B) Representative 3D model of the bioactive complex highlighting key hydrogen-bond interactions involving the termina βGal residue, which establishes recurrent contacts with Glu59, Ser123, Tyr126, and Asn57.

Energetics contributions. We employed isothermal titration calorimetry (ITC) to examine the energetic contribution at the basis of the interaction between mAb 2C7/OS and Tetra-1. The data were fitted using a one-site binding model to determine the thermodynamic parameters and the affinity constant. The binding stoichiometry was found to be 2.13, indicating that one mAb molecule interacted with two oligoes, in line with the two antigen-binding sites characteristic of an IgG antibody. The interaction between mAb 2C7 and OS was exothermic ($\Delta_bG° < 0$) and driven to a greater extent by the enthalpy contribution with a minor entropy contribution, with a negative enthalpy change ($\Delta_bH° = -27.3$ kJ/mol) and a favorable entropic term ($-T\,\Delta_bS° = -2.4$ kJ/mol). The dissociation constant (K_D) obtained was 6.35 μM, that indicates a strong binding affinity. (Figure 12A). In the same way, we had characterized the energetic profile of the interaction between mAb 2C7 and the Tetra-1 oligo. The binding event was found to be spontaneous ($\Delta_bG° < 0$) and enthalpy-driven, with $\Delta_bH° = -43.3$ kJ/mol (as reported in Table 9). These thermodynamic features are in line with an interaction dominated by hydrogen bonding and hydrophobic contacts at the binding interface. The resulting dissociation

constant (K_D) of 2 µM supports a moderate-to-strong affinity between mAb 2C7 and the synthetic oligo (Figure 12B).

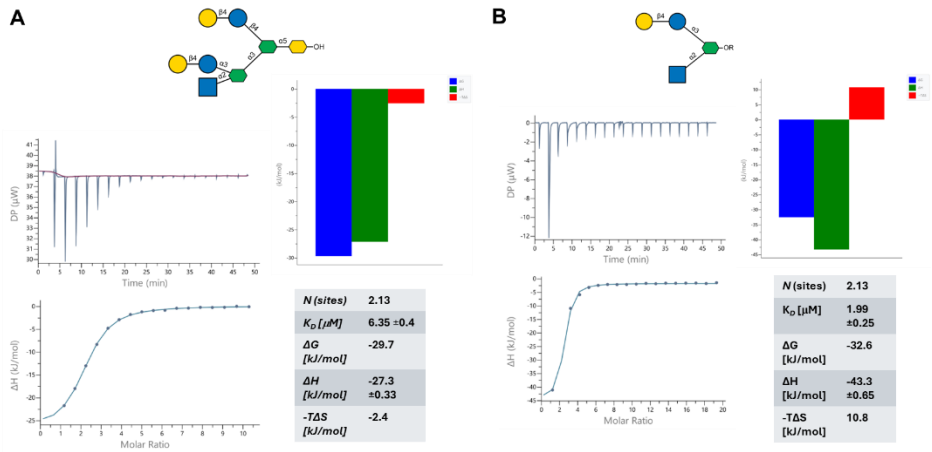


Figure VI.11 Isothermal titration calorimetry (ITC) analysis. Representative ITC thermograms and corresponding fitted binding isotherms obtained using a single-site binding model. The experiments reveal comparable stoichiometry (N ≈ 2) and favorable enthalpy-driven binding for both mAb2C7–OS (A) and mAb2C7–Tetra-1 (B) complexes. The thermodynamic parameters ($\Delta_bH°$, $\Delta_bS°$, and K_D) derived from curve fitting are reported in the accompanying tables, confirming strong and specific interactions between mAb2C7 and the carbohydrate epitopes.

6.3 Discussion

Neisseria gonorrhoeae infections continue to represent a serious public health threat, characterized by a persistent upward trend in annual cases. Monoclonal antibodies (mAbs) have emerged as a promising alternative to antibiotic therapies for the management of infectious diseases, exhibiting several advantages over conventional treatments. These include enhanced specificity, the capacity to minimize side effects, and the ability to selectively bind to the target antigen.

The conserved nature of the 2C7 epitope within the gonococcal LOS makes it a particularly promising target for antibody-based therapies. It is also important to consider that previous studies have shown that the 2C7 epitope, consisting of Lac-β-(1→4)-αHepI and Lac-α-(1→3)-HepII residues, is present in more than 95% of clinical isolates[17]. This high prevalence highlights the potential of mAb 2C7 to provide broad protection against diverse gonococcal strains.

Here, we adopted a multidisciplinary approach to comprehensively analyze the molecular interaction between the *Neisseria gonorrhoeae* 15253 core oligosaccharide (OS) and mAb 2C7. The integration of chemical, structural, and computational methodologies has provided critical insights into the structural and energetic determinants of this interaction.

The conservation of the 2C7 epitope in all LOS remarks its essential role in bacterial survival and immune evasion[33,34].

Western blot analysis of OS conjugated with CMP protein demonstrated a strong interaction with mAb 2C7, confirming the high specificity of the antibody for the conserved 2C7 epitope. NMR analysis yielded a detailed map of the OS binding epitope, and unveiled a well-defined recognition pattern. Significantly, the β-chain was identified as the predominant contributor to the binding, with the terminal βGal residue exhibiting the most pronounced STD NMR enhancements. In contrast, the α and γ-chain were found to establish least contribution

In this study, we delineated the molecular framework that governs the recognition of a conserved epitope exposed to the outer membrane of gram-negative pathogenic bacteria by antigen-specific immunoglobulins. The functional implications of these findings extend beyond the context of therapeutic development. The structural insights gained could provide important information on the design of mimetic compounds or conjugate vaccines that exploit the conserved features of the 2C7 epitope.

The conserved epitope 2C7 is a promising target for vaccine development, especially considering the limited success of traditional vaccine antigens. In contrast to the variable nature of traditional vaccine antigens, the conserved nature of the 2C7 epitope suggests its potential as a highly conserved and effective target for eliciting a protective immune response against common pathogenic strains of the *Neisseria* genus.

6.4 References

1. Organization, W.H., Antimicrobial resistance. 2023.

2. Sabtu, N.E., D. A.; Brown, N. M., Antibiotic resistance: what, why, where, when and how? . British Medical Bulletin 2015. 111: p. 105–116.

3. Uddin, T.M.C., A. J.; Khusro, A.; et al. , Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. . Journal of Infection and Public Health 2021. 14 (5): p. 601-609.

4. Stohl, E.A., et al., Neisseria gonorrhoeae Metalloprotease NGO1686 Is Required for Full Piliation, and Piliation Is Required for Resistance to H2O2- and Neutrophil-Mediated Killing. 2013. 4(4): p. 10.1128/mbio.00399-13.

5. Jerse, A.E., et al., Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. J Exp Med, 1994. 179(3): p. 911-20.

6. Cahoon, L.A.a.H.S.S., Focusing homologous recombination: pilin antigenic variation in the pathogenic Neisseria. . Mol Microbiol, 2011. 81(5): p. 1136-1143. .

7. Liu Y, F.B., Russell MW., New concepts in immunity to Neisseria gonorrhoeae: innate responses and suppression of adaptive immunity favor the pathogen, not the host. Front Microbiol., 2011. 2: p. 52.

8. Gibson, B.W., et al., Investigation of the structural heterogeneity of lipooligosaccharides from pathogenic Haemophilus and Neisseria species and of R-type lipopolysaccharides from Salmonella typhimurium by electrospray mass spectrometry. J Bacteriol, 1993. 175(9): p. 2702-12.

9. John, C.M., et al., The structural basis for pyocin resistance in Neisseria gonorrhoeae lipooligosaccharides. Journal of Biological Chemistry, 1991. 266(29): p. 19303-19311.

10. Gulati Sunita , S.J., Ram Sanjay , Rice Peter A. , Targeting Lipooligosaccharide (LOS) for a Gonococcal Vaccine. Frontiers in Immunology 2019. 10: p. 321.

11. Braun, D.C. and D.C. Stein, The lgtABCDE Gene Cluster, Involved in Lipooligosaccharide Biosynthesis in Neisseria gonorrhoeae, Contains Multiple Promoter Sequences. 2004. 186(4): p. 1038-1049.

12. Williams E, S.K., Fairley CK, Pollock GL, Hocking JS, McCarthy JS, Williamson DA., Neisseria gonorrhoeae vaccines: a contemporary overview. Clin Microbiol Rev, 2024. 37: p. 1.

13. Densen P, G.S., Rice PA Specificity of antibodies against Neisseria gonorrhoeae that stimulate neutrophil chemotaxis. Role of antibodies directed against lipooligosaccharides. The Journal of clinical investigation 1987. 80: p. 78–87.

14. Gulati, S., et al., Complement alone drives efficacy of a chimeric antigonococcal monoclonal antibody. PLOS Biology, 2019. 17(6): p. e3000323.
15. Chakraborti, S., et al., Bypassing Phase Variation of Lipooligosaccharide (LOS): Using Heptose 1 Glycan Mutants To Establish Widespread Efficacy of Gonococcal Anti-LOS Monoclonal Antibody 2C7. Infection and immunity, 2020. 88(2).
16. Gulati S, M.D.P., Mandrell R E, Jani D B, Rice P A. , Immunogenicity of Neisseria gonorrhoeae lipooligosaccharide epitope 2C7, widely expressed in vivo with no immunochemical similarity to human glycosphingolipids. . J. Infect. Dis. , 1996. 174: p. 1223-37.
17. Gulati, S., et al., Experimental Immunization with a Monoclonal Anti-Idiotope Antibody that Mimics the Neisseria gonorrhoeae Lipooligosaccharide Epitope 2C7. The Journal of Infectious Diseases, 1996. 174(6): p. 1238-1248.
18. Gulati, S., et al., Immunization against a Saccharide Epitope Accelerates Clearance of Experimental Gonococcal Infection. PLOS Pathogens, 2013. 9(8): p. e1003559.
19. Ram, S., et al., A Novel Sialylation Site on Neisseria gonorrhoeae Lipooligosaccharide Links Heptose II Lactose Expression with Pathogenicity. Infection and Immunity, 2018. 86(8): p. 10.1128/iai.00285-18.
20. Westphal, O.J.M.C.C., Bacterial lipopolysaccharides: extraction with phenol-water and further applications of the procedure. 1965. 5: p. 83.
21. Kittelberger, R., F.J.J.o.b. Hilbink, and b. methods, Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. 1993. 26(1): p. 81-86.
22. Leontein, K., B. Lindberg, and J.J.C.R. Lōnngren, Assignment of absolute configuration of sugars by glc of their acetylated glycosides formed from chiral alcohols. 1978. 62(2): p. 359-362.
23. Yamasaki, R., et al., Structural and immunochemical characterization of a Neisseria gonorrhoeae epitope defined by a monoclonal antibody 2C7; the antibody recognizes a conserved epitope on specific Lipo-oligosaccharides in spite of the presence of human carbohydrate epitopes. 1999. 274(51): p. 36550-36558.
24. Di Carluccio, C., et al., Investigation of protein-ligand complexes by ligand-based NMR methods. 2021. 503: p. 108313.

25. Case, D.A.B.-S., I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden , T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; SalomonFerrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, M.; Kollman, P. A., AMBER 2018 University of California, San Francisco. 2018.

26. Beernink, P.T., et al., Gonococcal Mimitope Vaccine Candidate Forms a Beta-Hairpin Turn and Binds Hydrophobically to a Therapeutic Monoclonal Antibody. 2024. 4(7): p. 2617-2629.

27. Ngan, C.H., et al., FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics, 2012. 28(2): p. 286-7.

28. Schrödinger, L. and W.J.S. DeLano, LLC, New York, NY, The PyMOL Molecular Graphics System, V2. 0.0. 2019.

29. Dam, T.K., et al., Isothermal titration calorimetry reveals differential binding thermodynamics of variable region-identical antibodies differing in constant region for a univalent ligand. 2008. 283(46): p. 31366-31370.

30. Unemo, M. and W.M.J.C.m.r. Shafer, Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. 2014. 27(3): p. 587-613.

31. Ventola, C.L.J.P. and therapeutics, The antibiotic resistance crisis: part 1: causes and threats. 2015. 40(4): p. 277.

32. Gulati, S., et al., Targeting lipooligosaccharide (LOS) for a gonococcal vaccine. Front Immunol 10: 321. 2019.

33. Apicella, M.A. and M.P. Jennings, Phase variation of bacterial surface glycosylated molecules in immune evasion, in Microbial Glycobiology. 2010, Elsevier. p. 837-846.

34. Varki, A.J.G., Biological roles of glycans. 2017. 27(1): p. 3-49.

35. Mandrell, R.E. and M.A.J.I. Apicella, Lipo-oligosaccharides (LOS) of mucosal pathogens: molecular mimicry and host-modification of LOS. 1993. 187(3-5): p. 382-402.

36. Dingjan, T., et al., Structural biology of antibody recognition of carbohydrate epitopes and potential uses for targeted cancer immunotherapies. Mol Immunol, 2015. 67(2 Pt A): p. 75-88.
37. Gómez-Redondo, M., et al., Bacterial polysaccharides: conformation, dynamics and molecular recognition by antibodies. 2020. 35: p. 1-11.
38. Lutsyk, V., et al., The Conformation of Glycosidic Linkages According to Various Force Fields: Monte Carlo Modeling of Polysaccharides Based on Extrapolation of Short-Chain Properties. 2024. 20(14): p. 6350-6368.
39. Woods, R.A., GA: University of Georgia, GLYCAM Web; Complex Carbohydrate Research Center. 2005.
40. Schrödinger, L.S.I.N.Y., NY, USA, Schrödinger Release 2022-3: LigPrep. 2021.
41. Roe, D.R., T.E.J.J.o.c.t. Cheatham III, and computation, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. 2013. 9(7): p. 3084-3095.
42. Humphrey, W., A. Dalke, and K.J.J.o.m.g. Schulten, VMD: visual molecular dynamics. 1996. 14(1): p. 33-38.
43. Pettersen, E.F., et al., UCSF Chimera—a visualization system for exploratory research and analysis. 2004. 25(13): p. 1605-1612.
44. Du, X., et al., Insights into protein–ligand interactions: mechanisms, models, and methods. 2016. 17(2): p. 144.

VII. General conclusions

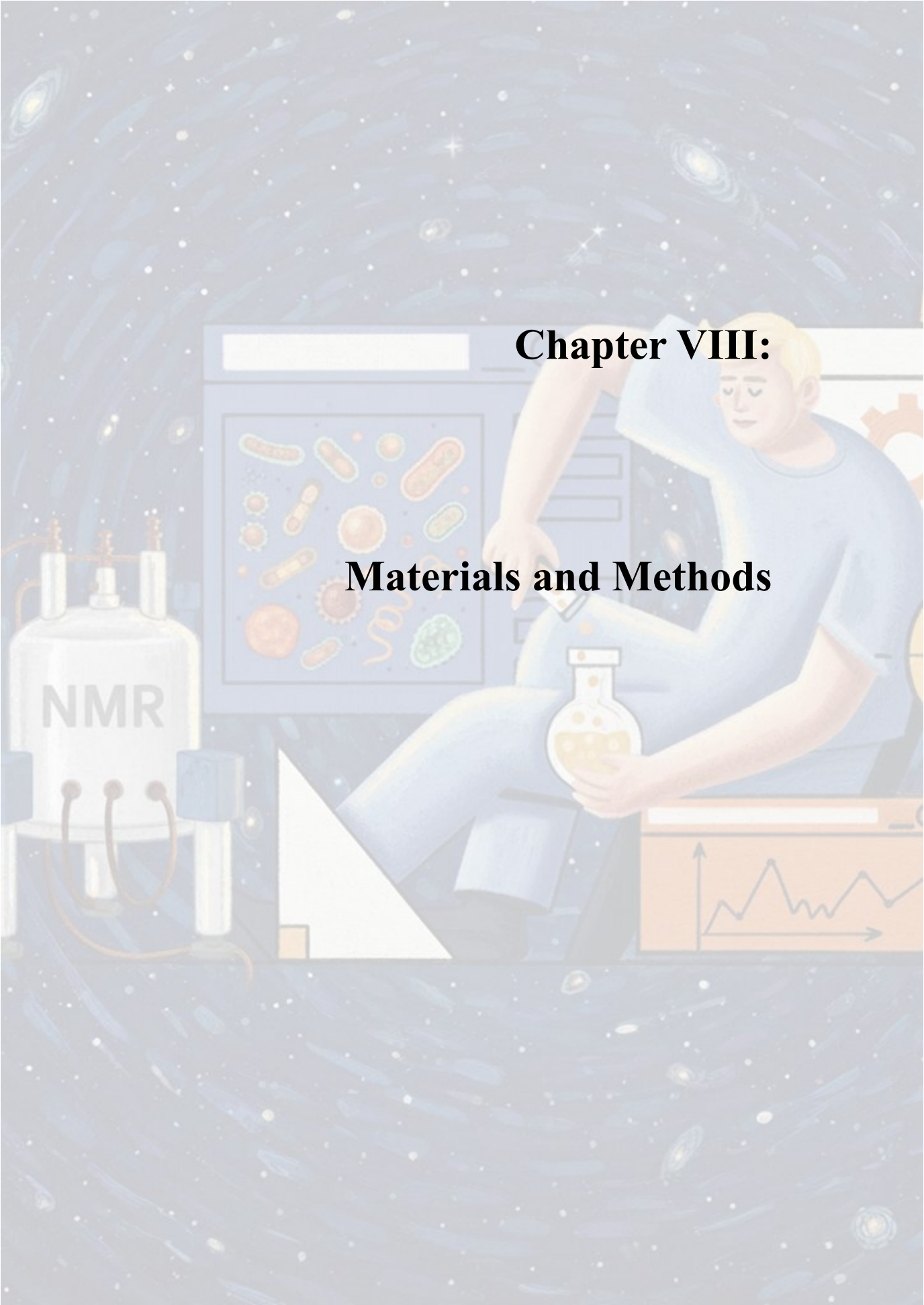
Glycans are fundamental components of all living systems, acting not only as structural elements of cell membranes but also as dynamic mediators of molecular communication. Their biological relevance lies in their capacity to participate in cell–cell recognition, signal modulation, immune regulation, and pathogen–host interactions. Among them, sialic acids occupy a distinctive position as terminal residues of glycan chains, functioning as molecular signatures of "*self*" and serving as crucial regulators of immune homeostasis through interaction with sialic acid–binding proteins (e.g. Siglecs). Once considered exclusively markers of endogenous "*self*", sialic acids have progressively emerged as molecular targets for mimicry strategies exploited by pathogens and tumor cells to evade immune detection. This duality—*self* and *non-self*, tolerance and subversion—defines the central role of Siglecs as molecular checkpoints in health and disease. Within this conceptual framework, the present thesis provides a comprehensive investigation of Siglec–sialoglycan interactions across different biological contexts, combining experimental and computational approaches to dissect the molecular principles governing specificity and affinity. Using Siglec-7 as a model system, this work explored its binding landscape toward both endogenous and exogenous sialylated ligands, elucidating how variations in linkage type, degree of sialylation, sulfation, and conformational flexibility modulate the receptor's recognition profile. Fluorescence spectroscopy, native mass spectrometry, NMR-based methods, and computational biology together revealed that Siglec-7 affinity is the product of a delicate balance between enthalpic and entropic contributions, largely driven by hydration and electrostatic rearrangements at the binding interface. This integrated perspective demonstrated that α2,6-linked glycans and branched, disialylated structures enhance stronger and more entropically driven binding, while α2,3-linked glycans tend to follow an enthalpy-dominated regime. Such findings underline that subtle changes in glycan topology can profoundly influence the energetic nature of recognition. On the endogenous side, the characterization of O-glycans, and ganglioside-derived

ligands highlighted the molecular logic by which Siglec-7 distinguishes self-associated sialylated epitopes, clarifying the structural features that enable specific, reversible binding under physiological conditions. Conversely, the investigation of bacterial LPS and CPS from *Neisseria meningitidis* and *Fusobacterium nucleatum* uncovered how pathogens exploit sialic acid–containing motifs to mimic host glycans and engage inhibitory Siglecs, thereby dampening immune activation. The case of *F. nucleatum* O-antigen, in particular, revealed a unique binding mode in which internal sialic acid residues establish electrostatic and polar contacts within the Siglec-7 binding pocket, promoting immune modulation through receptor engagement. These findings strengthen the notion that molecular mimicry represents an evolutionary strategy for immune escape, directly mediated by the structural versatility of bacterial glycans. The combination of biophysical and computational tools employed in this work has proven essential to capture both the static and dynamic aspects of recognition. The high level of consistency among these independent methodologies attests to the robustness of the multidisciplinary approach and validates computational modeling as a predictive and interpretative complement to experimental data.

Altogether, the results presented in this thesis refine the current view of Siglec-7 as a dynamic receptor capable of integrating structural and energetic cues to discriminate between endogenous and exogenous sialoglycans. This fine-tuned recognition process, dictated by both direct contacts and solvent-mediated effects, provides a molecular explanation for how Siglec-7 can simultaneously maintain self-tolerance and sense pathogenic signals. Beyond their biological significance, these insights hold practical value for translational research: the knowledge gained on binding thermodynamics, ligand topology, and water-mediated interactions can guide the rational design of synthetic glycomimetics, inhibitors, and antibody conjugates aimed at modulating Siglec-mediated immune checkpoints.

Among the exogenous ligands investigated, *Neisseria gonorrhoeae* represents a critical case study due to its increasing antibiotic resistance and urgent clinical

relevance. The conserved 2C7 epitope within its LOS, present in over 95% of clinical isolates, makes it an attractive target for antibody-based therapeutics. This work elucidated the molecular basis of the interaction between the LOS oligosaccharide and monoclonal antibody 2C7, revealing a stable and specific recognition pattern driven primarily by β-chain of the epitope. These findings provide a structural framework for understanding immune recognition of *N. gonorrhoeae* and highlight the potential of the 2C7 epitope as a template for the rational design of glycan-mimetic vaccines and targeted immunotherapies against resistant gonococcal strains.



Chapter VIII:

Materials and Methods

8.1 Materials and methods related to Chapter III

Expression and Purification of the Full Extracellular Domain (FED) of Siglec-7 in HEK293 Cells:

The Siglec-7 FED (Gly18–Gly354) with a C-terminal histidine tag was expressed in suspension-adapted human HEK293S GnTI− cells, as previously indicated[44]. Briefly, Siglec-7 FED was produced as a glycoprotein containing oligo-mannose-type glycans to avoid interferences from sialic acids during the binding with the ligands. ExCELL 293 serum-free cell culture medium (Merck) was used for cell line cultivation and transfection. The protein was purified from the cell culture supernatant by immobilized metal affinity chromatography (IMAC) followed by size-exclusion chromatography (SEC). The glycosylated Siglec-7 FED was used for the ligand-based NMR binding studies.

Expression and Purification of the Carbohydrate Recognition Domain (CRD) of Siglec-7 in E. coli Cells: The carbohydrate recognition domain (CRD) of Siglec-7 was cloned into the pET29b(+) vector between NdeI and XhoI restriction sites. The vector sequence was designed as previously published[45]. In brief, the coding region containing the Gly18–His148 sequence from the full-length mature protein was codon-optimized by Twist Bioscience. An additional modification to the lectin domain was performed by replacing Cys41 with a serine residue to avoid unspecific disulfide bridge formation.

The Siglec-7 CRD-carrying plasmid was transformed into BL21(DE3) Codon Plus RIPL or Rosetta 2 cells with the specific antibiotic selection and cultured at 37 °C, 180 rpm, in either LB or M9 culture medium, correspondingly supplemented with 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 0.2% v/v solution Q, 1 μg mL^{-1} thiamine, 1 μg mL^{-1} biotin, 1 g $(^{15}NH_4)_2SO_4$, and 3 g [U-^{13}C] glucose for isotopic labeling. The overexpression was induced at OD_{600} = 0.8 with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and incubated for 24 h at 25 °C under 180 rpm. The cells were harvested at 7500 rpm for 20 min using a Sorvall

RC 6 Plus Superspeed Centrifuge (Thermo Scientific), collected, and resuspended in lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and subsequently sonicated for 10 cycles (30 s ON, 2.5 min OFF) at 70% amplitude using a Vibra-Cell sonicator (Sonics & Materials). The lysate was ultracentrifuged for 40 min at 40 000 rpm with an Optima LE-80K Ultracentrifuge (Beckman).

Siglec-7 CRD was found incorporated into inclusion bodies. While the supernatant was discarded, the inclusion bodies were resuspended in 8 M urea lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and re-sonicated. The solubilization step was repeated to improve recovery yield. The soluble, unfolded fraction containing Siglec-7 CRD was further subjected to IMAC purification on a HisTrap FF (5 mL, GE Healthcare). The column was equilibrated with 20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole (pH 7.4), and the protein was loaded at 3 mL min^{-1} in closed cycle for 2 h. Elution was performed at 4 mL min^{-1} with high-concentration imidazole buffer (20 mM potassium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4) (Figure 1A,B).

To refold the unfolded protein, a series of dialysis steps were carried out. The protein was placed in a 10 kDa cut-off membrane, and each dialysis was performed overnight in 2 L of buffer:1st dialysis 20 mM potassium phosphate, 300 mM NaCl, 4 M urea;2nd dialysis 20 mM potassium phosphate, 300 mM NaCl, 2 M urea; 3rd dialysis – 20 mM potassium phosphate, 150 mM NaCl.

Refolded Siglec-7 CRD was collected and ultracentrifuged at 40 000 rpm for 40 min to remove aggregates, then filtered through a 0.2 μm micropore membrane. Final purification was achieved by size-exclusion chromatography on a HiLoad 26/60 Superdex 75 pg (GE Healthcare) column coupled to an AKTA Go FPLC system, equilibrated with 20 mM potassium phosphate, 50 mM NaCl, pH 7.4 (Figure 1C,D). The non-glycosylated Siglec-7 CRD was used for protein-based NMR binding studies.

Backbone Resonance Assignment of Siglec-7:

NMR protein experiments for backbone assignment were acquired on [U-^{15}N-^{13}C] Siglec-7 at 400 μM in 600 μL of aqueous buffer (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM protease inhibitors, 10% D_2O) in a 5 mm NMR tube. Triple resonance experiments (HNCA, HNCACB, HNCO, CBCAcoNH) were recorded at 298 K on a Bruker Avance™ NEO 900 MHz spectrometer equipped with a TCI cryo-probe. A 3D HNcaCO spectrum was recorded at 298 K on a Bruker Avance™ 500 MHz instrument. 93% of the amino acid sequence (from Y26 to T147) was assigned, excluding the five proline residues. Data were processed in TOPSPIN 4.1.1 and analyzed using CARA (Computer-Aided Resonance Assignment) software[46].

Protein-Based NMR Experiments:

For ligand-binding studies, 2D ^{1}H–^{15}N HSQC experiments were recorded on 200 μM [U-^{15}N] Siglec-7 CRD in 200 μL buffer (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM protease inhibitors, 10% D_2O) in a 3 mm NMR tube. Spectra were acquired on a Bruker Avance™ NEO 900 MHz spectrometer with a triple resonance TCI cryo-probe. Each HSQC used 32 scans, 2048 points in t_2, 128 increments in t_1, a recycle delay of 1.2 s, and a temperature of 298 K.

Interactions with ligands GD3, DSGb3α3, and DSGb3α6 were monitored by titrating increasing ligand concentrations (12.5–3200 μM). A new 2D ^{1}H–^{15}N HSQC spectrum was acquired after each addition. Data were processed with TOPSPIN 4.1.1 and analyzed in CARA. Chemical shift perturbations (CSP) were evaluated according to the standard equation[47–48].

$$\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + \left(\frac{\Delta\delta_N}{5}\right)^2}$$

Ligand-Based NMR Experiments:

Saturation Transfer Difference (STD) and transferred-NOESY NMR experiments were recorded on a Bruker AVANCE NEO 600 MHz spectrometer equipped with a cryo-probe. Data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in phosphate saline

deuterated buffer (10 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 10 mM NaN_3, pH 7.4) at 298 K.[D4] propionic acid, sodium salt (TSP, 1%) was used as an internal reference. A protein concentration of 25 μM was used, and the protein:ligand molar ratio was 1:50 for all mixtures.

STD NMR experiments were acquired with a shaped pulse train for saturation on the f_2 channel, alternating between on- and off-resonance with a 20 ms spin-lock pulse applied to suppress protein signals. The acquisition was set with 65 k data points and 104 scans. Protein resonances were selectively irradiated by 40 Gauss pulses with a length of 50 ms, using off-resonance pulse frequency at 40 ppm and on-resonance pulses at 7.5 and 0 ppm. The STD NMR spectra were carried out using a saturation time of 2 s. STD NMR effects were determined using the formula $(I_0 / I_{sat}) / I_0$, where I_{sat} is the relative intensity of the STD NMR signal and I_0 is the peak intensity of an unsaturated reference spectrum (off-resonance). The strongest STD NMR response was set to 100%, while all other STD signals were normalized to this value to provide ligands' epitope maps[49].

2D ^1H–^1H NOESY experiments were acquired using data sets of 4096 × 800 points and 100–600 ms as mixing times. Proton–proton cross-relaxation rates were calculated by integration of cross-peaks normalized against the corresponding diagonal peak. ^1H–^1H distances were calculated using the following equation:

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

where r_{ij} is the unknown distance to be estimated, r_{ref} is the reference inter-proton distance, σ_ref is the cross-relaxation rate of the NOE cross-peak of interest, and σ_{ij} is the cross-relaxation rate of the reference[50].

Fluorescence Analysis:

Steady-state fluorescence spectra were collected using a Fluoromax-4 spectrophotometer (Horiba Scientific, Edison, NJ, USA). Emission spectra were recorded from 310 to 450 nm upon excitation at 295 nm. Slit widths were

set to 5 nm for both excitation and emission wavelengths. A quartz cuvette with a 1 cm pathlength and a volume of 0.2 mL was used. The starting solution of Siglec-7 CRD, prepared at a fixed concentration of 4 μM in PBS buffer (pH 7.4), was titrated by adding small aliquots of each ligand. Experiments were carried out at 10, 25, and 35 °C. For all ligands analyzed, fluorescence intensity was quenched. The binding curve was obtained by fitting the data using nonlinear regression with a 1:1 binding model according to Ribeiro et al.[51], plotting the ratio F/F_0 versus total ligand concentration [L] in μM.

To determine the enthalpy change of binding, the temperature dependence of the binding constants was evaluated. A plot of $\ln(K_b)$ versus 1/T gives a straight line whose slope is equal to $-\Delta_bH°/R$, allowing the determination of the enthalpy change of binding, assuming $\Delta_bH°$ is independent of temperature within the explored range[52].

Isothermal Titration Calorimetry:

The ITC experiment was performed using a Nano ITC-III microcalorimeter (TA Instruments, New Castle, DE, USA). A 500 μM solution of ligand 1 was placed in the calorimetry vessel, and a 62.5 μM solution of Siglec-7 CRD was injected (injection volume: 15 μL). Under these conditions, the injected protein was fully bound to the ligand in the calorimeter vessel, yielding a series of similar heat peaks. The heat of dilution was evaluated by performing injections into buffer alone. The temperature was set to 25 °C, and the stirring speed to 250 rpm. The $\Delta_bH°$ was calculated by integrating the peaks, subtracting the heat of dilution, and normalizing by the amount of injected protein[39].

MD Simulations:

Glycans were generated on the GLYCAM website[53], and MD simulations were performed using AMBER 18[28–54]. The prmtop and inpcrd files were generated with the tLEaP module. Force fields were GLYCAM06j-1 for carbohydrate parameters and protein.ff14SB for SLBR-N. Complexes were prepared for MD simulations by adding counterions to neutralize the system using the Leap module and placing them in octahedral boxes with explicit TIP3P water molecules. MD simulations were run using the CUDA implementation of

PMEMD in AMBER18[54]. Minimization steps were performed using Sander. The smooth particle mesh Ewald method described electrostatic interactions, applying periodic boundary conditions with a grid spacing of 1 Å. The system underwent initial annealing over 25 ps from 100 to 300 K, followed by 50 ps equilibration at 300 K. A total of 10 000 frames were collected. Using the k-means algorithm implemented in the ptraj module of AMBER18, the trajectories were clustered to obtain representative poses. MD simulations were visualized with VMD[55].

RedMat:

The analysis was carried out considering the STD NMR effects and the best poses of the complexes between Siglec-7 and ligands GD3 and DSGb3α3. Parameters were set as follows: NMR spectrometer frequency, 600 MHz; protein concentration, 25 μM; ligand concentration, 1000 μM; cut-off distance, 10 Å; and dissociation constants derived from fluorescence analysis. The quality of the complexes and the agreement between theoretical and experimental data were evaluated from the R-factor (R-NOE) values (0.2 and 0.3 for ligands GD3 and DSGb3α3, respectively), calculated using the following equation:

$$R - NOE = \sqrt{\frac{\sum STD^{exp,i} - STD^{calc,i}{}^2}{\sum STD^{exp,i}{}^2}}$$

where STD_exp,i is the experimental STD value for a given proton *i*, and STD_calc,i is the STD value simulated using the RedMat algorithm[38].

Statistical Analysis—NMR Spectroscopy:

Free induction decay (FID) and spectral editing (SED) data from 1D and 2D NMR experiments were processed using Bruker Topspin software as described in the experimental procedures. For all homonuclear experiments, the data matrix was zero-filled in both dimensions (4K × 4K), followed by resolution enhancement using a cosine-bell function prior to Fourier transformation.

Statistical Analysis—Fluorescence Analysis:

Fluorescence measurements were performed in triplicate for each data point across two independent experimental series. Data were fitted using nonlinear regression analysis as described in the experimental procedures.

Statistical Analysis—Molecular Dynamics (MD) Simulations:

MD simulation trajectories were analyzed by clustering to identify and analyze dominant conformational states. Clustering analysis was performed using the cpptraj module, applying a k-means algorithm based on RMSD of the selected atoms (typically the ligand) to classify conformational states. To ensure computational efficiency, sieve clustering was employed to select a representative subset of frames. The algorithm was iterated multiple times to optimize cluster assignments, enabling identification of dominant conformations, structural variability, and representative structures for further validation.

Statistical Analysis—RedMat:

The validity of 3D complex structures and agreement between theoretical and experimental data were assessed using the R-factor (R-NOE), calculated as described in the experimental section and implemented in the RedMat software.

8.1.1 Supporting information

Supporting information is available from the Wiley Online Library or from the author.

8.2 Matherials and Methods related to Chapter IV

Healthy subjects and cell line: Healthy subjects enrolled in this study were enrolled according to the protocol approved by IRCCS Pascale, Institutional Ethical Review Board (CE: Protocollo n. 4/21, 2021). RPMI 8226 cell line were obtained by IRCCS SYNLAB SDN Biobank (10.5334/ojb.26) and cultured in RPMI supplemented with 10% fetal bovine serum and 1% Glutamine. For the interaction study, 3×10^4 cells/well were seeded in a 96-well plate and incubated at 37 °C and 5% CO_2. After o/n incubation, different concentrations of LPS from Salmonella, Fusobacterium nucleatum and OPS from Fusobacterium nucleatum were added (from 0.39 µg/mL to 200 µg/mL) for 1 h at 37 °C. After two PBS washes, cells were labeled with anti-Siglec-7 antibody PE-conjugated (#12-5759-42, Invitrogen). After appropriate

wash, a minimum of 10 000 events were recorded at Cytoflex cytofluorimeter (Beckman Coulter). For the interaction analysis, the normalized (with respect to the untreated sample) mean fluorescence intensities of the anti-Siglec-7 antibody were reported with respect to the concentrations of LPS and OPS. Experiments were repeated three times with similar results. The K_D values were calculated using GraphPad 6 software (One-site Total equation). For the blood cell separation, T-lymphocytes, B-lymphocytes, and NK-cells were purified from five healthy subjects using Easy Sep™ kits (#17961, #19044, #17995, STEMCELL). For cytofluorimetric analyses, PBMCs were labeled using CD45-KO, CD3-APC700, CD14-PC5.5, CD19-PC7, CD56-APC750, $CD45^{-}56^{-}19^{-}3$ tetrachrome, and HLADR-PC5 antibodies (Beckman Coulter). Kaluza software (Beckman Coulter) was used for the determination of the percentage of gated cells according to the gating strategy.Cytokine evaluation: For the cytokine expression evaluation, 200 ng/mL of LPS from Salmonella, Fusobacterium *nucleatum*, and OPS from *Fusobacterium nucleatum* were added to 5×10^5 PBMC from three healthy donors (male, median age 32 years) for o/n incubation at 37 °C. Cells were harvested and cytokine expression levels were evaluated in the medium using a LEGENDplex Human Inflammation Panel (13-plex), according to the manufacturer's instructions (740118, Biolegend), using Cytoflex (Beckman Coulter).

Protein expression and purification: The full extracellular domain (FED) of Siglec-7 (Gly18–Gly354) containing a C-terminal histidine tag was expressed in suspension-adapted human HEK293S $GnTI^{-}$ cells, as previously published[19]. The carbohydrate recognition domain (CRD) of Siglec-7 (Gly18–His148) containing a C-terminal histidine tag was expressed in LB and M9 culture medium. The protein resulted in the inclusion bodies, which were resuspended in 8 M urea lysis buffer; then, the soluble protein was subjected to IMAC purification using a HisTrap FF. The protein was refolded and then purified by size-exclusion chromatography on a HiLoad 26/60 Superdex 75 pg (GE Healthcare) coupled to an AKTA Go FPLC system. Details related to expression and purification of Siglec-7 in both E. coli and HEK293 cells have been submitted elsewhere[19]. HNCA, HNCACB, HNCO, and CBCAcoNH triple resonance experiments were recorded for the Siglec-7 CRD NMR

assignment at 298 K on a Bruker Avance NEO 900 MHz spectrometer equipped with a TCI cryo-probe. A 3D HNcaCO spectrum was recorded at 298 K on a Bruker Avance 500 MHz spectrometer[27]. The amino acid sequence from Y26 to T147 was assigned for 93% of residues, excluding the five prolines. Data acquisition and processing were performed with TOPSPIN 4.1.1 software, and spectra were analyzed using CARA (Computer-Aided Resonance Assignment) software[29].

Protein-based NMR experiments: Protein-based NMR titration of Siglec-7 with OPS was performed by recording 2D $^{1}H–^{15}N$ HSQC NMR experiments on an aqueous buffered solution of 200 µM [U–^{15}N] Siglec-7 CRD in 200 µL (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM protease inhibitors, 10% D_2O) in a 3 mm NMR tube. The spectra were recorded for the free protein and after the addition of increasing aliquots of OPS. A stock solution of 3.2 mg of OPS from F. nucleatum ATCC 10953 dissolved in 800 µL of Milli-Q water was prepared, and the solution was aliquoted into different fractions to reach 0.075, 0.075, 0.45, 0.6, 1.2, and 1.1 mg after lyophilization. Each of the lyophilized aliquots was subsequently redissolved in 200 µL of [U–^{15}N] Siglec-7 CRD (200 µM, or the previous titration point) to reach final ligand concentrations of 25, 50, 200, 400, 800, and 1160 µM, assuming an average molecular weight of 15 000 Da for the OPS. Details are shown in Table S1. Experiments were acquired on a Bruker Avance NEO 900 MHz spectrometer equipped with a triple resonance TCI cryo-probe. Spectra were acquired using 32 scans, 2048 data points in the direct dimension, 128 points in the indirect dimension, a recycle delay of 1.2 s, and a constant temperature of 298 K. Data acquisition and processing were performed with TOPSPIN 4.1.1 software, and spectra were analyzed using CARA[29].Chemical shift perturbations (CSP) were evaluated using the standard equation:[30]

$$\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + \left(\frac{\Delta\delta_N}{5}\right)^2}$$

All NMR experiments were recorded at 298 K on a Bruker AVANCE NEO 600 MHz spectrometer equipped with a cryo-probe. Data acquisition and processing were performed using TOPSPIN 4.1.1

software. Samples were prepared in phosphate-buffered saline (10 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 10 mM NaN_3, pH 7.4) at 298 K, using [D_4] propionic acid sodium salt (TSP, 0.05 mM) as the internal reference. The NMR experiments were conducted with a protein:ligand ratio of 1:50. Saturation Transfer Difference (STD) NMR experiments: STD NMR spectra were acquired using 32k data points, zero-filled to 64k before processing. Protein resonances were selectively irradiated with 40 G pulses lasting 50 ms, with the off-resonance pulse frequency set at 40 ppm and the on-resonance pulses at 0 ppm. To suppress protein signals, a 20 ms spin-lock pulse was applied. The acquisition parameters included 65k data points and 112 scans. STD NMR experiments were performed at a saturation time of 2 s. Ligand epitope mapping was achieved by calculating the ratio $(I_0 - I_{sat})/I_0$, where I_{sat} corresponds to the intensity of the STD NMR signal and I_0 to the intensity of the off-resonance spectrum. The strongest STD response was normalized to 100%, and all other proton STD signals were scaled accordingly. Control experiments were also performed in the absence of the protein to ensure the specificity of the observed STD signals. Transferred NOESY (tr-NOESY) experiments: Transferred NOESY spectra were recorded with data sets of 2048 × 512 points to analyze interproton distances within the ligand, with mixing times ranging from 100 to 400 ms. Carr–Purcell–Meiboom–Gill (CPMG) NMR experiments: CPMG experiments were recorded to measure T_2 spin–spin relaxation times of the ligand in the absence and presence of the protein. A pseudo-CPMG 2D sequence with water suppression using excitation sculpting with gradients was used. A fixed echo time of 3 ms and a recycle delay of 2 s were applied. T_2 signals were analyzed with Dynamic Center 2.7.1 software by fitting the equation $f(t) = I \exp(-t/T)$.

Parametrization: The non-parametrized AAT unit was parametrized using a custom protocol with Gaussian 09[31], performing a restrained electrostatic potential (RESP) charge calculation with a Hartree–Fock method and a 6-31G* basis set. The .prep and .frcmod files were generated by combining Antechamber and xLeap[32]. Trajectory analysis was conducted using the ptraj module in AMBER 18[33], and the molecular dynamics (MD) results were visualized with the VMD program[33]. Molecular mechanics calculations were performed using the MM3*

force field available in the MacroModel 8.0 software from the Maestro package. A dielectric constant of 80 was applied to simulate vacuum conditions, and the disaccharide structures were explored by incrementally varying the Φ and Ψ angles with an 18° grid step. Each (Φ, Ψ) point was optimized using 2000 conjugate gradient steps.

Molecular Dynamic simulation: MD simulations were conducted with the AMBER 18 suite, employing specific force fields: ff14SB for the protein[35], GLYCAM06j-1 for the saccharide portion of the ligands[36], and TIP3P for the water molecules. A GLYCAM-adapted force field was prepared for the AAT unit. Proteins were prepared using the Maestro Protein Preparation Wizard, which added missing hydrogens, adjusted protonation states of ionizable groups, and capped the termini. Systems were hydrated with a truncated-octahedral box of TIP3P water, with a 15 Å buffer, and counterions were added to neutralize the system. Input files were generated using the tleap module of AMBER 18. Initial energy minimization was performed using the sander module, while MD simulations were carried out with the pmemd module. Periodic boundary conditions and the particle mesh Ewald method were applied to represent electrostatic interactions, using a grid spacing of 1 Å. The initial minimization was conducted with the complex fixed, followed by minimization of the entire system. Subsequently, the system was gradually heated from 0 to 300 K, starting at constant volume and then transitioning to an isobaric ensemble. The temperature was maintained at 300 K for 50 ps with progressive energy minimizations and solute restraints. After equilibration, restraints were removed, and the simulations proceeded in an isothermal–isobaric ensemble during the production phase. System coordinates were saved every 2 ps, generating a set of 10 000 structures per complex for further analysis. Trajectories were analyzed using the ptraj module in AMBER 18, and the molecular dynamics results were visualized with VMD. Cluster analysis based on ligand RMSD was performed using the k-means algorithm in the ptraj module. The representative structure of the most populated cluster was used to illustrate the complex interactions. Hydrogen bond determination was carried out with the CPPTRAJ module[37–38], defining a hydrogen bond as having a maximum distance of 3 Å between donor and acceptor atoms and a minimum A–H–D angle of 135°. 3D images were prepared

using PyMOL, and dihedral conformation analysis was performed using a custom script that generated histograms of the most populated values during the simulation.

8.2.1 Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.5c00810.

8.3 Matherials and Methods related to Chapter V

Native mass spectrometry. Native mass spectrometry (nMS) experiments were performed on a **timsTOF SCP** instrument (Bruker Daltonics, Bremen, Germany), a trapped ion mobility mass spectrometer equipped with electrospray (ESI) and nano-electrospray (nESI) ionization sources. The protein sample was dissolved in ultrapure water (UPLC-MS grade, Biosolve). Desalting was achieved using Amicon Ultra-0.5 centrifugal filters (Merck Millipore Ltd.) with a 10 kDa molecular weight cut-off. Several washing steps were performed using 100 mM ammonium acetate (NH_4OAc), followed by pure water, to ensure complete removal of non-volatile salts. The final protein solution was prepared at a concentration of 50 µM in 50 mM NH_4OAc and subsequently diluted to 5 or 1 µM prior to injection. All spectra were recorded under **soft ionization conditions** to preserve non-covalent interactions and maintain the native state of the complex. Instrumental parameters were carefully optimized to ensure stable ion transmission and minimal gas-phase activation. The source temperature was set to 220 °C, and the capillary voltage was maintained at 1.2 kV in negative ion mode. Ion mobility separation was conducted within the trapped ion mobility (TIMS) region under helium buffer gas at controlled pressure. The trapping and release times were set to 1000 µs and 200 µs, respectively, with a trap entrance grid delta of 2 V.

Data analysis. Data were acquired and processed using **Bruker DataAnalysis** software packages. Quantitative evaluation of the binding interactions was performed by integrating the ion signal intensities corresponding to the free and bound protein species using MATLAB. For the equations used to determine the amount of bound ligand and association/dissociation constants, please refer to Chapter II.3.2.1.

8.4 Matherials and Methods related to Chapter VI

Isolation and purification of LOS from Neisseria gonorrhoeae strain 15253: Bacterial dried cells (5 g) were extracted using the hot phenol/water method[20]. The crude LPS (900 mg) was analyzed by SDS-PAGE with silver staining[19], confirming the presence of rough-type LPS exclusively in the water phase. To eliminate cellular contaminants, the sample underwent enzymatic treatments with RNase (Roth), DNase (Roth), and Proteinase K (Roth) at 37 °C and 57 °C, followed by extensive dialysis. For SDS-PAGE analysis, the purified LOS was prepared at a concentration of 1 mg/mL, boiled for 10 minutes, and then aliquots of 4 µL and 8 µL were loaded into wells of a 13.5% SDS gel with a 5% stacking gel. Electrophoretic separation was performed using a Mini-PROTEAN system (Bio-Rad Laboratories). To ensure the LOS extract was free of contaminants, an aliquot of the previously treated LOS, processed to remove nucleic acids and proteins, dialyzed, centrifuged, and purified by chromatography, was subjected to additional washing steps. These washes involved chloroform–methanol (1:2, v/v) and chloroform–methanol–water (3:2:0.25, v/v) mixtures, commonly used to remove phospholipids. No unexpected safety hazards or unusual risks were encountered during this work. *Monosaccharide analysis of Neisseria gonorrhoeae LOS:* The monosaccharide composition of *Neisseria gonorrhoeae* LOS was determined by analyzing the acetylated O-methylglycoside derivatives. These derivatives were obtained through methanolysis using HCl/MeOH (1.25 M) at 100 °C, followed by acetylation with acetic anhydride in pyridine at 85 °C for 30 minutes. All chemical analyses were carried out using gas-liquid chromatography (GLC) with an Agilent Technologies 6850A system (Santa Clara, CA, USA), equipped with a 5973N mass-selective detector and a Zebron ZB-5 capillary column (Phenomenex, 30 m × 0.25 mm i.d., 0.25 µm film thickness). Helium was used as the carrier gas at a flow rate of 1 mL/min. The temperature program for sugar analysis was as follows: an initial temperature of 150 °C held for 5 minutes, followed by a ramp from 150 °C to 280 °C at 3 °C/min, and a final hold at 280 °C for 5 minutes.

Separation of core OS from lipid A: An aliquot (100 mg) of LOS sample was treated with acetic acid at 100 °C for 5 hours to facilitate the separation of the lipid A and core OS portions. The resulting solutions were centrifuged at 7000 × g for 30 minutes at 4 °C. The resulting water phases, containing OS, were collected. The organic phase was vortexed and further centrifuged by washing three times with distilled water. The water-soluble fractions, containing the OS, were collected separately and lyophilized. To remove any traces of lipid A and reaction products, the sample was purified by gel filtration chromatography using Bio-Gel TSK 40.

MALDI-ToF MS analysis: Linear and reflectron MALDI-TOF MS spectra of LOS mild hydrolysis product (OS) were recorded on an AB SCIEX TOF/TOF 5800 Applied Biosystems mass spectrometer equipped with an Nd:YAG laser (λ = 349 nm), with a 3-ns pulse width and a repetition rate of up to 1000 Hz. For sample preparation, an aliquot of OS was dissolved in distilled water and mixed with 2,4,6-trihydroxyacetophenone (THAP) in a solution of acetonitrile with 0.1% TFA (1:1). Each spectrum in the MS experiments was obtained by accumulating 2000 laser shots. Each experiment was performed in duplicate. *Isolation of the OSdeAc fraction:* To investigate the structure of the core oligosaccharide (OS) from the LOS of *Neisseria gonorrhoeae* strain 15253, a 40 mg aliquot of LOS was subjected to O-deacylation. This was achieved by treating the sample with anhydrous hydrazine (2 mL) at 37 °C for 90 minutes. The reaction mixture was then cooled, poured into ice-cold acetone, and left to precipitate. The precipitate was collected by centrifugation at 5000 × g for 30 minutes, washed with ice-cold acetone, dried, dissolved in water, and lyophilized. The O-deacylated product was further N-deacylated by treatment with 4 M KOH at 120 °C for 16 hours. To remove salts, the sample was purified by gel filtration chromatography using a Sephadex G-10 column (Pharmacia, 50 × 1.5 cm).

*NMR spectroscopy:*The NMR experiments were recorded on a Bruker AVANCE NEO 600 MHz spectrometer equipped with a cryo-probe, and data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in 50 mM phosphate buffer at pH 7.4,

using 3 mm NMR tubes. NMR binding experiments were run using an mAb 2C7 protein concentration of 30 μM. Assignment of the resonances of the core OS oligosaccharide in complex with mAb 2C7 (150:1) was obtained by the analysis of the COSY, TOCSY, and ROESY spectra using TOPSPIN software version 4.3.0. *STD NMR analysis:* STD NMR experiments were acquired with 32k data points and zero-filled up to 64k prior to processing. The mAb 2C7 protein resonances were selectively irradiated using 40 Gauss pulses of 50 ms length, setting the off-resonance pulse frequency at 40 ppm and the on-resonance pulse at 6.5 and 7.5 ppm. Excitation sculpting with gradient pulses (ESGP) was applied for the suppression of water signals. The %STD displayed in the ligand epitope maps was obtained by the ratio of the STD signals in the STD spectra ($I_0 - I_{sat}$) and each relative peak intensity of the unsaturated reference spectrum (off-resonance, I_0), at a saturation time of 2 s. The highest STD signal was set to 100%, and all other STD were normalized to this value. Homonuclear 2D ROESY experiments were conducted using data sets of 4096 × 900 points and mixing times of 100–400 ms. The inter-proton cross-relaxation rates (σ_{ij}) were measured by integrating the ROE cross peaks of interest and normalizing against the corresponding cross peak on the diagonal in F1. The experimental distances (r_{ij}) were then obtained using the isolated spin pair approximation.

Molecular Mechanics and Molecular Dynamics simulation: Molecular mechanics (MM) calculations were carried out using the MM3* force field in a vacuum environment, with a dielectric constant of 80. For the disaccharide structures, the glycosidic torsion angles (Φ and Ψ) were systematically varied in 18° increments. At each (Φ, Ψ) point on the map, optimization was performed using 2000 P.R. Molecular dynamics (MD) simulations were conducted with the AMBER 18 software package[25], employing explicit water molecules. The AMBER ff14SB, GLYCAM06j-1, and TIP3P force fields were used for the protein residues, saccharide ligand, and solvent, respectively. Monosaccharide structures were downloaded from the GLYCAM website (www.glycam.org)[39]. Protein preparation involved adding missing hydrogen atoms, determining the protonation states of ionizable groups, and capping termini using the Maestro Protein Preparation Wizard[40].

Each system was hydrated with an octahedral TIP3P water box, buffered to 15 Å, and neutralized with counterions. Input files were generated using the *tleap* module in AMBER. Minimization steps were carried out using the *sander* module, while MD simulations were performed with the *pmemd* module. The initial structure was refined through energy minimization, applying the SHAKE algorithm to constrain C–H bonds and using a 1 fs integration step. Periodic boundary conditions were applied, and electrostatic interactions were modeled with the smooth particle mesh Ewald method, using a grid spacing of 1 Å. Minimization was performed in two stages: first, the complex was restrained while the solvent was minimized, followed by minimization of the entire system. The system was then gradually heated from 0 K to 300 K. Heating was conducted in two steps: from 0 K to 100 K under constant volume and from 100 K to 300 K under isobaric conditions. Once the target temperature of 300 K was reached, the system was equilibrated over 50 ps with progressive energy minimization and solute restraints. After equilibration, restraints were removed, and production runs were performed under isothermal–isobaric conditions. Coordinates were saved throughout the 100 ns production simulations, with trajectory snapshots recorded every 2 ps, generating 10 000 structures per complex for analysis. Non-standard residues, including Kdo, HepI, GlcN, and HepII, were parametrized using a custom protocol involving Restrained Electrostatic Potential (RESP) charge calculations through Hartree–Fock with a 6–31G* basis set. The Antechamber and xLeap modules were used to generate .prep and .frcmod files[41]. Trajectories were analyzed with the *ptraj* module in AMBER 18, and the MD results were visualized using the VMD program[42]. Cluster analysis of the trajectories was performed based on the ligand RMSD, using the k-means algorithm implemented in *ptraj*. The representative structure of the most populated cluster was selected to characterize the interactions within the complexes. Hydrogen bonds were identified using the CPPTRAJ module in AMBER 18[41], with a hydrogen bond defined by a donor–acceptor distance cutoff of 3 Å and an A–H–D angle cutoff of 135°. 3D images of the complexes were generated using UCSF Chimera[43]. Dihedral angle conformations were analyzed with a custom script that depicted torsional

variations during the MD simulations and generated histograms showing the most populated torsion values.

Isothermal Titration Calorimetry: Isothermal titration calorimetry (ITC) experiments were carried out at 25 °C using a MicroCal PEAQ-ITC instrument. mAb and OS extracts were both dialyzed against PBS pH 7.4 to avoid buffer mismatch between solutions. A solution of mAb at 30 μM was loaded into the cell, and for each measurement, OS extract was added to the antibody solution through a micro-syringe under a stirrer speed of 750 rpm. The titration was performed by adding 2 μL from a 1.6 mM OS solution for 19 injections using a time duration of 150 s. The data were analyzed and fitted using a one-binding-site model. The standard thermodynamic parameters of equilibrium were calculated according to the equation[44]: $\Delta_b G° = -RT \ln K_a = \Delta_b H° - T \Delta_b S°$.

References

IX. References

1. Varki, A.; Gagneux, P., Biological functions of glycans. **2017**.
2. Oswald, D. M.; Cobb, B. A. J. C. i., Emerging glycobiology tools: a renaissance in accessibility. **2018,** *333*, 2-8.
3. Gaudette, S.; Hughes, D.; Boller, M. J. J. o. v. e.; care, c., The endothelial glycocalyx: Structure and function in health and critical illness. **2020,** *30* (2), 117-134.
4. Schauer, R. J. C. o. i. s. b., Sialic acids as regulators of molecular and cellular interactions. **2009,** *19* (5), 507-514.
5. Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. J. N. r. M. c. b., Global view of human protein glycosylation pathways and functions. **2020,** *21* (12), 729-749.
6. Yu, R. K.; Tsai, Y.-T.; Ariga, T. J. N. r., Functional roles of gangliosides in neurodevelopment: an overview of recent advances. **2012,** *37* (6), 1230-1244.
7. Stanley, P. J. C. S. H. p. i. b., Golgi glycosylation. **2011,** *3* (4), a005199.
8. Varki, A. J. N., Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. **2007,** *446* (7139), 1023-1029.
9. Brockhausen, I.; Schachter, H.; Stanley, P. J. E. o. G. n. e., O-GalNAc glycans. **2009**.
10. Tran, D. T.; Ten Hagen, K. G. J. J. o. B. C., Mucin-type O-glycosylation during development. **2013,** *288* (10), 6921-6929.
11. Schnaar, R. L. J. J. o. A.; Immunology, C., Glycans and glycan-binding proteins in immune regulation: a concise introduction to glycobiology for the allergist. **2015,** *135* (3), 609-615.
12. Varki, A.; Schauer, R. J. E. o. g., 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Chapter 14. Sialic acids. **2009**.
13. Angata, T.; Varki, A. J. C. r., Chemical diversity in the sialic acids and related α-keto acids: an evolutionary perspective. **2002,** *102* (2), 439-470.
14. Hudak, J. E.; Canham, S. M.; Bertozzi, C. R. J. N. c. b., Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. **2014,** *10* (1), 69-75.
15. Stencel-Baerenwald, J. E.; Reiss, K.; Reiter, D. M.; Stehle, T.; Dermody, T. S. J. N. R. M., The sweet spot: defining virus–sialic acid interactions. **2014,** *12* (11), 739-749.
16. Büll, C.; den Brok, M. H.; Adema, G. J. J. B. e. B. A.-R. o. C., Sweet escape: sialic acids in tumor immune evasion. **2014,** *1846* (1), 238-246.
17. Varki, A. J. A. J. o. P. A. T. O. P. o. t. A. A. o. P. A., Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences, and implications for hominid evolution. **2001,** *116* (S33), 54-69.
18. Varki, A., Letter to the Glyco-Forum: 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.
19. Paulson, J. C.; Macauley, M. S.; Kawasaki, N. J. A. o. t. N. Y. A. o. S., Siglecs as sensors of self in innate and adaptive immune responses. **2012,** *1253* (1), 37-48.
20. Pomin, V. H.; Mulloy, B. J. P., Glycosaminoglycans and proteoglycans. MDPI: 2018; Vol. 11, p 27.
21. Mogensen, T. H. J. C. m. r., Pathogen recognition and inflammatory signaling in innate immune defenses. **2009,** *22* (2), 240-273.
22. Pinho, S. S.; Reis, C. A. J. N. R. C., Glycosylation in cancer: mechanisms and clinical implications. **2015,** *15* (9), 540-555.

23. Cress, B. F.; Englaender, J. A.; He, W.; Kasper, D.; Linhardt, R. J.; Koffas, M. A. J. F. m. r., Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. **2014,** *38* (4), 660-697.
24. Varki, A.; Etzler, M. E.; Cummings, R. D.; Esko, J. D., Discovery and classification of glycan-binding proteins. **2010**.
25. Rabinovich, G. A.; Van Kooyk, Y.; Cobb, B. A. J. A. o. t. N. Y. A. o. S., Glycobiology of immune responses. **2012,** *1253* (1), 1-15.
26. Varki, A. J. G., Biological roles of glycans. **2017,** *27* (1), 3-49.
27. Varki, A.; Angata, T. J. G., Siglecs—the major subfamily of I-type lectins. **2006,** *16* (1), 1R-27R.
28. Crocker, P. R. J. C. o. i. p., Siglecs in innate immunity. **2005,** *5* (4), 431-437.
29. Pillai, S.; Netravali, I. A.; Cariappa, A.; Mattoo, H. J. A. r. o. i., Siglecs and immune regulation. **2012,** *30*, 357-392.
30. Crocker, P. R.; Varki, A. J. I., Siglecs in the immune system. **2001,** *103* (2), 137.
31. Linnartz-Gerlach, B.; Mathews, M.; Neumann, H., Sensing the neuronal glycocalyx by glial sialic acid binding immunoglobulin-like lectins. *Neuroscience* **2014,** *275*, 113-24.
32. Angata, T.; Varki, A. J. M. a. o. m., Discovery, classification, evolution and diversity of Siglecs. **2023,** *90*, 101117.
33. Crocker, P. R.; Paulson, J. C.; Varki, A., Siglecs and their roles in the immune system. *Nature Reviews Immunology* **2007,** *7* (4), 255-266.
34. Boelaars, K.; van Kooyk, Y. J. T. i. c., Targeting myeloid cells for cancer immunotherapy: Siglec-7/9/10/15 and their ligands. **2024,** *10* (3), 230-241.
35. Smith, B. A.; Bertozzi, C. R. J. N. R. D. D., The clinical impact of glycobiology: targeting selectins, Siglecs and mammalian glycans. **2021,** *20* (3), 217-243.
36. Macauley, M. S.; Crocker, P. R.; Paulson, J. C. J. N. R. I., Siglec-mediated regulation of immune cell function in disease. **2014,** *14* (10), 653-666.
37. Choi, H.; Ho, M.; Adeniji, O. S.; Giron, L.; Bordoloi, D.; Kulkarni, A. J.; Puchalt, A. P.; Abdel-Mohsen, M.; Muthumani, K. J. F. i. O., Development of Siglec-9 blocking antibody to enhance anti-tumor immunity. **2021,** *11*, 778989.
38. Yang, D. J. T. S. O., The Potential of Siglec Receptors in Cancer Immunotherapy. **2024,** *1* (3), 76-89.
39. May, A.; Robinson, R.; Vinson, M.; Crocker, P.; Jones, E. J. M. c., Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 Å resolution. **1998,** *1* (5), 719-728.
40. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. J. J. o. t. A. C. S., Carbohydrate–aromatic interactions in proteins. **2015,** *137* (48), 15152-15160.
41. Ulyanova, T.; Shah, D. D.; Thomas, M. L. J. J. o. B. C., Molecular cloning of MIS, a myeloid inhibitory siglec, that binds protein-tyrosine phosphatases SHP-1 and SHP-2. **2001,** *276* (17), 14451-14458.
42. Attrill, H.; Imamura, A.; Sharma, R. S.; Kiso, M.; Crocker, P. R.; Van Aalten, D. M. J. J. o. B. C., Siglec-7 undergoes a major conformational change when complexed with the α (2, 8)-disialylganglioside GT1b. **2006,** *281* (43), 32774-32783.
43. Yu ZhenBao, Y. Z.; Maoui, M.; Wu LiangTang, W. L.; Banville, D.; Shen ShiHsiang, S. S., mSiglec-E, a novel mouse CD33-related siglec (sialic acid-binding immunoglobulin-like lectin) that recruits Src homology 2 (SH2)-domain-containing protein tyrosine phosphatases SHP-1 and SHP-2. **2001**.

44. Ishida-Kitagawa, N.; Tanaka, K.; Bao, X.; Kimura, T.; Miura, T.; Kitaoka, Y.; Hayashi, K.; Sato, M.; Maruoka, M.; Ogawa, T. J. J. o. B. C., Siglec-15 protein regulates formation of functional osteoclasts in concert with DNAX-activating protein of 12 kDa (DAP12). **2012,** *287* (21), 17493-17502.
45. Chen, G.-Y.; Brown, N. K.; Zheng, P.; Liu, Y. J. G., Siglec-G/10 in self–nonself discrimination of innate and adaptive immunity. **2014,** *24* (9), 800-806.
46. Avril, T.; Wagner, E. R.; Willison, H. J.; Crocker, P. R. J. I.; immunity, Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on Campylobacter jejuni lipooligosaccharides. **2006,** *74* (7), 4133-4141.
47. Carlin, A. F.; Uchiyama, S.; Chang, Y.-C.; Lewis, A. L.; Nizet, V.; Varki, A. J. B., The Journal of the American Society of Hematology, Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. **2009,** *113* (14), 3333-3336.
48. 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. J. F. i. I., Siglec-7 mediates immunomodulation by colorectal cancer-associated Fusobacterium nucleatum ssp. animalis. **2021,** *12*, 744184.
49. Khatua, B.; Roy, S.; Mandal, C. J. I. J. o. M. R., Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens. **2013,** *138* (5), 648-662.
50. Chang, Y.-C.; Nizet, V. J. G., The interplay between Siglecs and sialylated pathogens. **2014,** *24* (9), 818-825.
51. Kawasaki, Y.; Ito, A.; Kakoi, N.; Shimada, S.; Itoh, J.; Mitsuzuka, K.; Arai, Y. J. T. T. j. o. e. m., Ganglioside, disialosyl globopentaosylceramide (DSGb5), enhances the migration of renal cell carcinoma cells. **2015,** *236* (1), 1-7.
52. Jin, X.; Yang, G.-Y. J. P. i. L. R., Pathophysiological roles and applications of glycosphingolipids in the diagnosis and treatment of cancer diseases. **2023,** *91*, 101241.
53. Yamakawa, N.; Yasuda, Y.; Yoshimura, A.; Goshima, A.; Crocker, P. R.; Vergoten, G.; Nishiura, Y.; Takahashi, T.; Hanashima, S.; Matsumoto, K. J. S. r., Discovery of a new sialic acid binding region that regulates Siglec-7. **2020,** *10* (1), 8647.
54. Jasti, A. K.; Selmi, C.; Sarmiento-Monroy, J. C.; Vega, D. A.; Anaya, J.-M.; Gershwin, M. E. J. E. r. o. c. i., Guillain-Barré syndrome: causes, immunopathogenic mechanisms and treatment. **2016,** *12* (11), 1175-1189.
55. Schnaar, R. L. J. J. o. m. b., Gangliosides of the vertebrate nervous system. **2016,** *428* (16), 3325-3336.
56. Schnaar, R. L.; Gerardy-Schahn, R.; Hildebrandt, H. J. P. r., Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. **2014**.
57. Yu, R.; Tsai, Y.; Ariga, T.; Structures, M. Y. J. D. h. d. o. j., biosynthesis, and functions of gangliosides—An overview., 2011, 60. *60*, 537-544.
58. Svennerholm, L. J. J. o. n., Chromatographlc separation of human brain gangliosides. **1963,** *10* (9), 613-623.
59. Rahmann, H. J. B. B. R., Brain gangliosides and memory formation. **1995,** *66* (1-2), 105-116.
60. Wang, B. J. A. r. o. n., Sialic acid is an essential nutrient for brain development and cognition. **2009,** *29* (1), 177-222.
61. Fredman, P.; Lekman, A. J. N. r., Glycosphingolipids as potential diagnostic markers and/or antigens in neurological disorders. **1997,** *22* (8), 1071-1083.

62. Silipo, A.; Molinaro, A., The diversity of the core oligosaccharide in lipopolysaccharides. In *Endotoxins: Structure, Function and Recognition*, Springer: 2010; pp 69-99.
63. Raetz, C. R.; Whitfield, C. J. A. r. o. b., Lipopolysaccharide endotoxins. **2002,** *71* (1), 635-700.
64. Akira, S.; Uematsu, S.; Takeuchi, O. J. C., Pathogen recognition and innate immunity. **2006,** *124* (4), 783-801.
65. Welker, M. J. T. C. i. W., 2.8. 1 General characteristics of bacterial LPS. **2021**, 137.
66. Holst, O.; Molinaro, A.; Holst, O.; Brennan, P.; Itzstein, M. v.; Moran, A. J. M. G. A. P., San Diego, Chapter 3-Core region and lipid A components of lipopolysaccharides. **2010**, 29-55.
67. Erridge, C.; Bennett-Guerrero, E.; Poxton, I. R. J. M.; infection, Structure and function of lipopolysaccharides. **2002,** *4* (8), 837-851.
68. Preston, A.; Mandrell, R. E.; Gibson, B. W.; Apicella, M. A. J. C. r. i. m., The lipooligosaccharides of pathogenic gram-negative bacteria. **1996,** *22* (3), 139-180.
69. Feldkamp, C. S.; Carey, J. L., 2 - IMMUNE FUNCTION AND ANTIBODY STRUCTURE. In *Immunoassay*, Diamandis, E. P.; Christopoulos, T. K., Eds. Academic Press: San Diego, 1996; pp 5-24.
70. Schroeder Jr, H. W.; Cavacini, L. J. J. o. a.; immunology, c., Structure and function of immunoglobulins. **2010,** *125* (2), S41-S52.
71. Xu, J. L.; Davis, M. M. J. I., Diversity in the CDR3 region of VH is sufficient for most antibody specificities. **2000,** *13* (1), 37-45.
72. Schroeder, H. W., Jr.; Cavacini, L., Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology* **2010,** *125* (2 Suppl 2), S41-52.
73. KÖHler, G.; Milstein, C., Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **1975,** *256* (5517), 495-497.
74. De Castro, C.; Parrilli, M.; Holst, O.; Molinaro, A., Chapter Five - Microbe-Associated Molecular Patterns in Innate Immunity: Extraction and Chemical Analysis of Gram-Negative Bacterial Lipopolysaccharides. In *Methods in enzymology*, Fukuda, M., Ed. Academic Press: 2010; Vol. 480, pp 89-115.
75. De Castro, C.; Parrilli, M.; Holst, O.; Molinaro, A. J. M. i. e., Microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. **2010,** *480*, 89-115.
76. De Castro, C. Mass Spectrometry for Glycosciences. Ed. Glycopedia https://glycopedia.eu/echapter/introduction-5/.
77. Viegas, A.; Macedo, A. L.; Cabrita, E. J., Ligand-Based Nuclear Magnetic Resonance Screening Techniques. In *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols*, Roque, A. C. A., Ed. Humana Press: Totowa, NJ, 2010; pp 81-100.
78. Di Carluccio, C.; Forgione, M. C.; Martini, S.; Berti, F.; Molinaro, A.; Marchetti, R.; Silipo, A. J. C. r., Investigation of protein-ligand complexes by ligand-based NMR methods. **2021,** *503*, 108313.
79. Mayer, M.; Meyer, B., Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *Journal of the American Chemical Society* **2001,** *123* (25), 6108-6117.
80. Angulo, J.; Enríquez-Navas, P. M.; Nieto, P. M., Ligand–Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. **2010,** *16* (26), 7803-7812.

81. Harvey, D. J. J. M. S. R., Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: an update for 2019–2020. **2023,** *42* (5), 1984-2206.
82. Tamara, S.; den Boer, M. A.; Heck, A. J. J. C. r., High-resolution native mass spectrometry. **2021,** *122* (8), 7269-7326.
83. Gabelica, V., Determination of Equilibrium Association Constants of Ligand–DNA Complexes by Electrospray Mass Spectrometry. In *Drug-DNA Interaction Protocols*, Springer: 2009; pp 89-101.
84. Bastos, M.; Abian, O.; Johnson, C. M.; Ferreira-da-Silva, F.; Vega, S.; Jimenez-Alesanco, A.; Ortega-Alarcon, D.; Velazquez-Campoy, A. J. N. R. M. P., Isothermal titration calorimetry. **2023,** *3* (1), 17.
85. Ribeiro, M. M.; Franquelim, H. G.; Castanho, M. A.; Veiga, A. S. J. J. o. p. s. a. o. p. o. t. E. P. S., Molecular interaction studies of peptides using steady-state fluorescence intensity. Static (de) quenching revisited. **2008,** *14* (4), 401-406.
86. Organization, W. H., Antimicrobial resistance. 2023.
87. Sabtu, N. E., D. A.; Brown, N. M., Antibiotic resistance: what, why, where, when and how? . *British Medical Bulletin* **2015,** *111*, 105–116.
88. Uddin, T. M. C., A. J.; Khusro, A.; et al. , Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. . *Journal of Infection and Public Health* **2021,** *14 (5)*, 601-609.
89. Stohl, E. A.; Dale, E. M.; Criss, A. K.; Seifert, H. S., Neisseria gonorrhoeae Metalloprotease NGO1686 Is Required for Full Piliation, and Piliation Is Required for Resistance to H2O2- and Neutrophil-Mediated Killing. **2013,** *4* (4), 10.1128/mbio.00399-13.
90. Jerse, A. E.; Cohen, M. S.; Drown, P. M.; Whicker, L. G.; Isbey, S. F.; Seifert, H. S.; Cannon, J. G., Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. *The Journal of experimental medicine* **1994,** *179* (3), 911-20.
91. Cahoon, L. A. a. H. S. S., Focusing homologous recombination: pilin antigenic variation in the pathogenic Neisseria. . *Mol Microbiol* **2011,** *81(5)*, 1136-1143. .
92. Liu Y, F. B., Russell MW., New concepts in immunity to Neisseria gonorrhoeae: innate responses and suppression of adaptive immunity favor the pathogen, not the host. *Front Microbiol.* **2011,** *2*, 52.
93. Gibson, B. W.; Melaugh, W.; Phillips, N. J.; Apicella, M. A.; Campagnari, A. A.; Griffiss, J. M., Investigation of the structural heterogeneity of lipooligosaccharides from pathogenic Haemophilus and Neisseria species and of R-type lipopolysaccharides from Salmonella typhimurium by electrospray mass spectrometry. *Journal of bacteriology* **1993,** *175* (9), 2702-12.
94. John, C. M.; Griffiss, J. M.; Apicella, M. A.; Mandrell, R. E.; Gibson, B. W., The structural basis for pyocin resistance in Neisseria gonorrhoeae lipooligosaccharides. *Journal of Biological Chemistry* **1991,** *266* (29), 19303-19311.
95. Gulati Sunita , S. J., Ram Sanjay , Rice Peter A. , Targeting Lipooligosaccharide (LOS) for a Gonococcal Vaccine. *Frontiers in Immunology* **2019,** *10*, 321.
96. Braun, D. C.; Stein, D. C., The lgtABCDE Gene Cluster, Involved in Lipooligosaccharide Biosynthesis in Neisseria gonorrhoeae, Contains Multiple Promoter Sequences. **2004,** *186* (4), 1038-1049.
97. Williams E, S. K., Fairley CK, Pollock GL, Hocking JS, McCarthy JS, Williamson DA., Neisseria gonorrhoeae vaccines: a contemporary overview. *Clin Microbiol Rev* **2024,** *37*, 1.

98. Densen P, G. S., Rice PA Specificity of antibodies against Neisseria gonorrhoeae that stimulate neutrophil chemotaxis. Role of antibodies directed against lipooligosaccharides. *The Journal of clinical investigation* **1987,** *80*, 78–87.
99. Gulati, S.; Beurskens, F. J.; de Kreuk, B.-J.; Roza, M.; Zheng, B.; DeOliveira, R. B.; Shaughnessy, J.; Nowak, N. A.; Taylor, R. P.; Botto, M.; He, X.; Ingalls, R. R.; Woodruff, T. M.; Song, W.-C.; Schuurman, J.; Rice, P. A.; Ram, S., Complement alone drives efficacy of a chimeric antigonococcal monoclonal antibody. *PLOS Biology* **2019,** *17* (6), e3000323.
100. Chakraborti, S.; Gulati, S.; Zheng, B.; Beurskens, F. J.; Schuurman, J.; Rice, P. A.; Ram, S., Bypassing Phase Variation of Lipooligosaccharide (LOS): Using Heptose 1 Glycan Mutants To Establish Widespread Efficacy of Gonococcal Anti-LOS Monoclonal Antibody 2C7. *Infection and immunity* **2020,** *88* (2).
101. Gulati S, M. D. P., Mandrell R E, Jani D B, Rice P A. , Immunogenicity of Neisseria gonorrhoeae lipooligosaccharide epitope 2C7, widely expressed in vivo with no immunochemical similarity to human glycosphingolipids. . *J. Infect. Dis.* **1996,** *174*, 1223-37.
102. Gulati, S.; McQuillen, D. P.; Sharon, J.; Rice, P. A., Experimental Immunization with a Monoclonal Anti-Idiotope Antibody that Mimics the Neisseria gonorrhoeae Lipooligosaccharide Epitope 2C7. *The Journal of Infectious Diseases* **1996,** *174* (6), 1238-1248.
103. Gulati, S.; Zheng, B.; Reed, G. W.; Su, X.; Cox, A. D.; St. Michael, F.; Stupak, J.; Lewis, L. A.; Ram, S.; Rice, P. A., Immunization against a Saccharide Epitope Accelerates Clearance of Experimental Gonococcal Infection. *PLOS Pathogens* **2013,** *9* (8), e1003559.
104. Ram, S.; Gulati, S.; Lewis Lisa, A.; Chakraborti, S.; Zheng, B.; DeOliveira Rosane, B.; Reed George, W.; Cox Andrew, D.; Li, J.; St. Michael, F.; Stupak, J.; Su, X.-H.; Saha, S.; Landig Corinna, S.; Varki, A.; Rice Peter, A., A Novel Sialylation Site on Neisseria gonorrhoeae Lipooligosaccharide Links Heptose II Lactose Expression with Pathogenicity. *Infection and Immunity* **2018,** *86* (8), 10.1128/iai.00285-18.
105. Westphal, O. J. M. C. C., Bacterial lipopolysaccharides: extraction with phenol-water and further applications of the procedure. **1965,** *5*, 83.
106. Kittelberger, R.; Hilbink, F. J. J. o. b.; methods, b., Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. **1993,** *26* (1), 81-86.
107. Beernink, P. T.; Di Carluccio, C.; Marchetti, R.; Cerofolini, L.; Carillo, S.; Cangiano, A.; Cowieson, N.; Bones, J.; Molinaro, A.; Paduano, L. J. J. A., Gonococcal Mimitope Vaccine Candidate Forms a Beta-Hairpin Turn and Binds Hydrophobically to a Therapeutic Monoclonal Antibody. **2024,** *4* (7), 2617-2629.