University of Naples "Federico II"

Polytechnic and Basic Sciences School

Department of Chemical Sciences

Ph.D. School in Chemical Sciences – Cycle XXXVIII

Lectins expression and purification for biochemical and physico-chemical studies of their interaction with eukaryotic glycans

Alessia Stornaiuolo

Supervisor:

Prof. Alba Silipo

Examiner:

Prof. Angela Amoresano

Ad Andrea, che mi guardi da lassù. Sono certa che saresti fiero di me!

TABLE OF CONTESTS

List of Abbreviations

µL: Microliter

µM: Micromolar

AD: AutoDock

ADCC: Antibody-Dependent Cell-mediated Cytotoxicity

ADT: AutoDockTools

APS: Ammonium Persulfate

B: B cells

Ba: Basophils

BLI: Biolayer Interferometry

BMRB: Biological Magnetic Resonance Data Bank

CMAH: Cytidine Monophosphate-N-Acetylneuraminic Acid Hydroxylase

COSY: Correlation Spectroscopy

CPMG: Carr-Purcell-Meiboom-Gill

CPS: Capsular Polysaccharide

CRD: Carbohydrate Recognition Domain

Cryo-EM: Cryo-Electron Microscopy

CSM: Chemical Shift Mapping

CSP: Chemical Shift Perturbation

CTL: C-type lectin

CuAAC: Copper(I)-catalyzed Azide-Alkyne Cycloaddition

CV: Column Volumes

D_2O: Deuterated Phosphate Buffer

DAMP: Danger-Associated Molecular Pattern

DC: Dendritic Cells

DOSY: Diffusion-Ordered SpectroscopY

DTT: Dithiothreitol

E. coli: *Escherichia coli*

ECD: Extracellular domain

EDTA: Ethylenediaminetetraacetic acid

Eo: Eosinophils

EP: Evolutionary Programming

ER: Endoplasmic Reticulum

GA: Golgi apparatus

GBPs: Glycan-Binding Proteins

GBS: *Group B Streptococcus*

gg: gauche-gauche

gt: gauche-trans

HF: Hartree-Fock

HMBC: Heteronuclear Multiple Bond Correlation

HSQC: Heteronuclear Signle Quantum Correlation

Hz: Herz

Ig: Immunoglobulin

IPTG: Isopropyl β- d-1-thiogalactopyranoside

ITAM: Immunoreceptor tyrosine-based activatory motif

ITC: Isothermal Calorimetry

ITIM: Immunoreceptor tyrosine-based inhibitory motif

kDa: KiloDalton

L: Liter

LB: Luria Bertani

LOS: Lipooligosaccharide

LS: Local search

Lum epi: lumen epithelia cells

M: Molar

mAb: Monoclonal Antibody

MAG: Myelin Associated Glycoprotein

MAMP: Microbial-Associated Molecular Pattern

MC: Mast cells

MD: Molecular Dynamics

Mic: Microglia

mL: Milliliter

mM: Millimolar

MM: Molecular Mechanics

Mø: Macrophages

Mo: Monocytes

MS: Mass Spectrometry

MW: molecular weight marker

Myp: Myeloid progenitor

N: Neutrophils

NaCl: Sodium Chloride

NaP: Sodium Phosphate

NK: Natural-Killer cells

NMR: Nuclear Magnetic Resonance

NOE: Nuclear Overhauser Effect

NOESY: Nuclear Overhauser Effect Spectroscopy

Ocl: Osteoclasts

OD: Optical Density

PAMP: Pathogen-Associated Molecular Pattern

PBC: Periodic Boundary Conditions

PME: Particle Mesh Ewald

PR: Polak-Ribière

PRR: Pattern Recognition Receptor

PSB: Phosphate-Buffered Saline

RESP: Restrained ElectroStatic Potential

RMSD: Root Mean Square Deviation

ROESY: Rotating-frame Overhauser Effect Spectroscopy

RPM: Revolutions Per Minute

RT: Room Temperature

SAMP: Self-Associated Molecular Pattern

SAR: Structure-Activity Relationship

Schw: Schwann cells

SDS-PAGE: Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

Siglec: Sialic acid-binding immunoglobulin (Ig)-like lectin

sLex: Sialyl Lewis X

SNFG: Symbol nomenclature for glycans

SPR: Surface Plasmon Resonance

STD: Saturation Transfer Difference

T: T cells

TACA: Tumor-Associated Carbohydrate Antigen

TAE: Tris-acetate-EDTA

TEMED: N, N, N', N'-tetramethylethylenediamine

tg: trans-gauche

TOCSY: Total Correlation Spectroscopy

Tris: Tris(hydroxylmethyl)aminomethane

Tr-NOESY: Transferred-Nuclear Overhauser Effect Spectroscopy

Troph: Trophoblasts

TROSY: Transverse Relaxation Optimized Spectroscopy

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

WaterLOGSY: Water-Ligand Observed with Gradient Spectroscopy

Sugar Abbreviations

DSeDG: (di-galactosyl) diselenide

Fuc: Fucose

Gal: Galactose

GalNAc: N-acetyl galactosamine

Glc: Glucose

GlcA: Glucuronic Acid

GlcN: Glucosamine

GlcNAc: N-acetyl glucosamine

IdoA: Iduronic Acid

Kdn: 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid

LacNAc: N-acetyl lactosamine

Man: Mannose

Neu5Ac: N-acetyl neuraminic acid

Neu5Gc: N-glycolyl neuraminic acid

SeDG: seleno-digalactoside

SeDG-Bn: benzyl 3,3'-seleno-digalactoside

Sia: Sialic acid

TDG: Thiodigalactoside

Unsym(Se): Unsymmetrical seleno-digalactoside

Xyl: Xylose

Amino acids Abbreviations

The following is a list of standard abbreviations for the 20 common amino acids.

Ala (A): Alanine

Arg (R): Arginine

Asn (N): Asparagine

Asp (D): Aspartic Acid

Cys (C): Cysteine

Gln (Q): Glutamine

Glu (E): Glutamic Acid

Gly (G): Glycine

His (H): Histidine

Ile (I): Isoleucine

Leu (L): Leucine

Lys (K): Lysine

Met (M): Methionine

Phe (F): Phenylalanine

Pro (P): Proline

Ser (S): Serine

Thr (T): Threonine

Trp (W): Tryptophan

Tyr (Y): Tyrosine

Val (V): Valine

Abstract

The immune system's extraordinary capacity to discriminate between "self" and "non-self" and respond accordingly is fundamental to health. This intricate balance, crucial for inducing acquired immune responses, regulating inflammation, and establishing tolerance, is increasingly understood to be profoundly influenced by glycans. This post-translational process, found on host cell and pathogen proteins (glycoproteins) and lipids (glycolipids), exhibit natural heterogeneity, acting as crucial carriers of biological information. This information is precisely decoded by lectins, diverse protein families including C-type lectin receptors (CLRs), galectins, and Sialic acid-binding immunoglobulin (Ig)-like lectins (Siglecs), which trigger specific tolerogenic or immunogenic signaling pathways upon glycan recognition. Siglecs are cell surface receptors that recognize sialic acids. They are known modulators of immune responses, modulating immune cell response upon binding with sialylated ligands. These interactions mediate events such as cell adhesion and the inhibition or regulation of immune cell activation. Galectins are soluble proteins binding β-galactoside-containing glycans. They play critical roles in a myriad of physiological and pathological processes. Glycans and their complementary glycan-binding proteins are essential components in controlling both innate and adaptive immunity. Indeed, Sialic acid-Siglec interactions are associated with a broad spectrum of diseases, from autoimmunity to neurodegeneration and cancer. Similarly, Galectin-3 modulates a range of cellular interactions, such as immune response, cell adhesion, migration, and angiogenesis. Consequently, it is a key player in fibrosis and carcinogenesis, as its upregulation often correlates with poor prognosis. Given this context, strategies for rationally modulating lectin-glycan interactions in pathological processes offer significant therapeutic potential. Recent decades have seen the development of various inhibitors, including small molecules, multivalent saccharide ligands, peptides, and peptidomimetics,

offering alternatives against neoplastic and other diseases. This thesis project aimed to elucidate the molecular details of the binding between Siglec-8 and Galectin-3 with specific glycomimetics. The research involved the expression and purification of these proteins, followed by binding studies with sialic acid-containing glycans (for Siglec-8) and β-galactoside-containing glycans (for Galectin-3). Structural insights into these molecular recognition processes were provided by Nuclear Magnetic Resonance (NMR) spectroscopy, employing both ligand-based and protein-based approaches, complemented by biophysical and computational methods, such as docking and molecular dynamics simulations. This comprehensive approach facilitated a detailed examination of the three-dimensional complexes, which enabled the thorough characterization of their recognition and binding processes and provided a deeper understanding of the molecular mechanisms underlying their inhibitory or modulatory activities.

SECTION I – INTRODUCTION

Chapter 1: The importance of glycans and human lectins

1.1 The role of glycans and glycosylation in biological and pathological functions

Glycans, complex carbohydrate structures, are ubiquitous and essential components of living systems, typically found as glycoconjugates – primarily glycoproteins and glycolipids, though they can also exist as free oligosaccharides and polysaccharides. Glycans form a highly layer on the cell surface, known as the glycocalyx, which acts as a dynamic interface between the cell and its surrounding environment [1]. This layer serves as a protective barrier against physical forces and stresses, while also mediating crucial cell-cell interactions [2].

Although glycosylation is predominantly observed on cell surface lipids and proteins, numerous studies also document the presence of glycoconjugates within the nucleus and cytoplasm. Furthermore, the existence of glycans in nucleic acids has been postulated [3]. Thus, glycans are found in the three major classes of macromolecules (proteins, lipids, and nucleic acids) across all cellular compartments, playing indispensable roles in both health and disease.

Over the past decades, the involvement of glycans in a wide variety of biological processes has been extensively demonstrated. Broadly, these functions can be categorized into two main groups: i) intrinsic functions, which include structural and modulatory properties (e.g., nutrient storage), and ii) extrinsic functions, which primarily involve interactions with other molecular partners, especially Glycan-Binding Proteins (GBPs) such as lectins, and their molecular mimicry by pathogens to evade the immune system (**Figure 1.1**) [4-5]. Indeed, the mutual recognition between glycans and lectins is a highly regulated process within the immune system. Through their binding to lectins, glycans are integral to various

cellular mechanisms that contribute to both adaptive and innate immunity [6]. This research field continues to experience rapid growth, yielding fundamental discoveries and promising new applications in disease prevention, diagnosis, and therapy.

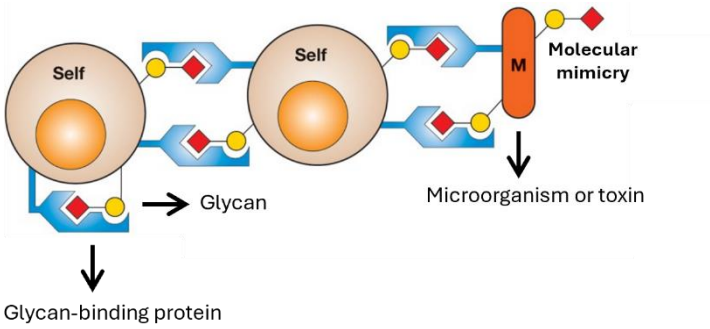


Figure 1.1. Schematic overview of interactions between glycans and glycan-binding proteins (GBPs). The crucial role of glycans in recognition events involves both host-endogenous and exogenous GBPs. Intrinsic recognition (events on the *left*) describes how host GBPs read "self" glycans to regulate normal physiological processes. Extrinsic recognition (events on the *right*) shows interactions with non-host GBPs, including pathogen- and environment-derived lectins. The diagram also highlights how these functional categories can overlap, and how the concept of molecular mimicry, where pathogens display host-like glycans, further complicates and exploits these biological roles. *Adapted from [4].*

Proteins are subject to numerous post-translational modifications during synthesis and maturation. The most abundant modification is glycosylation, a process common to all eukaryotic cells. Protein glycosylation involves successive enzymatic addition of different monosaccharides onto specific amino acid residues within the polypeptide chain, ultimately forming a complex oligosaccharide chain called glycan [7]. The individual monosaccharides commonly found in mammalian glycoproteins and glycolipids are illustrated in **Figure 1.2** [8].

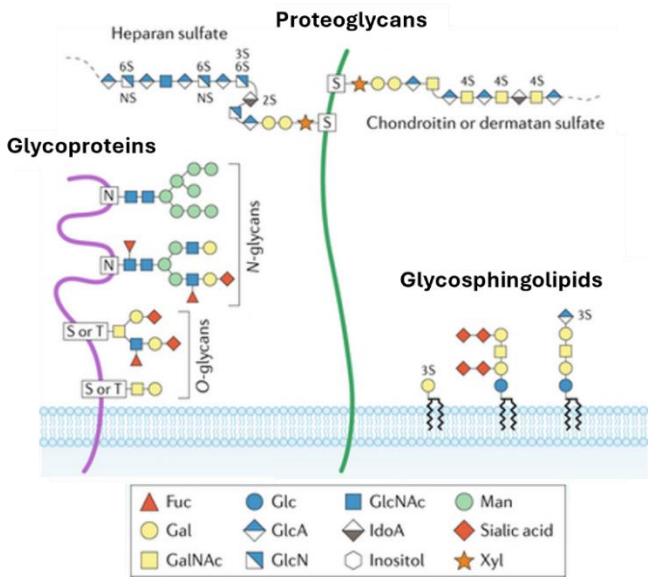


Figure 1.2. Schematic representation of the typical glycans on eukaryotic cells. Glycans can be attached to proteins (glycoproteins and proteoglycans) or to lipids (glycosphingolipids). At the *bottom* of the figure, the most common monosaccharides present in eukaryotic cells are represented as SNFG symbols: D-Glucose (Glc), Sialic Acid (Neu5Ac), D-Glucosamine (GlcN), L-Iduronic Acid (IdoA), D-Glucuronic Acid (GlcA), D-Xylose (Xyl), D-Mannose (Man), L-Fucose (Fuc), D-Galactose (Gal), D-*N*-Acetyl Galactosamine (Gal*NA*c), D-*N*-Acetyl Glucosamine (Glc*NA*c). *Adapted from [8].*

Glycans can exhibit linear or branched arrangements and may be linked by α- or β - glycosidic bonds at various positions, creating immense chemical diversity [9]. The initial addition of monosaccharides typically occurs in the Endoplasmic Reticulum (ER), with the process continuing through the *cis*-, *medial*-, and *trans*-Golgi apparatus (GA), where a diverse array of glycosyltransferases and glycosidases modify the glycan composition [8].

Among the different glycosylation patterns in glycoproteins, *N*-glycosylation and *O*-glycosylation are the most prevalent, commonly found on proteins trafficking through the cellular secretory pathway [7].

O-glycosylation begins with the addition of *N*-acetylgalactosamine (Gal*N*Ac) to the hydroxyl group of a serine or threonine residue, which can then be extended into diverse structures. Gal*N*Ac-type *O*-glycans, also known as mucin-type *O*-glycans due to their ubiquitous presence in mucins, are comprised of four major core structures (Core 1-4), alongside an alternative core based on the sialyl-T antigen (**Figure 1.3A**) [1, 10]. The high density of *O*-glycans found on mucins can serve as barrier against pathogens invasion. However, interactions between *O*-glycans, generally sialylated, and bacterial adhesins also lead to colonization on the host cell, causing infections and triggering the pathogenesis [6].

N-glycosylation, conversely, initiates with the attachment of *N*-acetylglucosamine (Glc*N*Ac) to the nitrogen atom of an asparagine side chain via a β1-*N* linkage, dependent on the recognition of a specific Asn-X-Ser/Thr sequence motif (where X is any amino acid except proline) [11]. The *N*-glycans structure contains a common $(Glc\mathit{N}Ac)_2$ $(Man)_3$ core to which other monosaccharides, including Mannose (Man), Glc*N*Ac, Galactose (Gal), Fucose (Fuc), and Sialic Acid (Neu5Ac), can be attached through various glycosidic linkages to form bi-, tri-, or tetra- antennary structures. Based on sugar composition, *N*-glycans are further classified into three major types: i) high-mannose, ii) hybrid-type, and iii) complex-type (**Figure 1.3B**) [1, 7].

The biosynthesis of glycans represents a remarkably intricate process. In clear contrast to template-dependent biopolymers like DNA, RNA, and proteins, the glycome's assembly is not determined by a genomic template. Consequently, the definitive glycan repertoire of a given cell arises from a confluence of factors, including the inherent protein folding, the accessibility of precursor monosaccharides, the expression profiles of specialized glycosidases and glycosyltransferases within the ER and GA, and the precise spatial arrangement of these enzymes in the Golgi [12].

Mammalian cells possess a vast enzymatic machinery, with over 200 genes encoding various glycosyltransferases and glycosidases. The transcription of numerous such enzymes is dynamically regulated at the genomic level, and their subcellular localization can fluctuate across different compartments, responding to cellular lifecycle stages, metabolic demands, and oxidative conditions. This adaptive regulation enables cells to fine-tune their glycan profiles in direct response to both endogenous and exogenous signals.

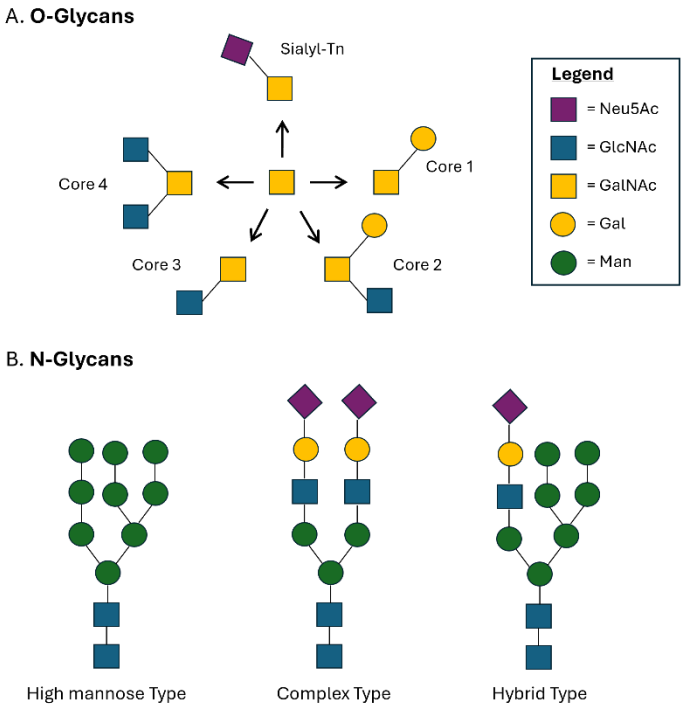


Figure 1.3. The principal structures of O-Glycans (A) and N-Glycans (B).

Moreover, the physiological context profoundly impacts glycome; indeed, pronounced alterations in glycosylation patterns are consistently observed across

a spectrum of pathological conditions, including autoimmune disorders, infectious diseases, chronic illnesses and cancer [7].

1.1.1 Glycans heterogeneity and functionality

The vast heterogeneity on complexity of glycans' structure drives to an enormous array of biological functions (**Figure 1.4**) [13].

As for example, glycans are crucial for cell recognition and signaling, serving as vital interaction sites for other cells, pathogens, and molecules [14]. This recognition is fundamental for immune responses and broader cellular interactions. Beyond their surface roles, glycosylation is essential for proper protein folding and stability, directly impacting overall protein function [15]. Glycans actively modulate antibody effector functions and facilitate pathogen recognition, thereby initiating immune responses [16]. Their interactions, particularly with GBPs like lectins, are key to maintaining the delicate balance between immune tolerance and robust immunity, enabling cells to distinguish "self" from "non-self" components [17].

Glycans also undergo dynamic changes during development, influencing cell differentiation and organogenesis, and acting as specific markers of various cell types and development stages. In disease contexts, altered glycosylation patterns are frequently observed in tumor progression (**Figure 1.4**). These modifications can promote cancer cell proliferation, invasion, and metastasis, notably assisting tumor cells in evading immune detection [18-19]. By interacting with diverse GBPs, glycans influence cell signaling and the tumor microenvironment, further contributing to disease progression and immune escape.

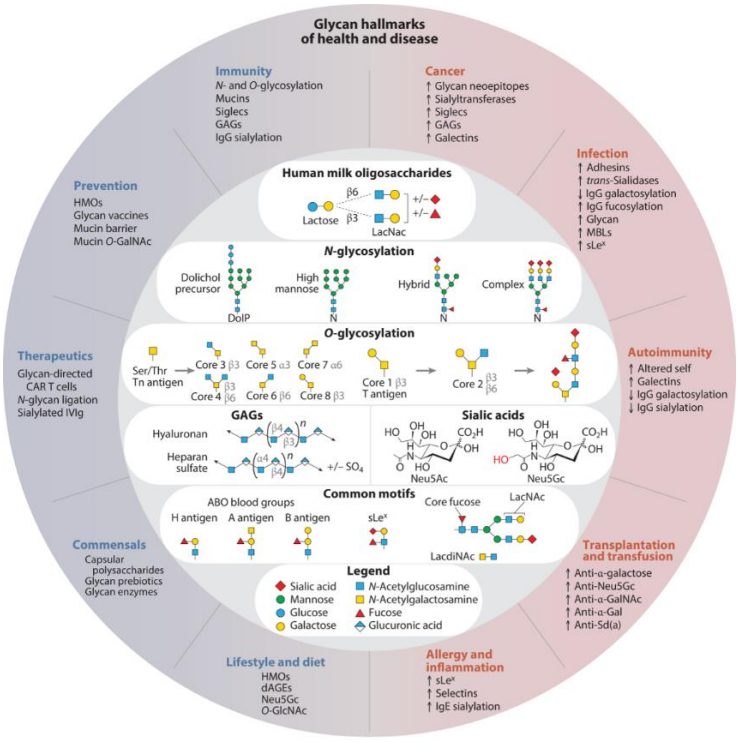


Figure 1.4. Glycans functions in immune homeostasis and pathological alterations. This divided circular diagram provides a comprehensive overview of the significance of glycans in health and disease. Specifically, the central section illustrates the primary glycan structures and motifs fundamental to biological function. The outer ring highlights the dual roles of glycans across various biological contexts: homeostasis and defense (on the *left*) and pathological alterations (on the *right*). *Reproduced from [13].*

1.2 Human Lectins. Decoders of the Glycan Code

While glycans themselves carry a vast amount of biological information through their intricate structures, this "Glycan code" must be interpreted and translated into cellular responses. This crucial role is primarily performed by lectins, a diverse superfamily of GBPs.

Lectins are defined by their ability to recognize and bind specific carbohydrate motifs in a non-catalytic manner, acting as key molecular partners in a myriad of biological processes [20]. These proteins function as sophisticated "decoders" of the glycan language, enabling cells to sense and respond to their glycosylated environment. Their specificity for glycan motifs allows them to mediate highly selective interactions, distinguishing between different cell types, physiological states, and even between host and pathogen [6].

Human lectins are broadly distributed across various cellular compartments and participate in a wide array of fundamental biological functions. In the context of the immune system, they are crucial. Lectins expressed by immune cells, such as C-type lectins, galectins, and Siglecs, act as critical pattern recognition receptors (PRRs). They recognize specific glycan patterns on pathogens, initiating innate immune responses and shaping adaptive immunity by influencing antigen presentation, T cell activation, and cytokine production [17]. Beyond pathogen recognition, lectins also play vital roles in maintaining immune homeostasis, regulating inflammatory processes, and mediating cell-cell adhesion and migration within the immune system [6].

However, the intricate interplay between lectins and glycans is not confined to healthy physiological processes. In various disease states, including cancer, autoimmune disorders, and chronic infections, the functions of lectins can be profoundly altered or even subverted. Aberrant glycan expression on diseased cells can lead to altered lectin binding, contributing to pathological outcomes such as tumor immune evasion, chronic inflammation, or dysregulated immune responses (**Figure 1.5**) [6, 18].

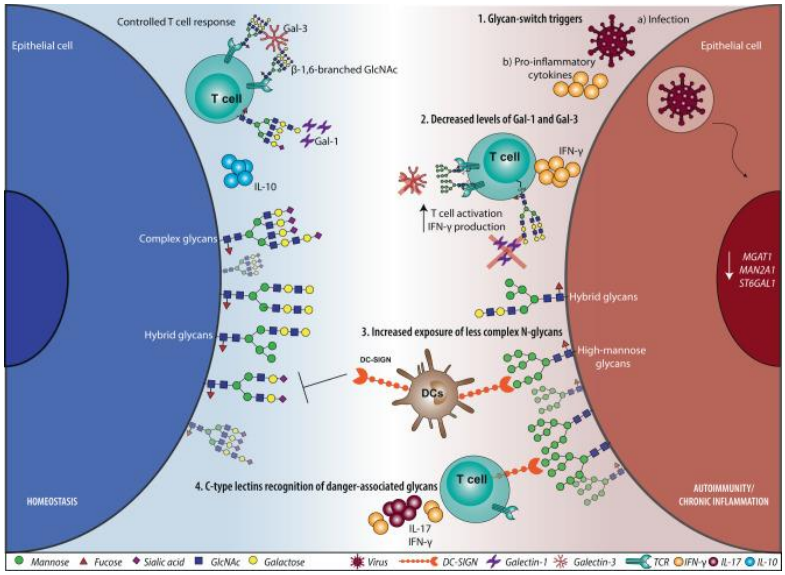


Figure 1.5. Glycan-lectin interactions in homeostasis *vs* dysregulated functions in autoimmunity and chronic inflammation. The crucial contrast between the state of homeostasis (*left*, blue) and dysregulated immune function (*right*, red) illustrates the underlying mechanism: how environmental factors, altered glycan biosynthesis, and changes in GBP levels lead to aberrant glycan exposure and recognition, ultimately contributing to immune dysregulation and severe pro-inflammatory outcomes. *Reproduced from [6].*

Understanding the precise mechanisms by which human lectins decode the glycan code is therefore paramount for unravelling disease pathogenesis and developing novel therapeutic strategies.

Generally, lectins classification often relied on their carbohydrate-binding specificity (e.g., mannose-binding, galactose-binding, sialic acid-binding lectins). However, a more comprehensive and molecular approach now combines amino acid sequence homology, structural folds of their carbohydrate-recognition domains (CRDs), and requirements for divalent cations (such as Ca^{2+}) for binding [21]. Based on these characteristics, animal lectins are primarily grouped into (**Figure 1.6**):

- **C-type lectins (CTLs):** lectins that are characterized by their dependence on Ca^{2+} ions for carbohydrate binding and often possess a conserved C-type lectin domain. This is a very large and functionally different family, including selectins, DC-Sign, Langerins and various pattern recognition receptors [22].
- **Galectins (S-type lectins):** this group includes soluble lectins. They recognize β-galactosides. Galectins are ubiquitous and can be found both intracellularly and extracellularly [23].
- **I-type lectins (Siglecs):** these lectins bind to sialic acid residues and possess immunoglobulin (Ig)-like domains in their extracellular regions. They play crucial roles in cell-cell interactions and immune regulation [24].
- **P-type lectins** (which bind mannose-6-phosphate), **R-type lectins** (characterized by a ricin B-chain-like fold), and **F-type lectins** (fucose-binding lectins): other important, though sometimes smaller or less broadly distributed [21].

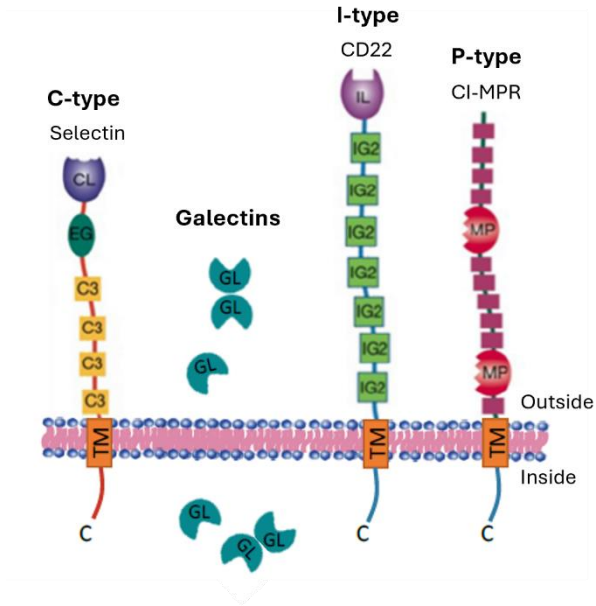


Figure 1.6. Structural classification of common animal lectins. C-type lectins necessitate Ca^{2+} for sialoglycans recognition; Galectins are soluble proteins that are thiol-activated and specific for β-galactosides; I-type lectins incorporate a carbohydrate recognition domain (CRD) with an immunoglobulin-like fold; P-type lectins target glycoproteins bearing mannose 6-phosphate. C-type lectin CRD (CL), Galectin CRD (GL), I-type lectin CRD (IL), P-type lectin CRD (MP), EGF-like domain (EG), complement regulatory repeats (C3), immunoglobulin C2-set domains (IG2), transmembrane segments (TM).

Lectins, as key glycan-binding proteins on immune cells, are essential for maintaining immune homeostasis. They recognize various molecular patterns from pathogens (PAMPs) and cellular damage (DAMPs), playing a critical role in both anti-infection responses and the progression of immune-related diseases [25].

1.2.1 Siglecs: structure, function and recognition mechanisms

Siglecs (Sialic acid-binding immunoglobulin-type lectins) or I-type lectins in humans are composed of 15 members. Based on the sequence conservation, Siglecs are divided into two subgroups: (i) classic Siglecs (comprising Sialoadhesine (Siglec-1), CD22 (Siglec-2), MAG (siglec-4 and Siglec-15); (ii) CD-33 related Siglecs (CD-33 (Siglec-3), Siglec-5 to -14 and Siglec-16) (**Figure 1.7**) [26].

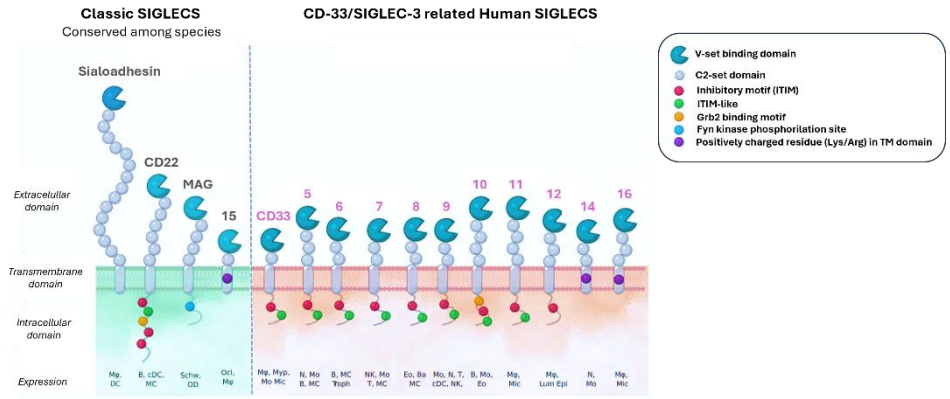


Figure 1.7. Schematic representation of the human Siglec receptors. The Siglec family is organized into "evolutionarily conserved" and "CD-33 related" subsets. The functional roles of Siglecs are determined by their intracellular components: specific features within the cytoplasmic tails and the transmembrane domains differentiate the members into either activatory or inhibitory proteins. At the *bottom* of the figure, the cell-expression patterns are shown: Mø, macrophages; DC, dendritic cell; B, B cells; MC, mast cells; Schw, Schwann cells; OD, oligodendrocytes; Ocl, osteoclasts; Myp, myeloid progenitor; Mo, monocytes; Mic, microglia; N, neutrophils; Troph, trophoblasts; NK, natural-killer cells; T, T cells; Eo, eosinophils; Ba, basophils; Lum epi, lumen epithelia cells. *Adapted from [26]*.

These receptors are defined by their specific ability to recognize and bind sialic acid residues, a type of negatively charged sugar often found at the outermost ends of glycan chains on cell surfaces.

Among the numerous identified sialylated glycan motifs, *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc), and 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (Kdn) are the most prevalent naturally occurring neuraminic acid (**Figure 1.8**). These sugars are typically linked via their C-2 anomeric carbon through α-(2,3) or α-(2,6) glycosidic bonds to hydroxyl groups on galactose or *N*-acetylgalactosamine residues. Sialyltransferases can further generate disialyl core structures which feature at least two sialic acid units connected by α-(2,3), α-(2,6), or α-(2,8) glycosidic linkages (**Figure 1.8**) [26-27].

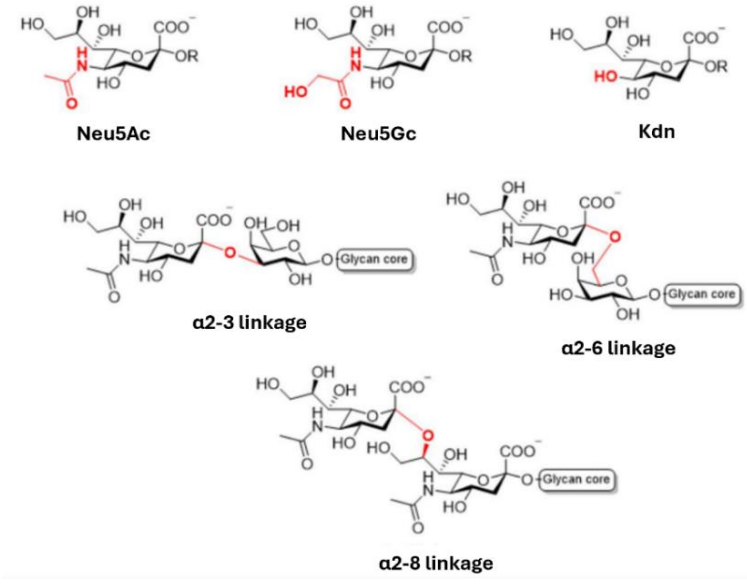


Figure 1.8. Chemical structures of the common mammalian sialic acids and their typical linkages to the subterminal glycans. *Adapted from [26]*.

Structurally, this family of proteins contains an extracellular N-terminal V-type Ig-like domain responsible for binding different sialoglycans. This V-set domain is directly connected to a varying number (from 1 to 16) of constant (C)-type Ig-like extracellular domains. The extracellular portion is followed by a single-pass transmembrane domain that anchors the protein to the cell membrane and connects to a variable number of cytosolic tails, strongly influencing the biological function of Siglecs.

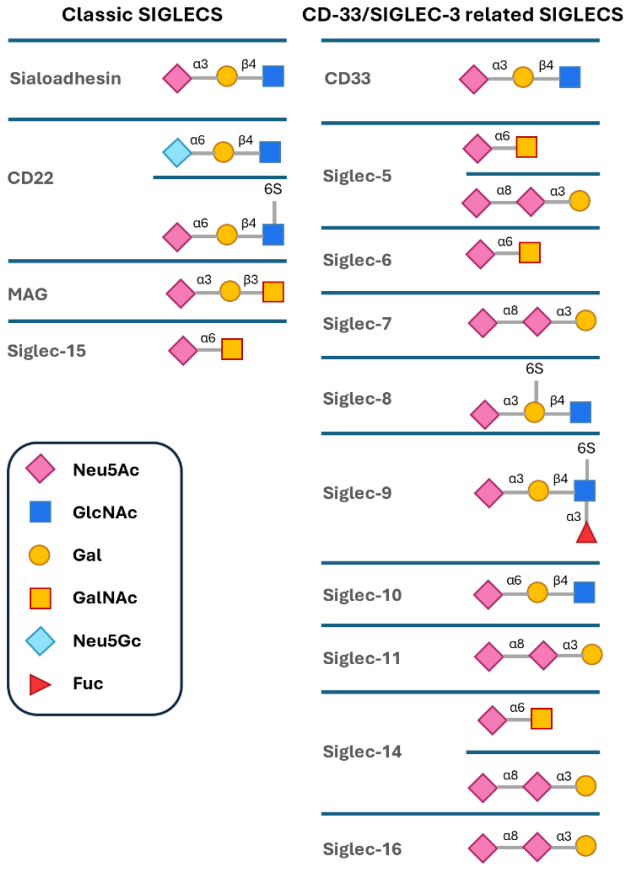
For most Siglecs the intracellular portion contains immunoreceptor tyrosine-bases inhibitory motifs (ITIMs), which serve to recruit phosphatases. Indeed, these ITIMs can inhibit immune cell stimulation through binding and activation of SH-2 domain containing protein tyrosine phosphatases such as SHP-1 and SHP-2 [28].

Conversely, in the case of Siglec-14, -15 and -16, the regulatory domains are immunoreceptor tyrosine-based activatory motifs (ITAMs) containing adapter proteins such as DAP12. These few Siglecs display activating signaling properties [29]. The association between these components derives from a positively charged residue in the Siglec's transmembrane segment engaging with a negatively charged residue on the adapter protein. This engagement activates the ITAM tyrosine signal, triggering the recruitment and subsequent activation of Syk family tyrosine kinases, ultimately leading to immune responses [24, 28, 30].

Additionally, a third category of Siglecs, notably Siglec-1 and Siglec-4, lack characteristic signaling motifs and possess neutral transmembrane domains. Therefore, they don’t directly participate in signaling; instead, their primary roles, like facilitating cell adherence, are carried out by sialic acid recognition [28, 31-33].

While all Siglecs bind sialic acid, each receptor displays a unique binding specificity. For instance, CD22 prefers α-(2-6)-sialoglycans [34-35], Sialoadhesin favors α2-3 [36], Siglec-7 and Siglec-11 target Neu4Ac-α-(2-8)-Neu5Ac structures [37-38]. Even sulfate group position matters, Siglec-8 displays high specificity for Neu5Ac-α-(2,3)-(6-O-sulfo)Gal-β-(1,4)-GlcNAc and Siglec-9 has a strong preference for this structure on glucose [39-41] (**Table 1.1**).

Table 1.1. Glycan binding specificities of human Siglecs.



Adapted from [26].

Found on various cell types within both the innate and adaptive immune systems, Siglecs play a critical role. Within innate immunity, they can inhibit NK cell cytotoxicity, prevent cell death in eosinophils and neutrophils, and block pathogen uptake (a mechanism often exploited by dangerous microbes). In adaptive immunity, Siglecs contribute to modulating T cell activation and polarization, alongside sustaining B cell tolerance to antigens [42].

The precise regulation of immune responses by Siglecs depends on a delicate interplay between how they bind to molecules that carry sialic acid. This process involves a dynamic balance between *cis*-interactions (with glycans on the same cell surface) and *trans*-interactions (with glycans on adjacent cells), a harmony finely tuned by ligand affinity and receptor availability (**Figure 1.9**) [42].

Typically, the high local concentration of sialylated glycans on immune cells means that Siglec binding pockets are predominantly occupied by *cis* associations, thereby outcompeting *trans* binding events [43]. Siglec-1 is an exception; its elongated structure places the ligand-binding domain far from the cell membrane, which allows it to largely bypass *cis* interference. However, various cellular processes, such as membrane sialidase activity, immune activation, or shifts in plasma membrane organization, can "unmask" these receptors, facilitating their engagement with *trans* ligands.

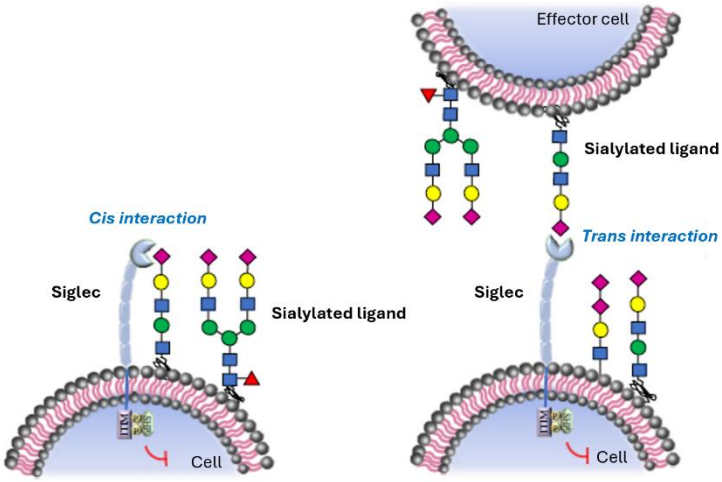


Figure 1.9. How Siglecs recognize their ligands. Siglecs can interact with a sialoglycans exposed on the same cell surface (*cis* interaction) or a different cell surface (*trans* interaction). *Adapted from [42]*.

A crucial aspect of Siglec function is their ability to distinguish between "*self*" (endogenous) and "*non-self*" (exogenous) ligands (**Figure 1.10**) [42].

Endogenous *self*-ligands, which are typically involved in *cis* binding, often possess a comparatively lower affinity. This stands in sharp contrast to the strong avidity of multivalent sialylated structures found on pathogens and tumor cells. This disparity can lead to the generation of more potent inhibitory signals from foreign *trans* ligands, effectively shifting the binding preference. Overall, *cis* binding by Siglecs is vital for maintaining cellular homeostasis, preventing indiscriminate cell-cell interactions that might trigger unwanted signaling. This form of *cis* signaling, often mediated by inhibitory Siglecs, can also influence whether an activating receptor associates or dissociates.

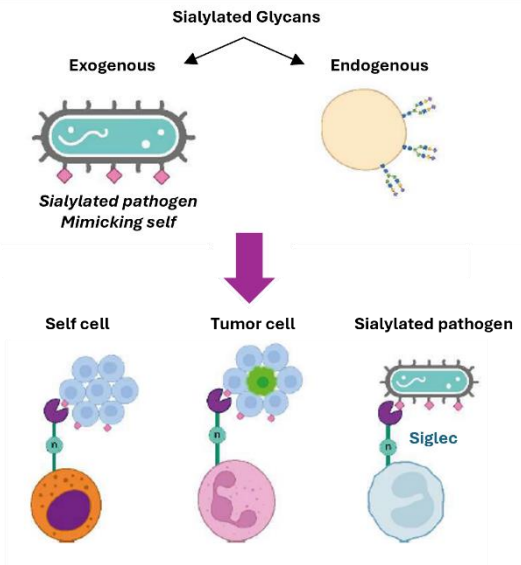


Figure 1.10. Overview of Siglec binding to exogenous and endogenous sialoglycan ligands. *Adapted from [42].*

This remarkable capacity for *self/non-self*-discrimination is unfortunately exploited by certain human pathogens, including *Group B Streptococcus* (GBS), various *Neisseria* species, and *Campylobacter jejuni*. These microbes have evolved to strategically evade the host's immune system by incorporating sialic acid onto their outer surfaces. These invaders are mistakenly recognized as "*self*" by mimicking host "self-associated molecular pattern" (SAMP) structures, like sialylated capsular polysaccharides (CPS) or lipooligosaccharides (LOS). This insidious strategy prevents immune system activation and promotes successful host colonization [44].

Similarly, some cancerous cells promote tumorigenesis and enhance growth by overexpressing sialylated glycans on their outer membrane, using this "*self*"-mimicry to evade immune surveillance. Consequently, given their pivotal roles in both pathogenic evasion and tumor progression, Siglecs represent attractive

targets for developing new therapeutic agents, including specific antibodies and glycomimetics, to treat a range of inflammatory, autoimmune, infectious, and cancer diseases.

Structurally, the V-set domain comprise approximately 70-110 amino acids and are characterized by antiparallel β-sheets, named A-G, interconnected by a single intradomain disulfide bond between B and E β-strands, allowing separation between the β-sheets (**Figure 1.11**) [42].

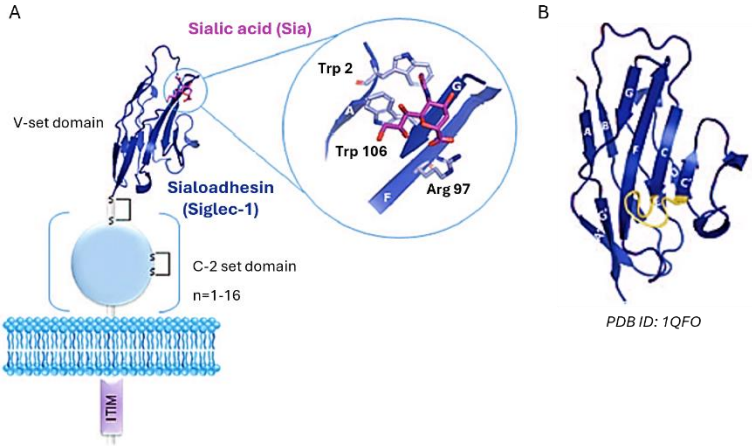


Figure 1.11. Structural features of Sialoadhesin/Siglec-1 V-set domain (*PDB ID: 1QFO*) and its mechanism of sialic acid recognition. (A) The conserved arginine (Arg97) located on the F strands is the critical determinant of Sia recognition by Siglecs. The conserved intra-sheet disulfide in Sialoadhesin broadens the Ig β-sandwich resulting in the exposure of two tryptophan residues (Trp2 and Trp106) on the A and G strands which establishes hydrophobic contacts with the N-acetyl group and glycerol side chain of Sia. (B) The variable CC' loop is highlighted in yellow. *Adapted from [42].*

Detailed analyses of 3D V-domain structures in complex with sialylated di/trisaccharides have yielded critical insights into their ligand recognition modes and specificity for sialic acids [45-51]. For most Siglecs, productive interactions

with sialoglycans are primarily limited to the sialic acid and its adjacent galactose residues, with no additional secondary binding sites yet identified.

The core sialic acid binding pocket is delineated by β-strands F and G, along with the C-C′ and C′-D loops. A key conserved Arginine residue in the F β-strands within this pocket is indispensable for forming a salt bridge with a carboxyl group of sialic acid, enabling specific sialoglycan recognition (**Figure 1.11**). Mutation of this Arginine leads to a significant reduction in binding affinity across all examined Siglecs.

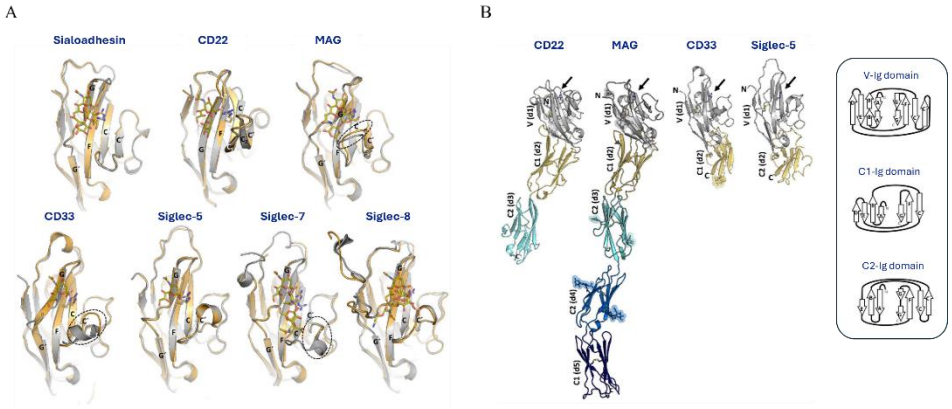


Figure 1.12. 3D structures of selected Siglecs. (A) The superposition of d1 domains from various Siglecs, including Sialoadhesin *(PDB ID: 1QFP and 1QFO)*, CD22 *(PDB ID: 5VKJ and 5VKM)*, MAG *(PDB ID: 5LFR and 5LF5)*, CD33 *(PDB ID: 5IHB and 5J06)*, Siglec-5 *(PDB ID: 2ZG2 and 2ZG3)*, Siglec-7 *(PDB ID: 1O7S and 2HRL)*, and Siglec-*8 (PDB ID: 2N7A and 2N7B)*. The structures are presented in both the unliganded (in grey) and liganded (in orange) forms. (B) Cartoon representations show the crystal structures of the extracellular regions of several Siglecs, specifically CD22 d1-d3 *(PDB ID: 5VKJ)*, MAG d1-d5 *(PDB ID: 5LFU)*, CD33 d1-d2 *(PDB ID: 5IHB)*, and Siglec-5 d1-d2 *(PDB ID: 2ZG2)*. The N-terminal d1 domain (in grey) is identified as a V-type Ig-like domain. An arrow marks the position of the crucial sialic acid binding pocket, which contains a conserved Arginine residue (in sticks). N-linked glycans and disulfide bonds are also visible, represented by sticks and spheres. An insert in the box provides a diagram of the secondary structure differences among V-, C1-, and

C2-type Ig foldings, detailing their distinct strand arrangements. *Adapted from [26]*.

Available structural data highlight those variations in the C-C′ loops and the G strand within the ligand-binding pocket are pivotal determinants of glycan specificity. The specific sequence and adopted conformation of the C-C′ loop directly influences binding preferences for glycan linkages and the galactose moiety. Intriguingly, in Siglec-8, the Arg56 and Gln59 side chains, located on the CC′ loop, are crucial for binding the sulfated Gal6S moiety through the formation of a salt bridge and a hydrogen bond, respectively [50]. While the binding pockets of Sialoadhesin, CD22, and Siglec-8 appear pre-formed to accommodate their ligands, other Siglecs such as CD33, MAG, and Siglec-7 undergo a conformational adjustment in their C-C′ loop upon ligand binding (**Figure 1.12A**).

Excluding CD22, the G strand in many Siglecs contains an inserted loop of variable length. Remarkably, the G-G′ loop of Siglec-8 is unusually long, comprising eleven residues compared to the typical five in most Siglecs. This extended loop allows the flexible side chains of Lys120 and Gln122 in Siglec-8 to interact with the GlcNAc portion of its preferred 6′S sLex ligand [50].

Furthermore, the C-type Ig domains within Siglec extracellular domains (ECDs) can adopt either C1 or C2 topological arrangements (**Figure 1.12B**). These distinct topologies, alongside differences in interdomain linker lengths, can notably alter the interfaces between Ig domains, potentially impacting ECD flexibility. As demonstrated by the crystal structure of MAG [47] and 3D electron microscopy reconstruction of CD22 [51], the ECD forms a semi-rigid, elongated structure. This architecture effectively projects the V-domain's ligand-binding pocket away from the cell surface, a conformation potentially beneficial for efficiently accommodating binding with both flexible *cis* and *trans* ligands [44, 51].

For their fundamental roles in immune regulation, Siglecs have emerged as significant therapeutic targets [52-54]. In particular, the restricted expression of these proteins on specific cell types offers an advantage for highly Siglec-targeting treatments. Siglec-8, for instance, has attracted considerable attention as a therapeutic target for asthma and allergies due to its selective presence on eosinophils (EO) and mast cells (MC) [55-58]. Numerous strategies are being explored to target Siglecs by exploiting these characteristics.

1.2.1.1 Siglec-8

Human Siglec-8 (Sialic acid-biding immunoglobulin-like lectin 8) is a type I transmembrane protein of 431 amino acids belonging to the CD33-related Siglec subgroup, found on chromosome 19q13.33-q13.41. [59].

The extracellular region of this protein contains a hydrophobic signal peptide and three Ig-like domains consisting of a unique N-terminal V-set domain and two C2-set domains. These two constant domains, located downstream of the V-set domain, provide structural support and contribute to the overall rigidity and presentation of the sialic acid-binding site. Following the transmembrane region, there is a cytoplasmic tail of 47 amino acids containing immunomodulatory motifs such as one immunoreceptor tyrosine-based inhibitory motif (ITIM) and one ITIM-like domain (**Figure 1.13**). Its architecture is precisely engineered for its roles in immune regulation.

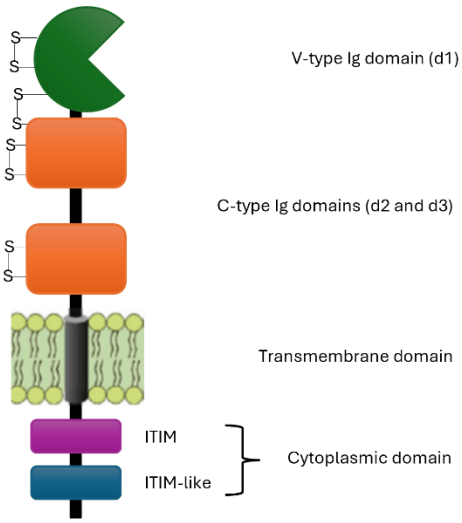


Figure 1.13. Graphical representation of Siglec-8 architecture. The extracellular domain is composed of a V-type Ig domain (d1) and two C-type Ig domains (d2 and d3), followed by a single transmembrane domain and a cytoplasmic domain, with one ITIM and one ITIM-like. Various disulfide bonds are formed between cysteine residues, including three intra-domain bonds and one inter-domain bond that links separate structural units.

In the extracellular domain Siglec-8 contains three potential *N*-linked glycosylation sites. Additionally, this crucial V-set domain is responsible for the sialic acid binding specificity. The available NMR structure [50] and X-Ray crystallographic structure [60] showed that this domain adopts a β-sandwich fold consisting of two antiparallel β-sheets (strands ABED and C'CFG) connected by one intra-sheet disulfide bond between adjacent β-strands B and E.

For Siglec-8, this domain exhibits a strong preference for α2,3-linked sialic acids. In particular, glycan microarray analyses have revealed high specificity for a unique sulfated and sialylated tetrasaccharide glycan epitope, termed 6'-sulfo sialyl Lewis^x (6'S sLe^x) or Neu5Acα2-3[6S]Galβ1-4[Fucα1-3]Glc*NA*c [41, 61] (**Figure 1.14**).

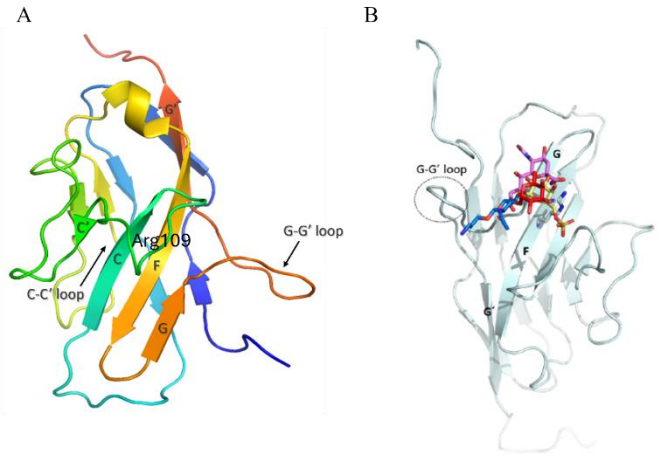


Figure 1.14. Structural features and ligand binding of the human Siglec-8 lectin domain. (A) The secondary structure of the unliganded human Siglec-8 lectin domain *(PDB 7QUI)* displays the CC' and GG' loops (indicated by arrows), G and F β-sheet strands which are critical components of the ligand recognition pocket and the position of the essential arginine (Arg109) which is crucial for sialic acid recognition. (B) In the Siglec-8 V domain with its high-affinity ligand 6'S sLe^x (*PDB ID: 2N7B*), based on NMR studies by Propster et al. [50], are highlighted the F and G strands of the β-sheet, forming the binding platform, and the GG′ loop, demonstrating its role as an extended, eleven-residue segment. This loop uses its long, flexible sidechains to engage with and stabilize the bound 6'S sLe^x ligand. The interaction enhances the domain's specificity for this sulfated ligand.

Siglec-8 is primarily expressed on human eosinophils and mast cells, with some expression also found on basophils. These cell types are central to allergic inflammation and parasitic infections. The primary function of this protein is to act as a negative regulator of inflammation, which it achieves by modulating the activity of those cells. Upon binding to its specific sialylated ligands, Siglec-8 triggers a unique intracellular signaling cascade.

One of the most well-characterized functions of Siglec-8 is the induction of apoptosis (programmed cell death) in eosinophils. This pro-apoptotic effect is

particularly significant in controlling chronic inflammatory conditions where these cells accumulate, such as eosinophilic asthma, hypereosinophilic syndromes, and atopic dermatitis. Beyond inducing cell death, Siglec-8 engagement also inhibits the activation and degranulation of mast cells and eosinophils, leading to a reduction in the release of pro-inflammatory mediators like histamine, leukotrienes, and cytokines.

In summary, the allergic response begins when mast cells and eosinophils are activated. This activation occurs through the cross-linking of IgE antibodies (anti-FcεRI) with their high-affinity receptor FcεRI [62]. The process triggers the release of inflammatory mediators like cytokines and chemokines, which in turn drive and regulate acute or chronic inflammation [63]. Since an abnormal allergic response can lead to various allergic and inflammatory diseases, the body employs inhibitory mechanisms. In these situations, Siglec-8 can temper this response by inhibiting histamine release. Upon activation, Siglec-8's ITIMs become phosphorylated, a process that specifically inhibits inflammatory mediator release from mast cells and induces apoptosis in eosinophils (**Figure 1.15**) [64].

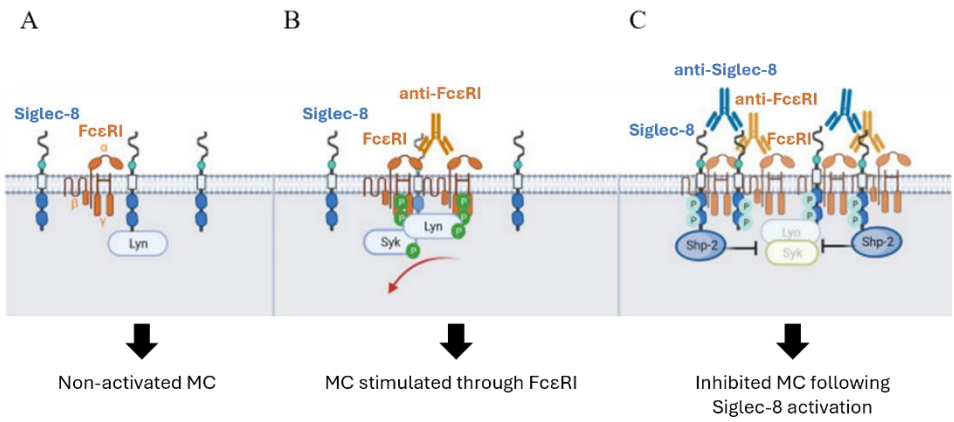


Figure 1.15. Siglec-8's role in inhibiting FcεRI signaling in mast cells (MC). (A) Siglec-8 and FcεRI are shown distributed on the membrane, with some overlap. (B) Cross-linking with anti-FcεRI (e.g., with MAR-1 mAB) and FcεRI

forms high density clusters, leading to ITAM phosphorylation, MC activation/degranulation, and partial exclusion of Siglec-8 form active complexes. (C) In the presence of Siglec-8 mAb, large complexes of both activating (FcεRI) and inhibitory (Siglec-8) receptors form. This process recruits phosphatases like SHP-2, resulting in potent inhibition of intracellular signaling. *Adapted from [64]*.

These distinct inhibitory actions strongly highlight the pivotal role of Siglec-8 in immuno-inhibition during allergic response. For these reasons, Siglec-8 represents a highly promising therapeutic target for treating inflammatory and allergic diseases, particularly those driven by the abnormal activity of eosinophils and mast cells.

Currently, therapeutic strategies primarily focus on two main approaches: monoclonal antibodies (mAb) and glycomimetics (**Figure 1.16**). Antibody-based therapeutic methods represent the most advanced approach in drug development targeting Siglec-8. These antibodies are designed to specifically bind to Siglec-8 on the cell surface, activating its inhibitory function or inducing cell death.

AK002 is a humanized, non-fucosylated monoclonal antibody, which binds Siglec-8 and enhances its efficacy through increased antibody-dependent cell-mediated cytotoxicity (ADCC) [65]. Phase 2 clinical trials have shown their ability to significantly reduce eosinophil counts and alleviate symptoms in patients with eosinophilic gastrointestinal diseases, such as eosinophilic gastritis and duodenitis [66]. It can inhibit the mast cells activity in patients with chronic urticaria, refractory to the conventional therapies [67]. This validates Siglec-8 as an effective pharmacological target. The mechanism of action for AK002 mirrors the inhibitory pathway previously illustrated in **Figure 1.15C** [62].

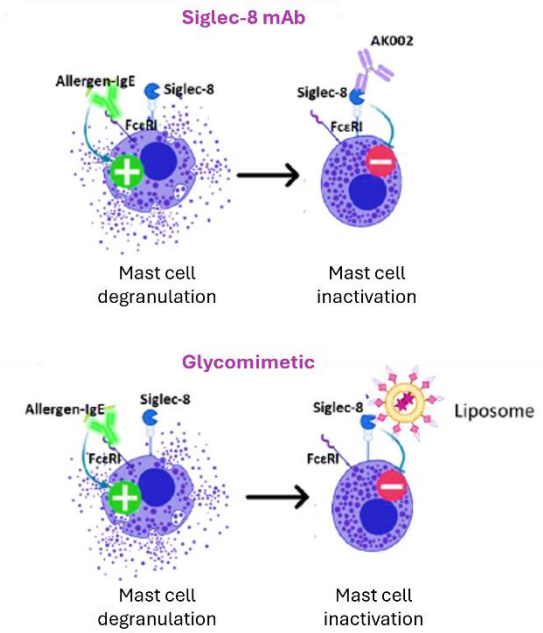


Figure 1.16. Two distinct therapeutic mechanisms targeting Siglec-8. *(top)* Antibody-mediated depletion: the AK002 anti-Siglec-8 antibody acts by binding the receptor, which results in the depletion of eosinophils and the suppression of mast cells degranulation. *(bottom)* Glycomimetic inhibition: liposomes displaying glycomimetic ligands for Siglec-8 achieve their therapeutic effect through inhibition of mast cells and eosinophils activation without inducing cell depletion.

However, anti-Siglec monoclonal antibodies (Abs) are now used to modulate Siglec-sialic acid signaling. Generally, they work by either depleting target cells or blocking Siglec-sialic acid interactions [26]. AK002, specifically targeting Siglec-8, exhibits both these actions depending on the cell type. The precise structure of the Siglec-8-AK002 complex was disclosed and the 3D structure revealed that the AK002 binds to Siglec-8 at a site distinct from its sialic acid binding pocket, thus not interfering with the protein's ability to interact with its natural glycan ligand. This finding was confirmed using a technique called Saturation Transfer Difference NMR (STD NMR). The experiment involved mixing a part of the Siglec-8 that was already bound to the AK002 (2C4 Fab

portion) with a natural Siglec-8 ligand, 3′SLN (a sialic acid-containing sugar, 3'-Sialyllactosamine). Despite the antibody being attached, the NMR signals for the 3′SLN sugar were still strong. This proves that the AK002 antibody does not block the Siglec-8 protein from binding to its natural sugar ligand [60].

Alternative therapeutic strategies for modulating specific Siglecs include glycomimetics, which are chemically modified sialic acid moieties. Naturally occurring sialylated glycan ligands for Siglecs typically exhibit diverse selectivity and weak monovalent affinity (ranging from 0.1-3 mM). While these glycan ligands achieve stronger binding through multivalent interactions [68-70], their structures nevertheless offer a foundation for designing more effective synthetic counterparts. Indeed, these drug-like compounds mimic native glycans but are engineered for improved affinities, enhanced bioavailability, and longer serum half-lives [71].

The design of these Siglec-targeting modified glycans centers on synthetically altered sialic acid scaffolds. While the carboxylic C1 position is crucial for Siglec binding, the rest of the scaffold (C2 to C9) can be modified (**Figure 1.17**) [26]. A key challenge is achieving sufficient potency for effective binding within the active site, which can be partially obscured by endogenous *cis* glycans on target cells.

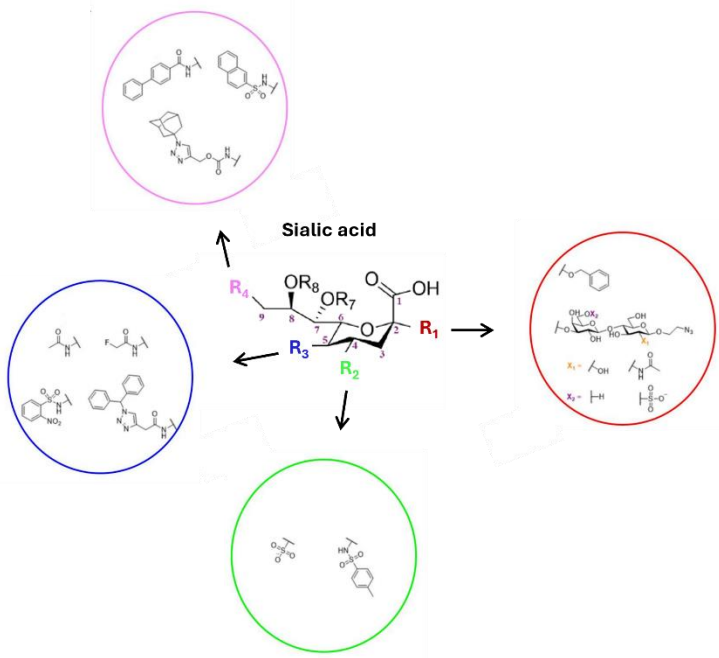


Figure 1.17. Structural representation of the sialic acid molecular platform and its various chemical substitutions (*R1-R4*). These modifications are key to designing high-affinity and specific glycans that effectively bind to Siglecs. *Adapted from [26].*

Early foundational work by Kelm et al. on CD22 (a related Siglec) provided the basis for this approach, showing that modifications like a C-9 and N-acetyl group or a C5-fluoroacetate could significantly boost binding affinity through optimized hydrogen bonding and lipophilic interactions [72]. These insights paved the way for designing a new class of unnatural glycans [73-75]. One notable example for Siglec-8 was shown in **Figure 1.18** [76].

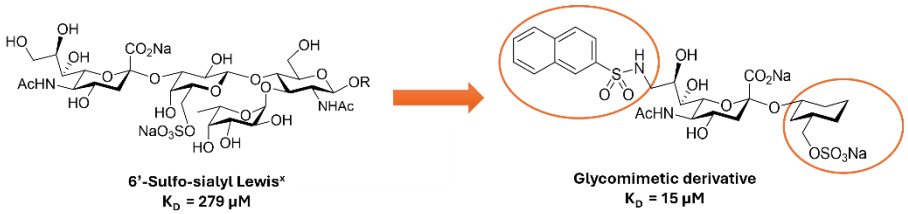


Figure 1.18. The chemical structure of the 6'-sulfo-sialyl Lewis^x and a high affinity glycomimetic. *(left)* The tetrasaccharide has been identified as a specific Siglec-8 ligand in glycan array screening. *(right)* The successful development of a glycomimetic derivative which maintains the neuraminic acid core, substitutes the Gal moiety with a carbocyclic mimetic and adds a naphthalene sulfonamide at C-9 to improve binding affinity. *Adapted from [76].*

Traditional structure-activity relationship (SAR) studies used for designing these compounds are often slow. To overcome this, a new high-throughput strategy utilizing the copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction has been developed. This method facilitates the synthesis and screening of large sialic acid analogue libraries, enabling the rapid incorporation of 1,2,3-triazole scaffolds and other substitutions at desired positions [77]. These analogues are printed as microarrays on glass slides and then tested with fluorescently labeled recombinant Siglecs. Rillahan et al. identified ligands for CD33 and Siglecs-5, -7, -9 and -10, in absence of structural information for the majority of the family members [77].

Recently, Professor James Paulson's research group identified a sulfonamide-based glycan analogue ligand that interacts specifically with Siglec-8. They synthesized a library of 156 sulfonamidesialoside analogues from basic (Neu5Acα2-3Galβ1-4GlcNAc) and 6′-O-sulfo Neu5Acα2-3Galβ1-4GlcNAc scaffolds (the latter preferentially recognized by Siglec-8). Systematic C9 modifications led to the identification of ^{NSA}Neu5Ac (Neu5Ac modified at C-9 with a 2-naphthyl sulfonyl group) as the prime ligand after screening against recombinant Siglec-8 [78]. Therefore, the best ligand, *9-N*-(2-naphthyl-sulfonyl)-

Neu5Acα2-3-[6-*O*-sulfo]-Galβ1-4GlcNAc (6'-*O*-sulfo^{NSA}Neu5Ac) combined the lead 2-naphthyl sulfonyl C-9 substituent with the preferred sulfated scaffold.

Targeted liposomes decorated with this 6'-*O*-sulfo^{NSA}Neu5Ac glycomimetic exhibited robust binding and uptake *in vitro*, showing high selectivity for cells expressing Siglec-8 or -F (the mouse functional ortholog of human Siglec-8). When administered to mice, they also effectively targeted Siglec-F-positive eosinophils *in vivo* (**Figure 1.19**) [78]. Moreover, when these liposomes were evaluated in a transgenic mouse model where mast cells expressed Siglec-8, they demonstrated the ability to recruit Siglec-8 to the IgE-FcεRI complex, thereby suppressing subsequent activation and desensitizing mast cells to antigen-induced degranulation [79].

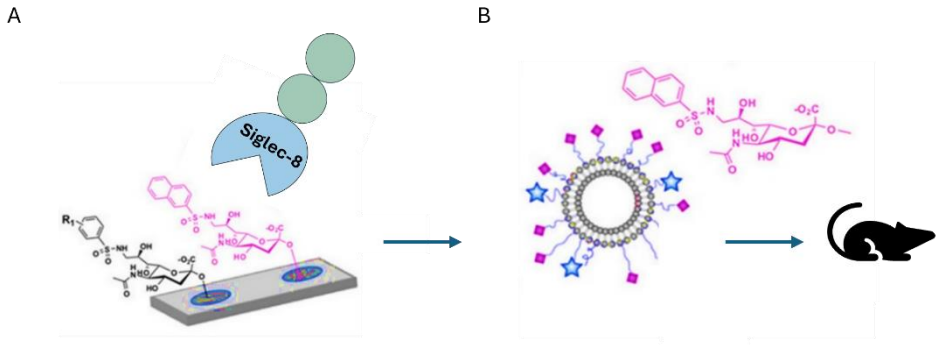


Figure 1.19. Schematic representation of the workflow for developing and validating Siglec-8-targeted glycomimetic liposomes. (A) Glycan array screening for potent glycomimetic Siglec-8 ligands. (B) Liposomes functionalized with high-affinity glycomimetic ligand for Siglec-8 intended for *in vivo* targeting. *Adapted from [78].*

These findings were significantly advanced by a recent structural study [60], which deciphered the molecular recognition features of Siglec-8 in complex with both a high-affinity sialoside analogue and a therapeutic antibody (**Figure 1.20**). By combining NMR spectroscopy and X-ray crystallography, this research

provided atomic-level detail on the binding mode of a high-affinity sialoside mimetic, specifically *9-N*-naphthylsulfonamide-Neu5Ac (^{NSA}Neu5Ac), to Siglec-8. The results demonstrated that the sialoside ring of ^{NSA}Neu5Ac binds to the canonical sialyl binding pocket common to the Siglec receptor family. Furthermore, the study also revealed the three-dimensional structure of Siglec-8 in complex with the fragment antigen-binding (Fab) portion of the AK002 antibody 2C4 (Lirentelimab). Understanding both the glycomimetic and antibody binding sites at this resolution is crucial for the rational design of even more potent and selective Siglec-8 inhibitors, offering key insights into the mechanism by which these analogues can suppress mast cell degranulation and thereby providing a structural basis for their specific therapeutic potential in allergic diseases.

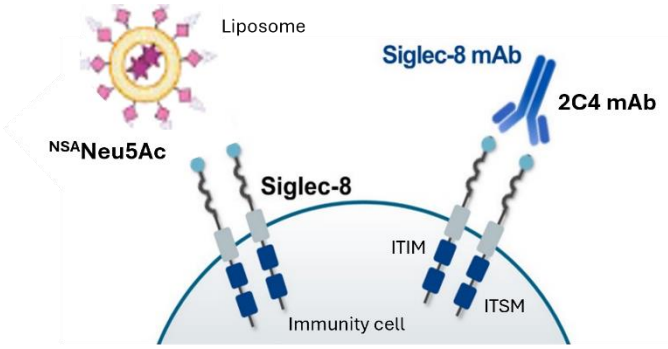


Figure 1.20. A diagram of the binding of the therapeutic monoclonal antibody (mAb) 2C4 and the glycomimetic ^{NSA}Neu5Ac (displayed on liposome) to Siglec-8 on the surface of an immunity cell. *Adapted from [62].*

1.2.2 Galectins: structure, function and binding mode

Galectins (Gals) are a ubiquitous family of β-galactoside-binding lectins, historically named S-type lectins because of their initial observed dependency on sulfhydryl groups from "free" cysteine residues. This family of proteins primarily recognizes β-galactosides within *N*- and *O*-glycan structures [80].

Fifteen distinct galectins have been identified to date, numbered sequentially based on their discovery and broadly categorized into three structural subgroups (**Figure 1.21**) [81-82]. The prototype galectins (e.g., Gal-1, -2, -5, -7, -10, -11, -13, -14 and -15) possess a single CRD and can exist as monomers (e.g., Gal-5, -7, and -10) or form homodimers (e.g., Gal-1, -2, -11, -13, -14, and -15). The tandem-repeat galectins (e.g., Gal-4, -6, -8, -9, and -12) are characterized by two homologous CRDs joined by a single polypeptide chain that acts as a linker. Distinctively, Galectin-3 forms its own subgroup as a chimeric protein, characterized by a single CRD fused to a unique non-lectin N-terminal domain rich in Gly-Pro-Tyr motifs, which facilitates its oligomerization [83].

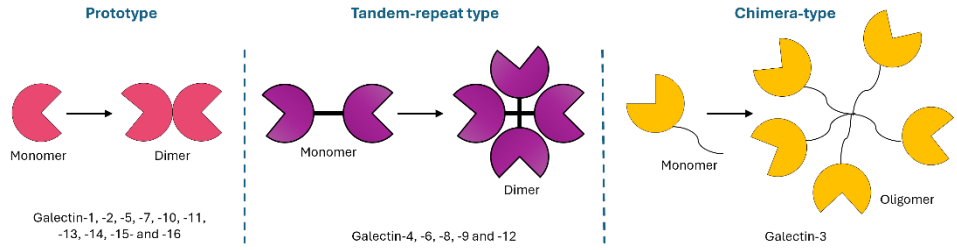


Figure 1.21. Schematic representation of different galectins structures and members of each family. Based on the CRD and its assembly, galectins can be divided into prototype, tandem-repeat type, and chimera-type.

Despite varying sequence identities, all galectins share a highly conserved CRD, typically comprising around 130 amino acids [84]. This domain adopts a

characteristic two-antiparallel β-sheet sandwich fold, often described as a "closing hand" shape. The "palm" is formed by S1 to S6 strands, while the "backhand" is formed by strands F1 to F5 (**Figure 1.22**).

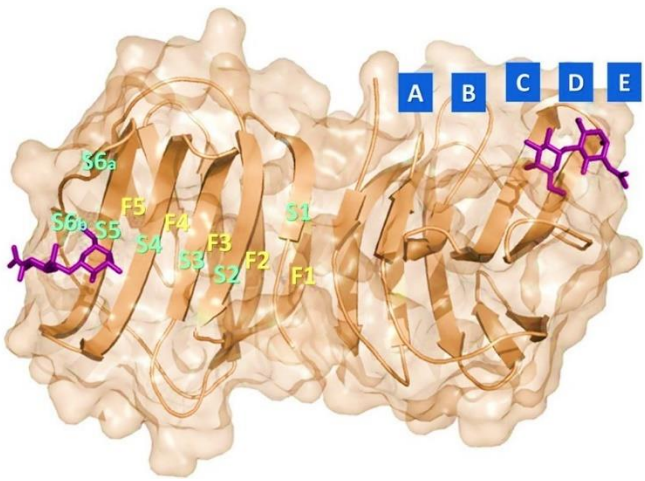


Figure 1.22. An illustration of the CRD of *h*Gal-1 (dimer) in complex with N-acetyllactosamine (*PDB ID: 1W6P*). The complex protein-ligand is displayed, highlighting both the subsites (A, B, C, D, E) of the binding pocket and the architecture of the β-strands (S1, S2, S3, S4, S5, S6, F1, F2, F3, F4, F5). *Adapted from [82].*

The carbohydrate binding pocket is located within the S-sheet, with specific involvement of strands S4, S5, and S6 mediating the core motif recognition [81]. In particular, the β-Gal-binding site is located on subsite C and may be further extended away through vicinal subsites A, B, D and E (**Figure 1.22**). These extensions display important variations between the different members of the family [82].

Protein-carbohydrate interactions, fundamental for intercellular communication, are mediated by hydrogen bonds, electrostatic forces, and crucial van der Waals interactions, including a stabilizing ring-stacking interaction between the

saccharide unit and a highly conserved Trp residue within the binding pocket (**Figure 1.23**). Additionally, water molecules found in the binding pocket play a role in ligand stabilization. Although monovalent binding to simple β-galactosides is typically weak, with dissociation constants (K_d) often in the millimolar range (0.1-3 mM), affinity significantly increases when binding to natural glycoconjugate ligands, with K_d values that can reach the submicromolar range [80].

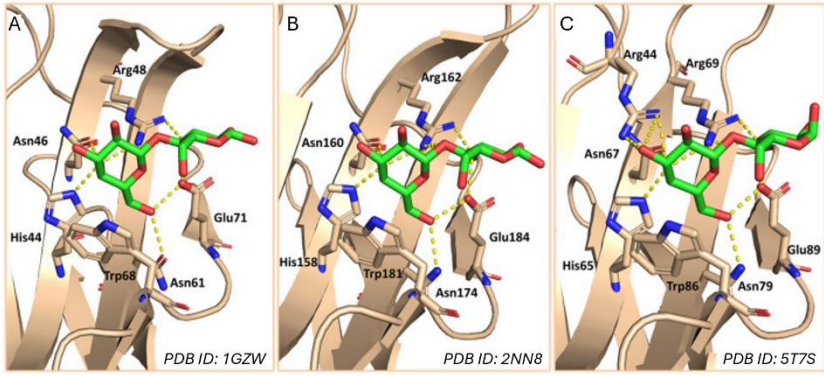


Figure 1.23. Binding sites for different Galectins. (A) *h*Gal-1 in complex with lactose, (B) *h*Gal-3 in complex with lactose, (C) the N-terminal CRD of *h*Gal-8 bound to lactose. The residues involved in the interaction are shown as sticks, while H-bonds as yellow dots. *Adapted from [82].*

Galectins are ubiquitously expressed across mammalian tissues, including most cells of both the innate and adaptive immune systems [85]. Their involvement in a wide array of biological processes, including apoptosis, cell proliferation, and inflammation, is well-established [81, 86]. Since galectins can recognize glycoconjugates containing β-galactoside present on the cell surface, in extracellular matrices, or in the lumen of intracellular vesicles, they serve as crucial soluble mediators in cell communication and signaling (**Figure 1.24**).

During inflammatory responses, galectins can be upregulated by various immune cells, such as $CD4^+$ and $CD25^+$ regulatory T cells, macrophages, dendritic cells (DCs), and natural killer (NK) cells [86]. While widely distributed, some galectins preferentially localize to specific tissues (e.g., Gal-7 in skin, Gal-12 in adipose tissue, Gal-10 in eosinophils and regulatory T cells) [87]. Beyond their extracellular roles, intracellular galectins can also bind ligands in a carbohydrate-independent manner, modulating different cellular functions [87, 88].

Galectins are integral regulators of immune cell development and homeostasis, impacting B cell maturation and differentiation in both central and peripheral immune compartments [89]. They also regulate T cell homeostasis, as their interaction and lattice formation with T cells can trigger signaling cascades that modulate T cell differentiation, functional activation, and the production of pro- and/or anti-inflammatory cytokines [85]. Consequently, because of galectins' significant involvement in immune homeostasis, any dysregulation in their normal activity can lead to beneficial or adverse effects in pathological conditions [85]. Indeed, their dysregulation has been associated with a wide range of diseases, such as cancer progression, autoimmune disorders, fibrosis, arthritis, obesity, cardiovascular diseases, allergies, and microbial infections [90-91].

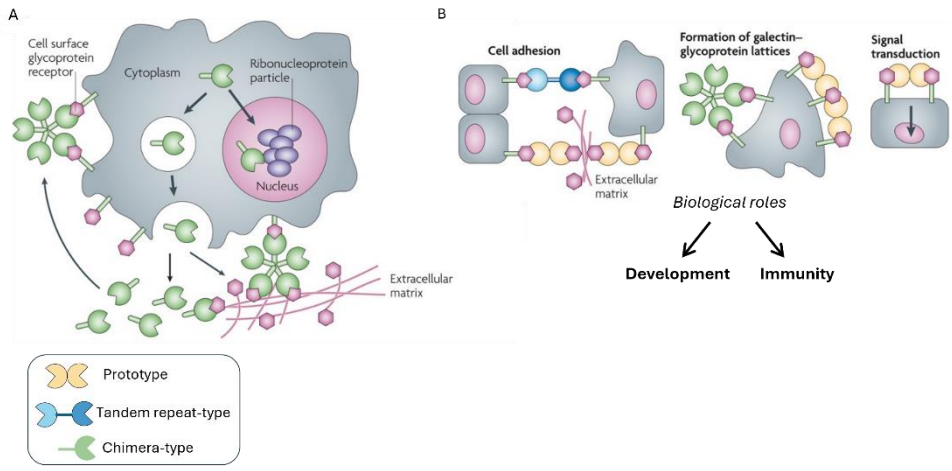


Figure 1.24. Exploring Galectins. (A) Subcellular localization and secretion. (B) Biological roles at the cell surface. *Adapted from [85].*

Galectins are particularly studied for their implication in cancer, as they are involved in numerous cancer hallmarks, including angiogenesis, tumor growth, immune cell adhesion and escape, as well as migration, invasion, and metastasis (**Figure 1.25**) [92]. Consequently, these findings have prompted significant efforts to develop galectin-targeting compounds that can hamper tumor progression.

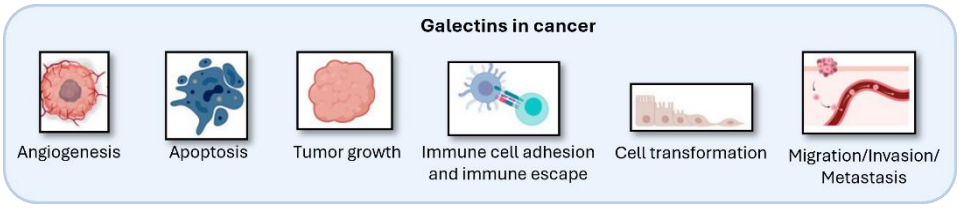


Figure 1.25. Galectins: central to cancer pathology. These proteins are implicated in key functional changes, including the promotion of tumor growth, angiogenesis, cell transformation, invasion, and metastasis, alongside regulating immune escape and immune cell adhesion, and modulating apoptosis. *Adapted from [92].*

Their capacity to act as PRRs enables galectins to directly engage with bacterial surface glycans, thereby modulating leukocyte function and the inflammatory response. Indeed, galectins can recognize glycans present on the surface of both Gram-positive and Gram-negative bacteria. This demonstrates host galectins' ability to directly bind glycoconjugates on bacterial surfaces, thereby either facilitating or inhibiting pathogen entry and subsequently regulating host innate and adaptive immunity [93].

For these all reasons, targeting galectins for clinical applications has become an intense area of research. Since most of their activities are associated with their carbohydrate-binding characteristics, the inhibition of their CRDs by glycomimetics capable of competing with natural ligands appears to be a feasible option. This approach not only promises to elucidate their exact functions, many of which are still unexplored, but also to facilitate the development of novel molecules for therapeutic intervention. Indeed, various polysaccharide-based inhibitors have been developed, with some currently being evaluated in Phase I or Phase II clinical trials for different cancer types [94-95].

1.2.2.1 Galectin-3

Galectin-3 (Gal-3) occupies a singular position within the galectin family, being the only member classified as a chimeric protein. Its distinctive structure comprises a single CRD situated at the C-terminal end, fused to a highly flexible N-terminal domain. This N-terminal region, encompassing approximately 100-150 amino acids across 7-14 repeating sequences, is notably rich in glycine, proline, tyrosine, and glutamine residues, yet lacks charged or large hydrophobic amino acids. It also contains phosphorylation sites at Ser6 and Ser12 [96-97]. The

CRD itself is intricately folded from around 135 amino acids, organized into different β-strands (named F1-F5 and S1-S6) (**Figure 1.26**) [98].

Crucially, the N-terminal domain drives the ordered organization of Gal-3 monomers into higher-order oligomers [83]. The unique ability of Gal-3 to shift between monomeric and aggregated forms is central to its function. Upon binding to glycan ligands on the cell surface, the Gal-3 structure undergoes conformational changes, enabling the self-assembly of its N-terminal regulatory domains. This process results in the formation of multivalent Gal-3 complexes, which are pivotal in mediating cellular activation or repression through three main mechanisms: receptor clustering, lattice formation, and cell-cell interactions [99].

Like Galectin-1, Gal-3 exhibits a preference for Lac*NA*c sequences found in both *N*- and *O*-glycans, with this specific recognition being essential for constructing various galectin-glycan lattices [100].

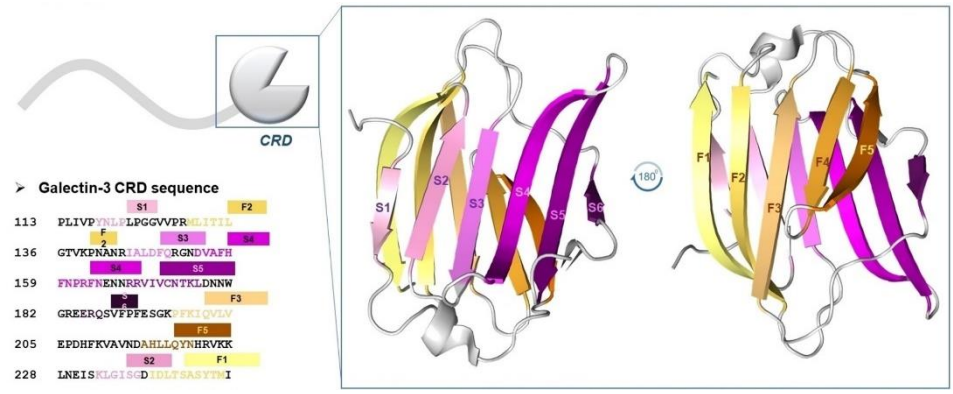


Figure 1.26. Galectin-3: A chimeric protein. The amino acid sequence and different three-dimensional views of Gal-3 CRD (*PDB 3ZSJ*). In the protein's crystal structure, the β-strands F1-F5 and S1-S6 of this domain are highlighted in orange and pink, respectively. *Reproduced from [98].*

The versatile roles of Galectin-3 are reflected in its widespread detection across various cell types, including activated macrophages, eosinophils, neutrophils, and mast cells, as well as in the epithelia of the gastrointestinal and respiratory tracts. Its functions involve vital biological processes such as cell growth, differentiation, apoptosis regulation, RNA splicing, and cell migration [101].

The multifunctionality of Gal-3 is further underscored by its variable subcellular localization; while primarily found in the cytoplasm, it can also be detected in the nucleus, on the cell surface, or within the extracellular environment [97]. In the cytosol, Gal-3 predominantly exerts an anti-apoptotic effect, whereas its presence in the extracellular compartment is linked to roles in cellular adhesion, organogenesis, immune system modulation, and tumorigenesis [102].

This protein is particularly significant in cancer, as its expression is often dysregulated in many tumors. It is involved in multiple stages of invasion and metastasis, including angiogenesis, cell-matrix interaction, dissemination through the bloodstream, and extravasation [103-104].

Notably, outside of mammalian systems, Gal-3 interacts with bacteria by recognizing bacterial glycans. This recognition occurs through a dual mechanism, involving either its C-terminal CRD or its non-carbohydrate-binding N-terminal domain [85].

Given its pervasive involvement in numerous disease processes, Gal-3 has become a compelling target for immunomodulation and tumor therapies. Its unique characteristic of existing as a monomer while retaining the capacity for aggregation provides ideal features for drug design [105]. Consequently, significant efforts have been directed towards synthesizing various Gal-3 inhibitors.

Strategies include the development of preorganized ligands, such as rigid histo-blood group antigens, which have demonstrated improved binding affinity to human Gal-3 due to favorable binding entropy [16]. Another effective approach involves introducing aromatic rings at the C3 position of the galactose residue in lactose, which enhances inhibitory potency by establishing additional favorable cation-π interactions with Gal-3 [105].

The optimal approach for the design of galectin inhibitors was the use of chemically modified natural galectin ligands, such as the disaccharides lactose (Lac) or *N*-acetyllactosamine (Lac*NA*c) (**Figure 1.27A and B**). Although Lac and Lac*NA*c are more promising scaffolds than monosaccharide galactose for the design of galectin inhibitors, they are both sensitive to enzymatic hydrolysis by glycosidases.

To overcome the poor bioavailability of lactose derivatives, thiosugars, and thiodigalactoside (TDG), have emerged as interesting building blocks for the design of potent galectin inhibitors, although with low selectivity (**Figure 1.27C**). TDG itself proved to be an intriguing ligand, as the substitution of the oxygen atom in the glycosidic linkage with a sulfur atom already led to improved enzymatic and hydrolytic stability [106]. Symmetrical modifications of TDG at C3 and C3′ positions introduce additional favorable protein-ligand interactions, resulting in increased affinity for galectins. A notable example is the monovalent inhibitor TD139 (Galecto Biotech, Copenhagen, Denmark) (**Figure 1.27D**) [94-95]. TD139 has demonstrated excellent affinity for Gal-3 (0.068 μM ± 0.01, determined by isothermal titration calorimetry), and has successfully completed Phase Ib/IIa clinical trials for idiopathic pulmonary fibrosis treatment [106].

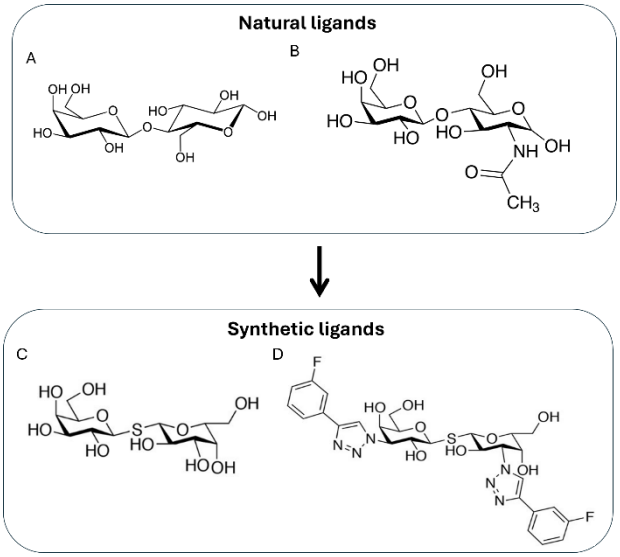


Figure 1.27. Developing Gal-3 inhibitors with enhanced affinity. From natural ligands: (A) Lac and (B) Lac*NA*c, to synthetic ligands: (C) TDG and (D) TD139.

Future efforts are increasingly focused on specifically designing inhibitors capable of selectively targeting and distinguishing between galectin members. The diverse activities of individual galectins in both normal physiological and pathological processes underscore the pressing need for potent and selective inhibitors [107]. However, the development of highly selective inhibitors presents a significant challenge due to the considerable structural similarities within the CRDs of different galectin members. For this reason, studies that provide detailed structural insights into galectin-ligand interactions are invaluable for the rational design of increasingly potent and selective inhibitors.

1.3 Objectives

In this PhD thesis, the molecular basis of human lectin-ligand interactions was investigated, focusing on two key proteins with significant roles in disease: Siglec-8 and Galectin-3. These lectins were expressed and purified in prokaryotic expression systems for detailed binding studies. A combination of biophysical and physicochemical techniques was employed, including NMR analysis from both the ligand and protein perspectives. The resulting data were integrated with computational studies to generate 3D models of the protein-ligand complexes.

Here is a summary of the chapters that comprise the work presented:

-Chapter 4: The first part of this thesis focused on Siglec-8, a protein that modulates immune responses in allergic and chronic inflammatory disorders, making it a key target for drug discovery. Starting with the known minimal glycan ligand (NeuAcα2-3(6-O-sulfo-Gal)), a series of sialylated triazole glycomimetics were designed and synthesized. The present study aimed to determine if triazoles, modified with sulfate, sulfonate, or carboxylic acid groups, could effectively replace the 6-O-sulfo-Gal moiety. Through a combination of NMR spectroscopy and ITC (Isothermal Titration Calorimetry), it was confirmed that the synthesized ligands were successfully bound to Siglec-8. The NMR analysis provided detailed insights into the molecular interactions between Siglec-8 and its glycomimetics. The findings reported provide a solid foundation for the future development of more potent Siglec-8 inhibitors.

-Chapter 5: The second part of this research shifted focus to Galectin-3, a β-galactoside-binding protein widely studied for its involvement in cancer progression and in fibrosis. Given the clinical need for effective inhibitors, the binding mechanism of a selenium-based Gal-3 inhibitor, a seleno-digalactoside modified with a benzyl group (SeDG-Bn), was investigated. To better understand

its high inhibitory efficacy at a molecular level, NMR spectroscopy was integrated with molecular dynamics simulations. The 3D model of the SeDG-Bn/Gal-3 complex, resulting from this combined approach, illustrates the structural basis for its enhanced binding affinity. The results of these analyses provide the foundation for the rational design of more potent and selective Galectin-3 inhibitors, thereby contributing to the development of new therapeutic approaches for conditions like cancer and fibrosis.

Chapter 2: Methodologies for studying glycan-lectin interactions

Understanding the intricate molecular interplay between proteins and their ligands is fundamental across all biological disciplines, particularly in drug discovery and disease pathology. A diverse repertoire of experimental and computational techniques has been developed to dissect these interactions, each offering unique insights into binding mechanisms, kinetics, thermodynamics, and structural details.

Experimental approaches range from high-resolution structural methods to functional assays. X-ray crystallography, NMR and Cryo-Electron Microscopy (Cryo-EM) [108] are powerful tools for providing atomic-level structures of protein-ligand complexes, revealing precise binding poses and conformational changes. ITC directly measures the heat changes associated with binding, providing comprehensive thermodynamic parameters like affinity, enthalpy, and entropy. Surface Plasmon Resonance (SPR) and Bio-Layer Interferometry (BLI) are label-free techniques used to monitor real-time binding events, yielding association and dissociation rates (k_a and k_d), which allow for the calculation of equilibrium dissociation constants (K_d). Other valuable methods include Fluorescence Spectroscopy and Mass Spectrometry (MS)-based techniques, all contributing to a comprehensive view of protein-ligand recognition.

In parallel, computational studies have become indispensable, complementing experimental data by offering atomic-level insights into dynamic processes and predictive capabilities. Techniques such as molecular docking predict potential binding poses of ligands within a protein's binding site, while molecular dynamics (MD) simulations calculate a time-resolved view of the dynamic behavior of the protein-ligand complex, including conformational changes, water networks, and interactions that govern the binding.

The present chapter aims to describe the principal techniques used for the thesis objectives, by focusing on NMR techniques [109] from both ligand and protein perspectives and molecular modeling methods.

2.1 NMR spectroscopy

NMR spectroscopy is a powerful technique extensively employed to investigate the conformation, dynamics, and interactions of glycans [110]. A wide array of NMR methodologies is available for this purpose. From a physical chemistry perspective, the binding of a protein to a ligand is characterized by a chemical exchange process between its free and bound states. This process is governed by the association rate constant (k_{on}) and the dissociation rate constant (k_{off}), as described in **Equation 2.1**.

The specific values of these rate constants dictate which NMR parameters and experiments are most suitable for monitoring the binding event and extracting crucial structural insights. Notably, the interaction can be observed from either the glycan's or the protein's perspective, by monitoring changes in their respective NMR parameters upon complex formation [111].

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

Equation 2.1. The equation illustrates the dynamic equilibrium maintained between three distinct species: the unbound protein, the unbound ligand, and the resulting protein-ligand complex. k_{on} represents the association rate constant, quantifying the speed at which the protein and ligand bind to form the complex, while k_{off} signifies the dissociation rate constant, indicating the rate at which the complex breaks apart into its free components.

In addition, the ratio (k_{off}/k_{on}) determines the rate constant K_D, as described in **Equation 2.2**:

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

The equilibrium dissociation constant (K_D) is a key determinant for classifying the chemical exchange between a molecule's free and bound states into two primary categories: slow exchange or fast exchange.

In the slow exchange regime, typically observed for K_D values around 10^{-5} M, the lifetime of the protein-ligand complex is sufficiently long so that it exceeds the timescale of the chemical shift difference between the free and bound forms. This temporal separation results in the distinct appearance of two separate NMR signals, each corresponding to one of the states.

Conversely, the fast exchange regime, often associated with lower K_D values, approximately 10^{-8} M, describes a situation where the interconversion between the free and bound forms occurs at a rate significantly faster than the chemical shift difference between these two states. In this scenario, the individual NMR signals fuse into a single, averaged peak, which represents a weighted average of the chemical shifts from both the free and bound environments [112] (**Figure 2.1**).

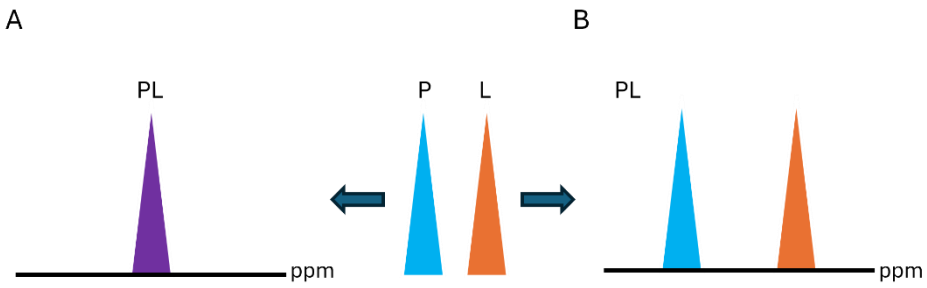


Figure 2.1. An illustration of NMR chemical exchange regimes in protein-ligand binding. (A) Fast exchange and (B) Slow exchange.

Since the efficacy of certain NMR techniques depends on the protein-ligand exchange regime, the meticulous selection of the most appropriate NMR experiment to investigate a specific protein-ligand interaction is crucial [113].

It is worth noting that most screening methods based on NMR have proven highly effective in identifying small molecules that interact with macromolecular receptors. These screening processes can be broadly categorized into either ligand-based or protein-based approaches, depending on whether the NMR signals of the ligand or the receptor are being primarily monitored.

2.1.1 Ligand-based NMR techniques for ligand characterization

Beyond the application of ligand-based NMR methods for exploring protein-ligand interactions, a prerequisite is the comprehensive NMR assignment of the ligand itself. This crucial step is achieved by combining both one-dimensional (1D) and two-dimensional (2D) NMR spectra.

Specifically, ^{1}H and ^{13}C NMR experiments are invaluable for elucidating various structural features of monosaccharides, including their composition, the anomeric

configuration (whether α or β), the specific substitution pattern, and the presence and nature of any non-glycidic substituents.

^{1}H NMR experiments yield critical information on signal *multiplicity* and scalar interactions between vicinal and geminal nuclei. These through-bond interactions are quantified by the coupling constant (J value), which serves as a powerful diagnostic tool for structural insights.

For example, a $^{3}J_{H1,H2}$ coupling constant exceeding 8 Herz (Hz) is typically indicative of a β-configured pyranose ring with either *gluco*- or *galacto*-configuration, whereas a value below 3 Hz is characteristic of an α-configured sugar.

Furthermore, the magnitude of the $^{1}J_{C1,H1}$ coupling constant provides a reliable indicator of anomeric configuration across a broad range of sugar moieties: values below 170 Hz suggest a β-anomer, while those above 170 Hz are diagnostic of an α-anomer.

The characteristic chemical shift (δ) regions for ^{1}H and ^{13}C are summarized in **Table 2.1**.

Table 2.1. ^{1}H and ^{13}C chemical shift values relevant for carbohydrates.

δ (ppm)	**^{1}H**
8.5 – 7.5	Ammide resonances
5.5 – 4.2	Anomeric protons
4.5 – 2.8	Sugar ring protons
2.6 – 1.8	α-methylene protons of deoxy sugars
1.0 – 2.0	Methyl protons of 6-deoxy sugars and of the acetyl groups
2.2 – 0.8	β, γ, ω protons of methyl and distal methylene groups
δ (ppm)	**^{13}C**
160 - 180	Carbonyl carbons
95 - 105	Anomeric carbons
60 – 80	Sugar ring carbons
45 - 60	Nitrogen bearing carbon signals
~ 30	Aliphatic methylene carbons of deoxy sugars
20 - 17	Methyl carbons of deoxy sugars, acetyl groups

To establish the complete sugar sequence, a suite of 2D NMR experiments is indispensable. Homogenous nuclear experiments, such as COSY (Correlation Spectroscopy) and TOCSY (Total Correlation SpectroscopY), are employed to correlate chemical shifts (resonances) of ^{1}H nuclei. COSY reveals correlations between geminal and vicinal protons, while TOCSY extends this to correlate protons within the entire spin network of a monosaccharide unit. NOESY (Nuclear

Overhauser Effect SpectroscopY) is utilized to measure spin cross-relaxation rates, thereby identifying protons that are spatially close to each other. When NOE signals are close to zero (as can occur for small oligosaccharides like many trisaccharides), ROESY (Rotating-frame Overhauser Effect SpectroscopY), which yields consistently positive signals due to spin-locked conditions, is often used as an alternative.

For carbon-proton correlations, HSQC (Heteronuclear Single Quantum Correlation) provides direct correlations between directly coupled ^{13}C and ^{1}H nuclei.

Following comprehensive ligand assignment, ligand-based NMR methods are then performed to study interaction. A significant advantage of this approach is its independence from the molecular weight of the receptor molecule.

However, the success of ligand-based NMR experiments depends on the exchange-mediated transfer of bound state information to the free state. Consequently, a critical requirement for these methods is that the protein-ligand interaction must fall within the fast exchange regime [114].

2.1.1.1 Saturation transfer difference (STD) NMR

Saturation Transfer Difference (STD) NMR spectroscopy is a most popular and versatile method for characterizing the specific binding region (epitope) on a ligand when in complex with a receptor in solution [115]. This technique relies on the transfer of magnetization from the protein to the ligand.

This method is typically acquired as a pseudo-2D experiment, generated by subtracting two one-dimensional NMR spectra [116]. The first spectrum, known as the *off-resonance* spectrum, is obtained by irradiating a spectral region far from

both protein and ligand signals (commonly around 100 ppm) to serve as a reference. The second, or *on-resonance* spectrum, involves selectively saturating protein signals (usually in the aromatic or aliphatic regions) using a low-power radio frequency pulse train applied over several seconds (referred to as *saturation time*) [117].

During the ligand's permanency within the protein's binding pocket, this saturation is transferred from the receptor to the interacting ligand via intermolecular saturation transfer. Critically, during the *on-resonance* period, the spin-lattice relaxation (R1) of the free ligand is slower than the dissociation rate constant (K_{off}).

The final STD spectrum is the result of subtracting the *off-resonance* data from the *on-resonance* data. In particular, ligand protons in intimate contact with the protein's binding site will receive greater magnetization through intermolecular ^{1}H-^{1}H cross-relaxation mechanisms. This results in amplified STD signals for these protons. Conversely, ligand protons situated far from the protein will receive minimal or no saturation, consequently exhibiting low or absent STD signals (**Figure 2.2**) [115].

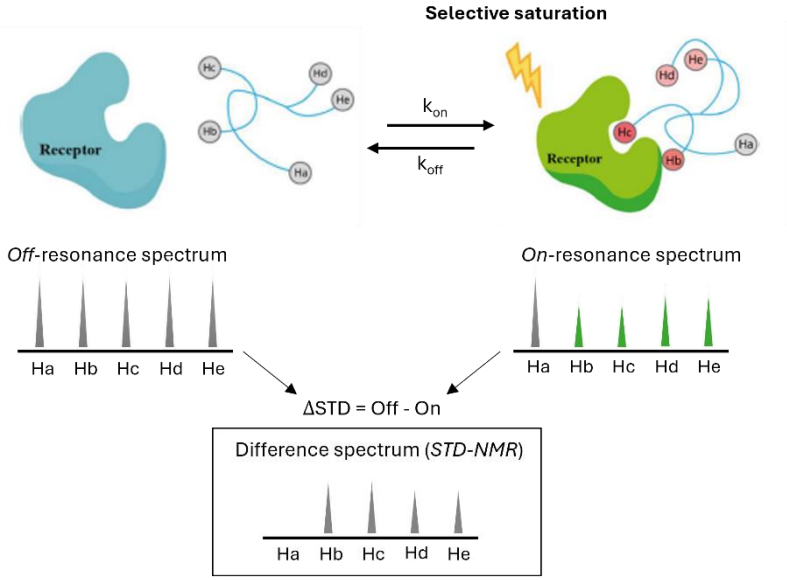


Figure 2.2. Schematic representation of STD NMR technique. The *off-resonance* spectrum, acquired by irradiating away from both protein and ligand signals, shows no intensity reduction in the ligand signals (Ha-Hb-Hc-Hd-He). In contrast, the *on-resonance* spectrum involves selective saturation of the receptor using radiofrequency (RF) pulses; magnetization is then transferred to ligand protons via intermolecular NOE. Consequently, only ligand signals from protons involved in the binding process exhibit STD signals in the resulting spectrum. *Adapted from [115]*.

The intensity of STD signals can be quantitatively determined using the following formula (**Equation 2.3**):

$$I_{STD} = \frac{I_0 - Isat}{I_0}$$

where I_0 is the signal intensity from the reference (*off-resonance*) experiment, and I_{sat} represents the signal intensity from the saturated (*on-resonance*) experiment. By analyzing the saturation degree of individual ligand protons, the precise

interacting epitope on the ligand can be elucidated. Once the most pronounced STD signal is normalized to 100%, the relative STD percentages for all other interacting protons can be calculated.

STD NMR emerges as an invaluable tool for elucidating the binding epitope of a ligand when engaged with a macromolecule. This capability is exceptionally beneficial in translational research, such as the design of novel therapeutic agents like small-molecule drugs or mimetics. Notably, the technique requires low concentration of the macromolecule (in the micromolar range), but critically, a significant molar excess of the ligand is required (usually higher of 1:50 protein: ligand molar ratio).

2.1.2 Protein-based NMR techniques for protein characterization

2.1.2.1 Chemical shift perturbation (CSP): mapping protein-ligand interactions

The presence of a ligand within a protein's binding site directly or indirectly alters the chemical environment of the involved amino acids. Since the chemical shift, a fundamental NMR parameter, is sensitive to this chemical environment, ligand binding affects the chemical shifts of protein nuclei. These perturbations are most pronounced for residues directly engaged in binding but can also extend to those that exhibit conformational adjustments or manifest additional motions upon interaction. This phenomenon is termed Chemical Shift Perturbation (CSP) [118-119].

The detection and evaluation of CSP upon ligand addition are most effectively performed using ^{1}H-^{15}N correlation experiments. Specifically, ^{1}H,^{15}N-HSQC (Heteronuclear Single-Quantum Coherence) spectra function as a unique

"fingerprint" for any protein [120-121]. The cross-peaks within this spectrum represent the amide NH groups of the protein backbone and, in some instances, the lateral chains (e.g., Lys, Arg, His, Trp). Given the unique chemical environment defined by a protein's sequence and 3D structure, the resulting HSQC spectrum is highly characteristic and distinct for each protein, reflecting its distinctive dispersion and specific $^{1}H/^{15}N$ chemical shifts.

A critical prerequisite for accurate CSP analysis is that both the protein and ligand are dissolved in the same buffer, and all measurements throughout the titration are acquired under identical conditions (e.g., temperature, pH). This process is crucial because chemical shifts, particularly those of amide protons, are highly sensitive to these experimental parameters.

The CSP analysis of a protein-ligand complex starts with acquiring a reference $^{1}H,^{15}N$-HSQC spectrum of the apo (ligand-free) protein. Subsequently, a series of $^{1}H,^{15}N$-HSQC spectra are recorded as increasing concentrations of the ligand are added, ideally until the protein's binding site achieves complete saturation. The ligand's presence alters the chemical environment of nuclei within the binding pocket, leading to observable CSPs in the corresponding peaks.

Depending on the k_{off} of the interaction, the behavior of NMR peaks during titration varies (**Figure 2.3**). In the *fast exchange* limit, where the chemical exchange rate between free and bound states is rapid relative to the chemical shift timescale, the $^{1}H,^{15}N$ cross-peaks gradually shift from their initial free position towards the final position corresponding to the fully saturated protein. The chemical shift observed represents a population-averaged value between the free and bound forms, leading to a linear shift as ligand concentration increases. Conversely, in the *slow exchange* rate, two distinct peaks are simultaneously observed in the HSQC spectrum: one for the free state and a separate one for the bound state. Their relative intensities are dependent on the binding affinity and the

amount of ligand added. As the titration proceeds, the peak corresponding to the free state gradually decreases in intensity, while that for the bound state increases, until the free state's peak eventually disappears [119].

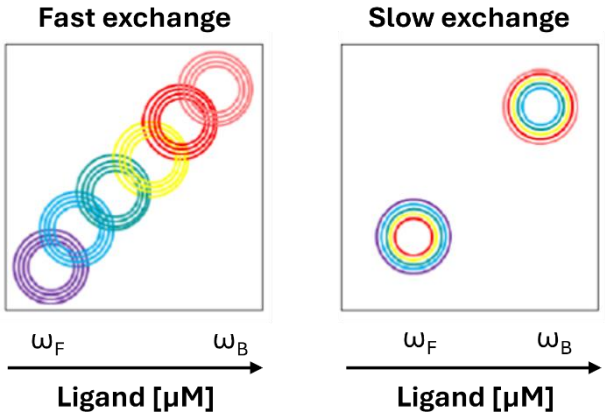


Figure 2.3. Chemical shift perturbation in protein-based NMR experiments. *Fast exchange*: increasing ligand concentrations cause the ^{15}N and ^{1}H frequencies to undergo a gradual chemical shift perturbation, transitioning smoothly from the unbound (ω_F) to the bound (ω_B) peak position. *Slow exchange*: the HSQC spectrum simultaneously displays two distinct cross-peaks: one for the free protein and another for the bound state.

While most chemical shift changes directly delineate the protein's binding site, it's important to note that conformational changes in amino acids can also induce shifts in resonance frequencies. Thus, a peak shift is not always only indicative of direct proximity to the binding interface; it can also provide information about allosteric changes within the protein's structure upon ligand binding.

The range of protein and ligand concentrations used for titration is selected to ensure a wide spectrum of fractional saturations, ideally achieving full saturation

of the protein by the ligand. Thus, based on the equilibrium presented in prior section (**Equation 2.1**):

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

from the established total concentrations of the ligand and protein, binding affinities can be derived using these basic algebraic equations (**Equations 2.4** and **2.5**):

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

The observed chemical shifts (**Equation 2.6**) are determined by the shifts of the free and bound states, in conjunction with the species' respective molar fractions:

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

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

And then (**Equation 2.7**):

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

Here, $\Delta\delta_{obs}$ represents the measured chemical shift change relative to the free state, while $\Delta\delta_{max}$ denotes the maximum possible shift upon saturation (typically determined through fitting, as direct experimental measurement is often impractical). This equation enables the determination of K_D by fitting the observed chemical shift values obtained across various protein and ligand concentrations. Standard software facilitates this fitting process.

2.2 Computational approaches

Computational tools are essential for studying biomolecular systems, especially proteins and glycans in free and bound states. There is a range of software available that can be used to prepare input files for proteins and glycans, to run molecular dynamics simulations and then analyzed the results (**Figure 2.4**) [122].

These computational resources, which will be explored further in subsequent sections, offer advanced methods for building, visualizing, and predicting protein/glycan structures, modeling protein-ligand complexes, and interpreting molecular dynamics outcomes [122]. Ultimately, this approach highlights the critical role of computational tools in protein-glycan research, demonstrating their power to provide significant insights into the binding kinetics, energetics, and structural factors that govern specific molecular interactions.

Beyond the field of computational and bioinformatic tools, which are designed to guide the elucidation of structural and conformational details and apply molecular dynamics simulations to study proteins and glycans in free and bound states, 3D complexes have also been extensively investigated using highly advanced and adaptable techniques like NMR, X-ray crystallography, and MS [14, 123-124]. These experimental techniques have been developed to acquire comprehensive information about the structural and conformational characteristics of glycans and proteins. This thesis will be focused on the application of molecular docking and dynamics simulations to integrate with experimental data in the study of protein-glycan interaction processes.

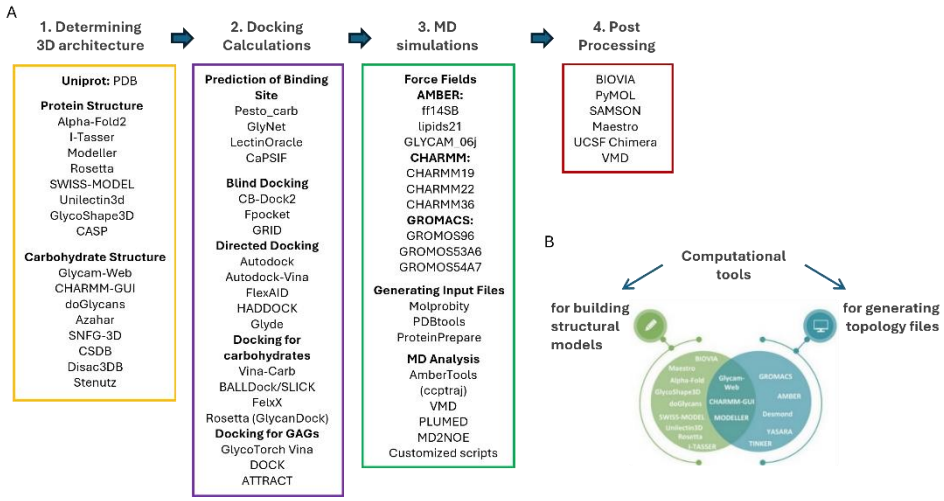


Figure 2.4 Overview of computational approaches used to study protein-glycan interactions. (A) The essential steps involved in computational studies include the construction or selection of suitable 3D structures for glycans and proteins, the modeling of their complex, the performance and analysis of molecular dynamics (MD) simulations, and the subsequent processing and visualization of the simulation outcomes. (B) The various computational tools, available for generating structural models and topology files for both glycans and proteins. *Adapted from [122].*

Molecular docking is a widely employed computational technique in structure-based drug design. Its primary utility is based on its ability to predict the binding conformation of a small molecule within a protein's specific binding site, thereby offering a model of their interaction.

Typically, computational docking protocols aim to identify the various potential binding poses of a small molecule within a given receptor, adjusting the ligand's shape to achieve the most favorable fit. The subsequent crucial step involves scoring the energy of these resulting binding poses, essentially evaluating the energy of the ligand-target complex.

To conduct a docking calculation, the 3D structure of the protein of interest is indispensable. This structure can be obtained through experimental methods, such as X-ray crystallography or NMR spectroscopy, or it can be derived from homology modeling techniques or prediction such as AlphaFold. Information regarding the protein binding site is an important, though not essential, requirement. For instance, in this thesis, docking calculations were performed using CB-Dock2 [125] to automatically predict binding modes without information about binding sites, followed by refined docking with AutoDock-Vina [126].

Molecular Dynamics (MD) simulations stand as a key tool in the computational investigation of biological molecules, offering intricate details on their dynamics and conformational shifts, both individually and within complexes.

MD simulations are a powerful tool for probing the behavior of carbohydrates, both free and bound. They are frequently used in conjunction with NMR data (such as NOE and residual dipolar coupling-based experiments) to provide a thorough description of ligand conformational behavior and to build validated 3D models of protein-glycan complexes. Following an overview of the key molecular modeling tools utilized throughout this thesis, **Figure 2.5** provides a concise summary of these techniques.

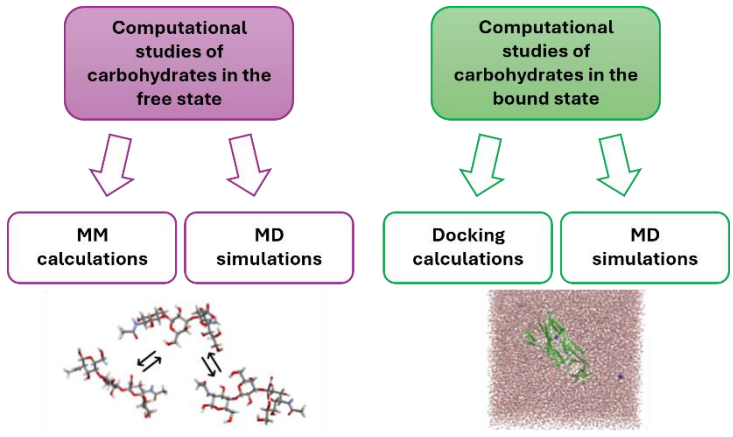


Figure 2.5. Computational methodologies commonly used for the study of carbohydrates and their interaction with receptors. The diagram shows the techniques used for analyzing carbohydrates in the free state (MM calculations and MD simulations) and in the bound state (Docking calculations and MD simulations).

Section I - References

1. D.M. Oswald, B.A. Cobb. “Emerging glycobiology tools: A renaissance in accessibility.” *Cellular immunology*, 333 (2018) 2-8.
2. L. Möckl. “The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation.” *Frontiers in Cell and Development Biology*, 8 (2020) 253.
3. Y. Xie, P. Chai, N.A. Till, H. Hemberger, C.G. Lebedenko, J. Porat, C.P. Watkins, R.M. Caldwell, B.M. George, J. Perr, C.R. Bertozzi, B.A. Garcia, R.A. Flynn. “The Modified RNA Base Acp3U Is an Attachment Site for N-Glycans in GlycoRNA.” *Cell*, 187 (2024) 5228-5237.e12.
4. A. Varki. “Biological Roles of Glycans.” *Glycobiology*, 27 (2017) 3-49.
5. A.F. Carlin, S. Uchiyama, Y.C. Chang, A.L. Lewis, V. Nizet, A. Varki. “Molecular Mimicry of Host Sialylated Glycans Allows a Bacterial Pathogen to Engage Neutrophil Siglec-9 and Dampen the Innate Immune Response.” *Blood*, 113 (2009) 3333-3336.
6. S.S. Pinho, I. Alves, J. Gaifem, G.A. Rabinovich. “Immune Regulatory Networks Coordinated by Glycans and Glycan-Binding Proteins in Autoimmunity and Infection.” *Cellular & Molecular Immunology*, 20 (2023) 1101-1113.
7. K.T. Schjoldager, Y. Narimatsu, H.J. Joshi, H. Clausen. “Global View of Human Protein Glycosylation Pathways and Functions.” *Nature Reviews Molecular Cell Biology*, 21 (2020) 729-749.
8. C. Reily, T.J. Stewart, M.B. Renfrow, J. Novak. “Glycosylation in Health and Disease.” *Nature Reviews Nephrology*, 15 (2019) 346-366.
9. J.B. Goh, S.K. Ng. “Impact of Host Cell Line Choice on Glycan Profile.” *Critical Reviews in Biotechnology*, 38 (2018) 851-867.

10. P. Nair-Gupta, M. Diem, D. Reeves, W. Wang, R. Schulingkamp, K. Sproesser, B. Mattson, B. Heidrich, M. Mendonça, J. Joseph, J. Sendecki, B. Foulk, G. Chu, D. Fink, Q. Jiao, S.J. Wu, K. Packman, Y. Elsayed, R. Attar, F. Gaudet. "A Novel C2 Domain Binding CD33xCD3 Bispecific Antibody with Potent T-Cell Redirection Activity Against Acute Myeloid Leukemia." *Blood Advances*, 4 (2020) 906-919.
11. D. Dutta, C. Mandal, C. Mandal. "Unusual Glycosylation of Proteins: Beyond the Universal Sequon and Other Amino Acids." *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1861 (2017) 3096-3108.
12. P. Fisher, J. Thomas-Oates, A.J. Wood, D. Ungar. "The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus." *Frontiers in Cell and Developmental Biology*, 7 (2019) 1-11.
13. J.Y. Zhou, B.A. Cobb. "Glycans in Immunologic Health and Disease." *Annual Review of Immunology*, 39 (2021) 511-536.
14. A. Gimeno, P. Valverde, A. Ardá, J. Jiménez-Barbero. "Glycan Structures and Their Interactions with Proteins. A NMR View." *Current Opinion in Structural Biology*, 62 (2020) 22-30.
15. C.J. Schultz. "Glycoproteins, Plant." In *Encyclopedia of Biological Chemistry (Second Edition)*, Academic Press (2013) 465-468.
16. A. Valverde, A. Ardà, N.C. Reichardt, J. Jiménez-Barbero, A. Gimeno. "Glycans in Drug Discovery." *Medicinal Chemistry Communications*, 10 (2019) 1678.
17. T.H. Mogensen. "Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses." *Clinical Microbiology Reviews*, 22 (2009) 240-273.
18. S.S. Pinho, C.A. Reis. "Glycosylation in Cancer: Mechanisms and Clinical Implications." *Nature Reviews Cancer*, 15 (2015) 540-555.

19. H.H. Khan, B. Shi, Y. Tian, T. Wang, S. Hussain, F.U. Khan, Z. Khan, B. Ashfaq, H.A.T. Ahmad. “Glycan Regulation in Cancer, Nervous and Immune System: A Narrative Review.” *Biomedical Research and Therapy*, 4 (2019) 3113-3120.
20. A. Varki, P. Gagneux. “Biological Functions of Glycans.” In A. Varki et al. (Eds.), *Essentials of Glycobiology*, 3 (2017) 77-88.
21. M.E. Taylor, K. Drickamer, R.L. Schnaar, M.E. Etzler, A. Varki. “Discovery and Classification of Glycan-Binding Proteins.” In A. Varki, R.D. Cummings, J.D. Esko et al. (Eds.), *Essentials of Glycobiology*, 3rd ed. Cold Spring Harbor Laboratory Press (2017).
22. M. Scur, B.D. Parsons, S. Dey, A.P. Makrigiannis. “The Diverse Roles of C-type Lectin-Like Receptors in Immunity.” *Frontiers in Immunology*, 14 (2023) 1126043.
23. F.T. Liu, S.R. Stowell. “The Role of Galectins in Immunity and Infection.” *Nature Reviews Immunology*, 23 (2023) 479-494.
24. M.S. Macauley, P.R. Crocker, J.C. Paulson. “Siglec-Mediated Regulation of Immune Cell Function in Disease.” *Nature Reviews Immunology*, 14 (2014) 653-666.
25. B. Lepenies, R. Lang. “Editorial: Lectins and Their Ligands in Shaping Immune Responses.” *Frontiers in Immunology*, 10 (2019) 2379.
26. M.P. Lenza, U. Atxabal, I. Oyenarte, J. Jiménez-Barbero, J. Ereño-Orbea. “Current Status on Therapeutic Molecules Targeting Siglec Receptors.” *Cells*, 9 (2020) 2691.
27. T. Angata, A. Varki. “Chemical Diversity in the Sialic Acids and Related α-Keto Acids: An Evolutionary Perspective.” *Chemical Reviews*, 102 (2002) 439-469.
28. S. Pillai, I.A. Netravali, A. Cariappa, H. Mattoo. “Siglecs and Immune Regulation.” *Annual Review of Immunology*, 30 (2012) 357-392.

29. H. Arase, L.L. Lanier. “Specific Recognition of Virus-Infected Cells by Paired NK Receptors.” *Reviews in Medical Virology*, 14 (2004) 83-93.
30. P.R. Crocker, J.C. Paulson, A. Varki. “Siglecs and Their Roles in the Immune System.” *Nature Reviews Immunology*, 7 (2007) 255-266.
31. P.R. Crocker. “Siglecs in Innate Immunity.” *Current Opinion in Pharmacology*, 5 (2005) 431-437.
32. R.H. Quarles. “Myelin-Associated Glycoprotein (MAG): Past, Present and Beyond.” *Journal of Neurochemistry*, 100 (2007) 1431-1448.
33. A.S. O'Neill, T.K. van den Berg, G.E. Mullen. “Sialoadhesin - A Macrophage-Restricted Marker of Immunoregulation and Inflammation.” *Immunology*, 138 (2013) 198-207.
34. S. Kelm, R. Schauer, J.C. Manuguerra, H.J. Gross, P.R. Crocker. “Modifications of Cell Surface Sialic Acids Modulate Cell Adhesion Mediated by Sialoadhesin and CD22.” *Glycoconjugate Journal*, 11 (1994) 576-585.
35. L.D. Powell, D. Sgroi, E.R. Sjoberg, I. Stamenkovic, A. Varki. “Natural Ligands of the B Cell Adhesion Molecule CD22β Carry N-Linked Oligosaccharides with α-2,6-Linked Sialic Acids That Are Required for Recognition.” *Journal of Biological Chemistry*, 268 (1993) 7019-7027.
36. B. E. Collins, M. Kiso, A. Hasegawa, M. B. Tropak, J. C. Roder, P. R. Crocker, R. L. Schnaar. “Binding Specificities of the Sialoadhesin Family of I-Type Lectins: Sialic Acid Linkage and Substructure Requirements for Binding of Myelin-Associated Glycoprotein, Schwann Cell Myelin Protein, and Sialoadhesin.” *Journal of Biological Chemistry*, 272 (1997) 16889-16895.
37. H. Begleiter, B. Porjesz. “Evoked Brain Potentials as Indicators of Decision-Making.” *Science*, 187 (1975) 754-755.

38. T. Yamaji, T. Teranishi, M.S. Alphey, P.R. Crocker, Y. Hashimoto. "A Small Region of the Natural Killer Cell Receptor, Siglec-7, Is Responsible for Its Preferred Binding to α2,8-Disialyl and Branched α2,6-Sialyl Residues. A Comparison with Siglec-9." *Journal of Biological Chemistry*, 277 (2002) 6324-6332.
39. E.M. Rapoport, G.V. Pazynina, M.A. Sablina, P.R. Crocker, N.V. Bovin. "Probing Sialic Acid Binding Ig-Like Lectins (Siglecs) with Sulfated Oligosaccharides." *Biochemistry (Moscow)*, 71 (2006) 496-504.
40. M. A. Campanero-Rhodes, R. A. Childs, M. Kiso, S. Komba, C. Le Narvor, J. Warren, D. Otto, P. R. Crocker, T. Feizi. "Carbohydrate Microarrays Reveal Sulphation as a Modulator of Siglec Binding." *Biochemical and Biophysical Research Communications*, 344 (2006) 1141-1146.
41. H. Tateno, P.R. Crocker, J.C. Paulson. "Mouse Siglec-F and Human Siglec-8 Are Functionally Convergent Paralogs That Are Selectively Expressed on Eosinophils and Recognize 6′-Sulfo-Sialyl Lewis X as a Preferred Glycan Ligand." *Glycobiology*, 15 (2005) 1125-1135.
42. C. Di Carluccio, R.E. Forgione, A. Molinaro, P.R. Crocker, R. Marchetti, A. Silipo. "Exploring the Fascinating World of Sialoglycans in the Interplay with Siglecs." In *Carbohydrate Chemistry, Volume 45*. Royal Society of Chemistry (2020) 31-55.
43. A. Varki. "Since There Are PAMPs and DAMPs, There Must Be SAMPs? Glycan "Self-Associated Molecular Patterns" Dampen Innate Immunity, but Pathogens Can Mimic Them." *Glycobiology*, 21 (2011) 1121-1124.
44. S. Varchetta, D. Mele, A. Lombardi, B. Oliviero, S. Mantovani, C. Tinelli, M. Spreafico, D. Prati, S. Ludovisi, G. Ferraioli, C. Filice, A. Aghemo, P. Lampertico, F. Facchetti, F. Bernuzzi, P. Invernizzi, M. U. Mondelli. "Lack of Siglec-7 Expression Identifies a Dysfunctional Natural Killer Cell Subset

Associated with Liver Inflammation and Fibrosis in Chronic HCV Infection." *Gut*, 65 (2016) 1998-2006.

45. H. Attrill, H. Takazawa, S. Witt, S. Kelm, R. Isecke, R. Brossmer, T. Ando, H. Ishida, M. Kiso, P. R. Crocker, D. M. van Aalten. "The Structure of Siglec-7 in Complex with Sialosides: Leads for Rational Structure-Based Inhibitor Design." *Biochemical Journal*, 397 (2006) 271-278.
46. A.P. May, R.C. Robinson, M. Vinson, P.R. Crocker, E.Y. Jones. "Crystal Structure of the N-Terminal Domain of Sialoadhesin in Complex with 3′ Sialyllactose at 1.85 Å Resolution." *Molecular Cell*, 1 (1998) 173-182.
47. M. F. Pronker, S. Lemstra, J. Snijder, A. J. Heck, D. M. Thies-Weesie, R. J. Pasterkamp, B. J. Janssen. "Structural Basis of Myelin-Associated Glycoprotein Adhesion and Signalling." *Nature Communications*, 7 (2016) 13584.
48. M.A. Zhuravleva, K. Trandem, P.D. Sun. "Structural Implications of Siglec-5-Mediated Sialoglycan Recognition." *Journal of Molecular Biology*, 375 (2008) 437-447.
49. M.S. Alphey, H. Attrill, P.R. Crocker, D.M.F. Van Aalten. "High Resolution Crystal Structures of Siglec-7. Insights into Ligand Specificity in the Siglec Family." *Journal of Biological Chemistry*, 278 (2003) 3372-3377.
50. J.M. Propster, F. Yang, S. Rabbani, B. Ernst, F.H.-T. Allain, M. Schubert. "Structural Basis for Sulfation-Dependent Self-Glycan Recognition by the Human Immune-Inhibitory Receptor Siglec-8." *Proceedings of the National Academy of Sciences of the United States of America*, 113 (2016) E4170–E4179.
51. J. Ereño-Orbea, T. Sicard, H. Cui, M. T. Mazhab-Jafari, S. Benlekbir, A. Guarné, J. L. Rubinstein, J. P. Julien. "Molecular Basis of Human CD22 Function and Therapeutic Targeting." *Nature Communications*, 8 (2017) 764.

52. S. Duan, J.C. Paulson. “Siglecs as Immune Cell Checkpoints in Disease.” *Annual Review of Immunology*, 38 (2020) 365-395.
53. M.K. O’Reilly, J.C. Paulson. “Siglecs as Targets for Therapy in Immune-Cell-Mediated Disease.” *Trends in Pharmacological Sciences*, 30 (2009) 240-248.
54. A. Bärenwaldt, H. Läubli. “The Sialoglycan-Siglec Glyco-Immune Checkpoint - A Target for Improving Innate and Adaptive Anti-Cancer Immunity.” *Expert Opinion on Therapeutic Targets*, 23 (2019) 943-956.
55. J.A. O’Sullivan, D.J. Carroll, Y. Cao, A.N. Salicru, B.S. Bochner. “Leveraging Siglec-8 Endocytic Mechanisms to Kill Human Eosinophils and Malignant Mast Cells.” *Journal of Allergy and Clinical Immunology*, 141 (2018) 1774-1785.e7.
56. B.S. Bochner. “Siglec”ting the Allergic Response for Therapeutic Targeting.” *Glycobiology*, 26 (2016) 546-552.
57. T. Kiwamoto, N. Kawasaki, J.C. Paulson, B.S. Bochner. “Siglec-8 as a Drugable Target to Treat Eosinophil and Mast Cell-Associated Conditions.” *Pharmacology & Therapeutics*, 135 (2012) 327-336.
58. B. A. Youngblood, E. C. Brock, J. Leung, R. Falahati, B. S. Bochner, H. S. Rasmussen, K. Peterson, C. Bebbington, N. Tomasevic. “Siglec-8 Antibody Reduces Eosinophils and Mast Cells in a Transgenic Mouse Model of Eosinophilic Gastroenteritis.” *JCI Insight*, 4 (2019) 126219.
59. H. Floyd, J. Ni, A.L. Cornish, Z. Zeng, D. Liu, K.C. Carter, J. Steel, P.R. Crocker. “Siglec-8. A novel eosinophil-specific member of the immunoglobulin superfamily.” *Journal of Biological Chemistry*, 275 (2000) 861-866.
60. M.P. Lenza, U. Atxabal, C. Nycholat, I. Oyenarte, A. Franconetti, J.I. Quintana, S. Delgado, R. Núñez-Franco, C.T. Garnica Marroquín, H. Coelho, L. Unione, G. Jiménez-Oses, F. Marcelo, M. Schubert, J.C. Paulson, J. Jiménez-Barbero, J. Ereño-Orbea. "Structures of the Inhibitory Receptor

Siglec-8 in Complex with a High-Affinity Sialoside Analogue and a Therapeutic Antibody." *JACS Au*, 3 (2022) 204-215.

61. B. S. Bochner, R. A. Alvarez, P. Mehta, N. V. Bovin, O. Blixt, J. R. White, R. L. Schnaar. "Glycan array screening reveals a candidate ligand for Siglec-8." *Journal of Biological Chemistry*, 280 (2005) 4307-4312.
62. B. A. Youngblood, J. Leung, R. Falahati, J. Williams, J. Schanin, E. C. Brock, B. Singh, A. T. Chang, J. A. O'Sullivan, R. P. Schleimer, N. Tomasevic, C. R. Bebbington, B. S. Bochner. "Discovery, Function, and Therapeutic Targeting of Siglec-8." *Cells*, 10 (2020) 1-14.
63. M. Krystel-Whittemore, K.N. Dileepan, J.G. Wood. "Mast cell: A multi-functional master cell." *Frontiers in Immunology*, 6 (2016) 1-12.
64. W. Korver, A. Wong, S. Gebremeskel, G.L. Negri, J. Schanin, K. Chang, J. Leung, Z. Benet, T. Luu, E.C. Brock, K. Luehrsen, A. Xu, B.A. Youngblood. "The Inhibitory Receptor Siglec-8 Interacts with FcεRI and Globally Inhibits Intracellular Signaling in Primary Mast Cells Upon Activation." *Frontiers in Immunology*, 13 (2022) 833728.
65. B. A. Youngblood, E. C. Brock, J. Leung, R. Falahati, P. J. Bryce, J. Bright, J. Williams, L. D. Shultz, D. L. Greiner, M. A. Brehm, C. Bebbington, N. Tomasevic. "AK002, a Humanized Sialic Acid-Binding Immunoglobulin-Like Lectin-8 Antibody that Induces Antibody-Dependent Cell-Mediated Cytotoxicity against Human Eosinophils and Inhibits Mast Cell-Mediated Anaphylaxis in Mice." *International Archives of Allergy and Immunology*, 180 (2019) 91-102.
66. E. S. Dellon, K. A. Peterson, J. A. Murray, G. W. Falk, N. Gonsalves, M. Chehade, R. M. Genta, J. Leung, P. Khoury, A. D. Klion, S. Hazan, M. Vaezi, A. C. Bledsoe, S. R. Durrani, C. Wang, C. Shaw, A. T. Chang, B. Singh, A. P. Kamboj, H. S. Rasmussen, M. E. Rothenberg, I. Hirano. "Anti-Siglec-8

Antibody for Eosinophilic Gastritis and Duodenitis." *New England Journal of Medicine*, 383 (2020) 1624-1634.

67. S. Altrichter, P. Staubach, M. Pasha, H. Rasmussen, B. Singh, A. Chang, A. Kamboj, J. Bernstein, F. Siebenhaar, M. Maurer. "P152 Clinical Activity of AK002, an Anti-Siglec-8 Antibody, in Multiple Forms of Uncontrolled Chronic Urticaria." *Annals of Allergy, Asthma & Immunology*, 123 (2019) S27-S28.
68. R.T. Lee, Y.C. Lee. "Affinity enhancement by multivalent lectin carbohydrate interaction." *Glycoconjugate Journal*, 17 (2000) 543-551.
69. T.K. Dam, T.A. Gerken, C.F. Brewer. "Thermodynamics of multivalent carbohydrate-lectin cross-linking interactions: importance of entropy in the bind and jump mechanism." *Biochemistry*, 48 (2009) 3822-3827.
70. O. Blixt, B.E. Collins, I.M. van den Nieuwenhof, P.R. Crocker, J.C. Paulson. "Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein." *Journal of Biological Chemistry*, 278 (2003) 31007-31019.
71. R. Hevey. "Strategies for the development of glycomimetic drug candidates." *Pharmaceuticals*, 12 (2019) 0055.
72. S. Kelm, R. Brossmer, R. Isecke, H.J. Gross, K. Strenge, R. Schauer. "Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues." *European Journal of Biochemistry*, 255 (1998) 663-672.
73. S. M. van Rossenberg, L. A. Sliedregt, R. Autar, C. Piperi, A. P. Van Der Merwe, T. J. van Berkel, J. Kuiper, E. A. Biessen. "A Structure-Function Study of Ligand Recognition by CD22β." *Journal of Biological Chemistry*, 276 (2001) 12967-12973.
74. S. Kelm, J. Gerlach, R. Brossmer, C.P. Danzer, L. Nitschke. "The ligand-binding domain of CD22 is needed for inhibition of the B cell receptor signal,

as demonstrated by a novel human CD22-specific inhibitor compound." *Journal of Experimental Medicine,* 195 (2002) 1207-1213.

75. N. R. Zaccai, K. Maenaka, T. Maenaka, P. R. Crocker, R. Brossmer, S. Kelm, E. Y. Jones. "Structure-guided design of sialic acid-based Siglec inhibitors and crystallographic analysis in complex with sialoadhesin." *Structure*, 11 (2003) 557-567.
76. B.S. Kroezen, G. Conti, B. Girardi, J. Cramer, X. Jiang, S. Rabbani, J. Müller, M. Kokot, E. Luisoni, D. Ricklin, O. Schwardt, B. Ernst. "A Potent Mimetic of the Siglec-8 Ligand 6'-Sulfo-Sialyl Lewisx." *ChemMedChem,* 15 (2020) 1706-1719.
77. C.D. Rillahan, E. Schwartz, R. McBride, V.V. Fokin, J.C. Paulson. "Click and pick: identification of sialoside analogues for siglec-based cell targeting." *Angewandte Chemie International Edition in English*, 51 (2012) 11014-11018.
78. C.M. Nycholat, S. Duan, E. Knuplez, C. Worth, M. Elich, A. Yao, J. O'Sullivan, R. McBride, Y. Wei, S.M. Fernandes, Z. Zhu, R.L. Schnaar, B.S. Bochner, J.C. Paulson. "A Sulfonamide Sialoside Analogue for Targeting Siglec-8 and -F on Immune Cells." *Journal of the American Chemical Society*, 141 (2019) 14032-14037.
79. S. Duan, B. M. Arlian, C. M. Nycholat, Y. Wei, H. Tateno, S. A. Smith, M. S. Macauley, Z. Zhu, B. S. Bochner, J. C. Paulson. "Nanoparticles Displaying Allergen and Siglec-8 Ligands Suppress IgE-Fc ε RI - Mediated Anaphylaxis and Desensitize Mast Cells to Subsequent Antigen Challenge." *The Journal of Immunology*, 206 (2021) 116-126.
80. R.D. Cummings, F.-T. Liu, G.A. Rabinovich, S.R. Stowell, G.R. Vasta. "Galectins." In *Essentials of Glycobiology*, Cold Spring Harbor: New York, (2022) 869-889.

81. C.P. Modenutti, J.I.B. Capurro, S. Di Lella, M.A. Martí. "The Structural Biology of Galectin-Ligand Recognition: Current Advances in Modeling Tools, Protein Engineering, and Inhibitor Design." *Frontiers in Chemistry*, 7 (2019) 823.
82. S. Bertuzzi, J.I. Quintana, A. Ardá, A. Gimeno, J. Jiménez-Barbero. "Targeting Galectins with Glycomimetics." *Frontiers in Chemistry*, 8 (2020) 593.
83. S. Sato, C. St-Pierre, P. Bhaumik, J. Nieminen. "Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs)." *Immunological reviews*, 230 (2009) 172-87.
84. M. Prado Acosta, B. Lepenies. "Bacterial glycans and their interactions with lectins in the innate immune system." *Biochemical Society Transactions*, 47 (2019) 1569-1579.
85. G.R. Vasta. "Roles of galectins in infection." *Nature Reviews Microbiology*, 7 (2009) 424-438.
86. L. Johannes, R. Jacob, H. Leffler. "Galectins at a glance." *Journal of Cell Science*, 131 (2018) jcs208884.
87. G.A. Rabinovich, Y. van Kooyk, B.A. Cobb. "Glycobiology of immune responses." *Annals of the New York Academy of Sciences*, 1253 (2012) 1-15.
88. F.-T. Liu, G.A. Rabinovich. "Galectins: regulators of acute and chronic inflammation." *Annals of the New York Academy of Sciences*, 1183 (2010) 158-182.
89. B. Rossi, M. Espeli, C. Schiff, L. Gauthier. "Clustering of pre-B cell integrins induces galectin-1-dependent pre-B cell receptor relocalization and activation." *The Journal of Immunology*, 177 (2006) 796-803.

90. R.P.M. Dings, M.C. Miller, R.J. Griffin, K.H. Mayo. “Galectins as Molecular Targets for Therapeutic Intervention.” *International Journal of Molecular Sciences*, 19 (2018) 905.
91. M. Hara, M. Niwa, K. Noguchi, T. Kanayama, A. Niwa, M. Matsuo, Y. Hatano, H. Tomita. “Galectin-3 as a Next-Generation Biomarker for Detecting Early Stage of Various Diseases.” *Biomolecules*, 10 (2020) 389.
92. C. Shimada, R. Xu, L. Al-Alem, M. Stasenko, D.R. Spriggs, B.R. Rueda. “Galectins and Ovarian Cancer.” *Cancers*, 12 (2020) 1421.
93. M. A. Campanero-Rhodes, I. Kalograiaki, B. Euba, E. Llobet, A. Ardá, J. Jiménez-Barbero, J. Garmendia, D. Solís. “Exploration of Galectin Ligands Displayed on Gram-Negative Respiratory Bacterial Pathogens with Different Cell Surface Architectures.” *Biomolecules*, 11 (2021) 595.
94. P.G. Traber, H. Chou, E. Zomer, F. Hong, A. Klyosov, M.-I. Fiel, S.L. Friedman. “Regression of Fibrosis and Reversal of Cirrhosis in Rats by Galectin Inhibitors in Thioacetamide-Induced Liver Disease.” *PLOS ONE*, 8 (2013) e75361.
95. P.G. Traber, E. Zomer. “Therapy of Experimental NASH and Fibrosis with Galectin Inhibitors.” *PLOS ONE*, 8 (2013) e83481.
96. N. Clementy, E. Piver, A. Bisson, C. Andre, A. Bernard, B. Pierre, L. Fauchier, D. Babuty. “Galectin-3 in Atrial Fibrillation: Mechanisms and Therapeutic Implications.” *International Journal of Molecular Sciences*, 19 (2018) 976.
97. A. Krześlak, A. Lipińska. “Galectin-3 as a multifunctional protein.” *Cellular & Molecular Biology Letters*, 9 (2004) 305-28.
98. M.P. Lenza, C. Di Carluccio, F. Nieto-Fabregat, L. Pirone, R. Russo, S. Di Gaetano, D. Capasso, A. Stornaiuolo, A. Iadonisi, M. Saviano, R. Marchetti, E. Pedone, A. Silipo. “Structural Insights Into Galectin-3 Recognition of a Selenoglycomimetic.” *ChemMedChem*, 20 (2025) e202401005.

99. N.C. Henderson, T. Sethi. “The regulation of inflammation by galectin-3.” *Immunological Reviews,* 230 (2009) 160-171.
100. O.B. Garner, L.G. Baum. “Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling.” *Biochemical Society Transactions,* 36 (2008) 1472-7.
101. L. Díaz-Alvarez, E. Ortega. “The Many Roles of Galectin-3, a Multifaceted Molecule, in Innate Immune Responses against Pathogens.” *Mediators of Inflammation,* 2017 (2017) 9247574.
102. J. Ochieng, V. Furtak, P. Lukyanov. “Extracellular functions of galectin-3.” *Glycoconjugate Journal,* 19 (2002) 527-35.
103. Y. Takenaka, T. Fukumori, A. Raz. “Galectin-3 and metastasis.” *Glycoconjugate Journal,* 19 (2002) 543-9.
104. A. Fortuna-Costa, A.M. Gomes, E.O. Kozlowski, M.P. Stelling, M.S.G. Pavão. “Extracellular Galectin-3 in Tumor Progression and Metastasis.” *Frontiers in Oncology,* 4 (2014) 138.
105. R. Tellez-Sanz, L. Garcia-Fuentes, A. Vargas-Berenguel. “Human galectin-3 selective and high affinity inhibitors. Present state and future perspectives.” *Current Medicinal Chemistry,* 20 (2013) 2979-2990.
106. T.J. Hsieh, H.Y. Lin, Z. Tu, T.C. Lin, S.C. Wu, Y.Y. Tseng, F.T. Liu, S.T.D. Hsu, C.H. Lin. “Dual thio-digalactoside-binding modes of human galectins as the structural basis for the design of potent and selective inhibitors.” *Sci. Rep.*, 6 (2016) 1-9.
107. Y.C. Chan, H.Y. Lin, Z. Tu, Y.H. Kuo, S.D. Hsu, C.H. Lin. "Dissecting the Structure-Activity Relationship of Galectin-Ligand Interactions." *Int. J. Mol. Sci.*, 19 (2018) 392.
108. M. Atanasova, H. Bagdonas, J. Agirre. “Structural glycobiology in the age of electron cryo-microscopy.” *Curr. Opin. Struct. Biol.*, 62 (2020) 70-78.

109. P.M. Nieto. “The use of NMR to study transient carbohydrate-protein interactions.” *Front. Mol. Biosci.,* 5 (2018) 1-7.
110. A. Gimeno, P. Valverde, A. Ardá, J. Jiménez-Barbero. "Glycan structures and their interactions with proteins. A NMR view." *Curr. Opin. Struct. Biol.*, 62 (2020) 22-30.
111. A. Ardá, J. Jiménez-Barbero. "The recognition of glycans by protein receptors. Insights from NMR spectroscopy." *Chem. Commun.*, 54 (2018) 4761-4769.
112. W. Becker, K.C. Bhattiprolu, N. Gubensäk, K. Zangger. "Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy." *Chemphyschem.*, 19 (2018) 895-906.
113. A.D. Gossert, W. Jahnke. "NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules." *Prog Nucl Magn Reson Spectrosc.*, 97 (2016) 82-125.
114. R. Marchetti, S. Perez, A. Arda, A. Imberty, J. Jimenez-Barbero, A. Silipo, A. Molinaro. "“Rules of Engagement” of Protein-Glycoconjugate Interactions: A Molecular View Achievable by using NMR Spectroscopy and Molecular Modeling." *ChemistryOpen*, 5 (2016) 274-296.
115. C. Di Carluccio, R.E. Forgione, S. Martini, F. Berti, A. Molinaro, R. Marchetti, A. Silipo. "Investigation of protein-ligand complexes by ligand-based NMR methods." *Carbohydr Res.*, 503 (2021) 108313.
116. M. Mayer, B. Meyer. "Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor." *J. Am. Chem. Soc.*, 123 (2001) 6108-611.
117. P.M. Angulo, P.M. Enríquez-Navas, P.M. Nieto. "Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates Chemistry." *Chemistry – A European Journal*, 26 (2010) 7803-7812.

118. A. Ardá, A. Canales, F.J. Cañada, J. Jiménez-Barbero. “Chapter 1: Carbohydrate–Protein Interactions: A 3D View by NMR, in Carbohydrates in Drug Design and Discovery.” (2015) 1-20.
119. M.P. Williamson. "Using chemical shift perturbation to characterise ligand binding." *Prog Nucl Magn Reson Spectrosc.*, 73 (2013) 1-16.
120. W. Braun, C. Bösch, L.R. Brown, N. Go, K. Wüthrich. "Combined use of proton-proton overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon." *BBA - Protein Struct.*, 667 (1981) 377-396.
121. L.P. McIntosh, F.W. Dahlquist. "Biosynthetic Incorporation of 15N and 13C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins." *Q. Rev. Biophys.*, 23 (1990) 1-38.
122. F. Nieto-Fabregat, M.P. Lenza, A. Marseglia, C. Di Carluccio, A. Molinaro, A. Silipo, R. Marchetti. “Computational toolbox for the analysis of protein-glycan interactions.” *Beilstein J. Org. Chem.*, 20 (2024) 2084-2107.
123. R. Maldonado-Hernández, O. Quesada, J.A. González-Feliciano, A. Baerga-Ortiz, J.A. Lasalde-Dominicci. “Identification of the native Torpedo californica nicotinic acetylcholine receptor's glycan composition after a multi-step sequential purification method using MALDI-ToF MS.” *Proteomics*, 24 (2024) e2300151.
124. B. Pandey, S. S, A. Chatterjee, V. Mangala Prasad. “Role of surface glycans in enveloped RNA virus infections: A structural perspective.” *Proteins*, 93 (2025) 93-104.
125. Y. Liu, M. Grimm, W.T. Dai, M.C. Hou, Z.X. Xiao, Y. Cao. "CB-Dock: a web server for cavity detection-guided protein-ligand blind docking." *Acta Pharmacol. Sin.*, 41 (2020) 138-144.

126. O. Trott, A.J. Olson. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." *J. Comput. Chem.*, 31 (2010) 455-461.

SECTION II – EXPERIMENTAL SECTION

Chapter 3: Materials and methods

3.1 Recombinant plasmids generation for proteins expression

The protein expression process for this study involved plasmids initially synthesized by Eurofins. Following this, the gene sequences were subcloned into a range of vectors. The CRD of Galectin-3 and Siglec-8 were expressed and purified in our laboratory.

For expression in *E. coli*, vectors from the pET family were preferred (details in **Table 3.1**). The sequence of human Siglec-8 CRD (V domain) (Uniprot: Q9NYZ4, aa 17-155, C42S) was inserted into pET-43.1(a) plasmid, adhering to the construct previously described by Propster et al. [1]. To facilitate subsequent protein purification, the C-terminal domain was engineered with a thrombin cleavage site and a His_6-tag. On the other hand, the Galectin-3 CRD (Uniprot: P17931, aa 112-250) was cloned into the pETM11 expression vector with a N-terminal polyHis-tag that is removable by TEV protease.

Table 3.1. pET plasmids employed in this PhD thesis.

Plasmid	Description	Size (bp)	Resistance
pET-43.1(a)	T7 lac promoter/ N and C terminal His-Tag/ Thrombin recognition site	7272	Ampicillin
pETM11	T7 lac promoter/ N terminal His-Tag/ TEV protease cleavage site	6092	Kanamycin

Adapted from [2].

3.1.1 Transformation into prokaryotic cells

Siglec-8 and Galectin-3 were successfully expressed within *E. coli* cells. A major advantage of employing this bacterial expression system is its capacity to yield substantial quantities of ^{15}N-labeled proteins, which are essential for NMR experiments. For the transformation, plasmids were introduced into competent *E. coli* cells (**Table 3.2**), adhering to the specific protocols provided by their respective manufacturers.

Table 3.2. Different *E. coli* strains employed in this PhD thesis.

Strain	Genotype	Application	Plasmid transfected	Ref
TOP10	F- mcrA (mrr-hsdRMS-mcrBC) 80lacZ M15 lacX74 recA1 ara 139 (ara-leu)7697 galUgalKrpsL (StrR) endA1 nupG	DNA plasmidic amplification	all	[3]
BL21 (DE3) GOLD	*E. coli* F - ompT hsdSB (rB-, mB-) gal dcm (DE3)	Protein expression	Galectin-3	[4]
Rosetta-gami B (DE3)	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F′[lac+ lacIq pro] gor522::Tn10 trxB pRARE2 (CamR, StrR, TetR)	Protein expression	Siglec-8	[4]

Initially, 100 ng of plasmid DNA was incubated with the cells for 30 min. The cell membranes were then made permeable using a heat shock at 42 °C for 45 sec. Following this brief thermal treatment, 300 µL of Luria-Bertani (LB) broth was added, and the cells were grown for 1 hour and 30 minutes. To select successful

transformants, 150 µL of the cell suspension was then plated onto an LB agar plate containing the appropriate antibiotics, followed by overnight incubation at 37 °C. The presence of antibiotics promoted the growth of only those cells that had successfully incorporated the plasmid.

3.1.2 Isolation of the expression plasmid by Mini-preparation

The isolation of the expression plasmid or DNA extraction was performed by Miniprep Kit (Qiagen) using the protocol suggested by provider.

Briefly, a single colony of recombinant bacteria was inoculated into 10 mL of LB medium supplemented with the appropriate antibiotic. This culture was incubated overnight at 37°C with constant agitation. The following day, the bacterial culture was centrifuged, and the resulting pellet was resuspended in Solution 1 (cell resuspension buffer). Cell lysis was then performed using Solution 2 (cell lysis buffer), followed by neutralization with Solution 3 (neutralization buffer). After a subsequent centrifugation step, the supernatant, containing the plasmid DNA, was loaded onto a silica column to capture the DNA. Following several washing steps to remove impurities, the DNA was eluted from the column using 20 µL of ultrapure water (H_2O mq). The concentration and purity of the isolated plasmid DNA were then assessed by spectrophotometric analysis using a NanoDrop instrument.

3.1.3 Enzymatic digestion

Restriction enzyme digestion is a fundamental molecular biology technique employed to precisely cut DNA sequences at specific recognition sites, which are unique to each enzyme. This method allows for the effective isolation or

preparation of desired DNA fragments (inserts) and vectors for subsequent cloning steps. For this study, the restriction enzymes listed in **Table 3.3** were utilized.

Table 3.3. Restriction enzymes used in this PhD thesis.

Enzyme	Restriction site
NdeI	5'..CA↓TATG..3' 3'..GTAT↑AC..5'
XhoI	5'..C↓TCGAG..3' 3'..GAGCT↑C..5'

Each digestion reaction was prepared using four key components:

1. DNA: the amount of plasmid DNA (vector or insert) was determined based on the initial concentration of each sample.
2. Restriction enzymes: they were added, with one unit defined as the amount required to completely digest 1 µg of DNA.
3. Buffer: 10X Tango Buffer, specifically optimized for the enzymes used, was diluted to a final concentration of 2X in the reaction.
4. Nuclease-free H_2O: this component was added as needed to bring the reaction to the desired final volume.

The digestion was carried out at a total volume of 20 µL. Both enzymes used operate optimally at an incubation temperature of 37 °C. Following the setup, all digestion reactions were incubated at 37 °C overnight to ensure complete cleavage and subsequently inactivated by heating the reactions to 80 °C.

3.1.4 DNA bands: purification and ligation

After electrophoresis, the DNA bands of interest were excised from the agarose gel. The DNA within these bands was then purified using a DNA purification kit (Qiagen) based on centrifugation. This process began with a "binding buffer" that completely dissolved the agarose gel. The dissolved solution was then washed in several steps with a membrane "wash buffer". Finally, the purified DNA was eluted using 20 µL of ultrapure water (H_2O mQ). The concentration of the isolated nucleic acid was determined using spectrophotometric analysis with a NanoDrop instrument.

To construct a complete vector containing the gene of interest, the ligation technique was employed. Specifically, T4 DNA ligase was used, which catalyzes the joining of two DNA strands by forming phosphodiester bonds between the 5'-phosphate and 3'-hydroxyl groups of adjacent nucleotides. To perform the ligation between the insert and the vector, the following formula (**Equation 3.1**) was used:

$$\text{ng insert} = \frac{\text{ng vector} \,.\, K_b \text{ insert}}{K_b \text{ vector}} \,.\, x$$

For each ligation run, a ratio of 3:1 (insert: vector) was utilized. The final ligation reaction mixture consisted of the vector, insert, T4 DNA ligase, a specific 10X ligase buffer, and nuclease-free H_2O to bring the reaction to its final volume. The entire ligation reaction was performed at a volume of 20 µL. Following assembly, the reaction was incubated overnight at 16 °C without agitation.

3.2 Proteins expression

3.2.1 Unlabeled proteins expression

For this study, a single bacterial colony, isolated from an agar plate, was first cultivated overnight in 100 mL of LB medium containing the appropriate antibiotics, under gentle shaking conditions. The following day, a 20 mL aliquot of this starter culture was used to inoculate larger flasks, each holding 1 L of fresh LB medium with antibiotics. These larger cultures were then grown at 37 °C with constant agitation (180 revolutions per minute, rev min^{-1}) for approximately 3/4 hours.

Once the optical density at 600 nm (OD_{600}) reached between 0.6 and 0.8, protein expression was induced. Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to the cultures at concentrations optimized for each protein: 1 mM for Siglec-8 expression and 0.5 mM for Galectin-3 expression. The induction period for Siglec-8 was extended for 24 hours at 20 °C, maintaining 180 rev min^{-1} agitation. In contrast, Galectin-3 induction was performed for 16 hours at 22 °C.

Following the expression phase, the bacterial cultures were harvested via centrifugation. The pellet was collected, frozen and stored at -20°C.

3.2.2 Labeled ^{15}N proteins expression

To express the uniformly ^{15}N-labeled CRD lectins, transformed cells were initially cultured in 100 mL of a minimal medium. This medium consisted of M9 supplemented with $^{15}NH_4Cl$ as the only nitrogen source, along with Biotin, Thiamine, and other essential elements detailed in **Tables 3.4** and **3.5**. These cultures were incubated overnight at 37 °C with gentle shaking at 180 rpm. Following this initial growth, a 10 mL aliquot was transferred to 1 L of the

complete minimal medium. The subsequent growth conditions replicated those outlined in the preceding section (3.2.1).

Table 3.4. Complete composition of M9 media.

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

Table 3.5. Salts composition.

Stock of salts (1L)	
60g	Na_2HPO_4
30g	KH_2PO_4
5g	NaCl

3.3 Proteins purification

3.3.1 CRD Siglec-8 purification

The bacterial cell pellet was resuspended in a lysis buffer containing 20 mM Tris pH 8, 300 mM NaCl, and a complete protease inhibitor cocktail (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roche). Bacterial cells were sonicated on ice for 12 cycles (10 seconds on, 59 seconds off, at 60% amplitude) to extract the expressed protein. Insoluble material was removed by centrifugation for 40' at 4 °C at 15000 rpm.

The resulting supernatant, containing the soluble proteins, was then filtered through a 0.45 µm pore-size syringe filter directly before loading on a prepacked Ni^{2+}-NTA agarose column (HisTrap FF, 5 ml) (GE Healthcare). This column has been pre-equilibrated with 5 column volumes (CV) of 20 mM Tris pH 8 and 300 mM NaCl (Buffer A) connected to an AKTA system (GE Healthcare). After washing with 20-30 CV of Buffer A, the target protein was then eluted with a linear gradient (0, 8, 16, and 100%) of 20 mM Tris pH 8, 300 mM NaCl, and 500 mM Imidazole (Buffer B). The eluted fraction was collected and concentrated to a minimum volume of 1 mL using an Amicon filter with a 10 kDa cut-off. Subsequently, a gel filtration chromatography step was performed using a Superdex 75 10/300 column (GE Healthcare). The protein was eluted into the final buffer containing 20 mM Sodium Phosphate, and 150 mM NaCl, pH 7.4.

SDS-PAGE analysis confirmed that the Siglec-8 V domain eluted as a monomer with a molecular weight of 16 kDa from the size exclusion step. Finally, the purified protein was concentrated further using an Amicon filter (10 kDa cut-off) to a final concentration of 2 mg/mL (100 µM) for NMR analysis. Any remaining protein was, in some instances, flash-frozen with liquid nitrogen and stored at -20 °C.

3.3.2 CRD Galectin-3 purification

The protocol by Pirone et al. [5] was employed for the purification of CRD Galectin-3. In particular, the harvested cell pellet was resuspended in 30 mL of lysis buffer containing 20 mM Sodium Phosphate, 500 mM NaCl, 10 mM Imidazole pH 7.0. The cells were then subjected to sonication, followed by centrifugation for 30 minutes.

The resulting supernatant was subsequently subjected to affinity chromatography onto an HisTrapTM column (5 mL, Ge Healthcare) previously equilibrated in lysis buffer. The desired protein was then eluted with 300 mM Imidazole. Finally, fractions containing the recombinant protein were loaded onto a Superdex 75 10/300 column (GE Healthcare), which was equilibrated in 20 mM Sodium Phosphate, 150 mM NaCl pH 7.0 and connected to an AKTA system (GE Healthcare) [6].

3.4 Binding studies by NMR analysis (related to Chapter 4 and 5)

All NMR spectra for study of ligand-protein interactions were acquired on a Bruker 600-MHz AVANCE NEO instrument equipped with a cryogenically cooled probe at 298 K.

3.4.1 ^{1}H,^{15}N-HSQC based titration experiments for analyzing the binding of lectins to glycan ligands

To analyze the interactions between Siglec-8 V domain and Galectin-3 and the ligands, from the protein point of view, ^{1}H,^{15}N-TROSY experiments were performed, using systematically increasing ligand concentrations. Ligands for Siglec-8 were added stepwise to reach protein:ligand molar ratios of 1:30 and 1:8.

For Galectin-3 titrations, one ligand was titrated up to a 1:20 ratio, while another was titrated up to 1:5. This approach allowed evaluating the most perturbed amino acids on proteins upon ligand binding. The Siglec-8 ligands evaluated in this thesis were kindly provided by Prof. Paul Murphy, a collaborator on this project. The ligands evaluated for Galectin-3 were supplied by Prof. Alfonso Iadonisi.

Uniformly ^{15}N-labeled Siglec-8 and Galectin-3 were dissolved in 500 µl of phosphate buffer (20 mM sodium phosphate, 40 mM NaCl, pH 7.4) at a concentration of 200 µM and 100 µM, respectively, using 90% H_2O and 10% D_2O. For this purpose, standard ^{1}H,^{15}N-HSQC experiments were performed with 256 (t1) and 2048 (t2) complex data points in ^{15}N and ^{1}H dimensions, respectively. All the experiments were recorded at a temperature of 293 K to avoid protein precipitation. NMR data were analyzed with CcpNmr Analysis V3.2.

3.4.1.1 Protein-based NMR: CSP analysis of proteins titration

CcpNmr Analysis software V3.2 was used for data processing [7]. Average chemical shift changes (δ_{average}) were calculated by using the following equation (**Equation 3.2**) [8]:

$$\delta_{\text{average}} = \sqrt{\frac{1}{2}\left[\Delta\delta^2{}_H + (0.14 \cdot \Delta\delta^2{}_N)\right]}$$

The amino acid backbone of each protein was assigned using data deposited into the Biological Magnetic Resonance Data Bank (BMRB) (bmrb 4909 for Galectin-3 and bmrb 25798 for Siglec-8).

3.4.1.2 Dissociation constant calculation by NMR titration

The equilibrium dissociation constant (K_d) was determined using NMR titration. Changes in the chemical shift ($\Delta\delta$) were monitored as the ligand concentration increased. K_d was then calculated by numerical fitting to the following equation (**Equation 2.7**):

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

In this equation, $\Delta\delta_{obs}$ represents the observed chemical shift difference from the free state, while $\Delta\delta_{max}$ is the maximum shift difference observed when the ligand concentration saturates the protein. Both $\Delta\delta_{max}$ and K_d were obtained through fitting. $[P]_t$ and $[L]_t$ refer to the total concentrations of protein and ligand, respectively.

3.4.2 Ligand-based NMR: STD NMR experiments

For the STD experiments all the samples were dissolved in deuterate buffer (PBS pH 7.4). All the STD NMR experiments were performed at 298 K and were acquired with 32 k data points and zero-filled up to 64 k data points prior to processing. Protein resonances were selectively irradiates using 40 Gauss pulses with a length of 50 ms, with the off-resonance pulse frequency set at 40 ppm and the on-resonance pulse at 0 ppm, with a relaxation delay D1 of 2 s.

A standard 1:50 protein: ligand molar ratio was used, with a protein concentration at 50 µM. The 1D STD sequence from Bruker library with spoil and T2 filter (stddiff.3) was employed for the STD NMR experiments. Selective saturation of the protein resonances (on-resonance spectrum) was performed by irradiating at 0.75 ppm, for a total saturation time of 2 seconds. For the reference spectrum (off-

resonance), the irradiation frequency was set at 100 ppm. The STD NMR spectra were obtained by subtracting the on-resonance from the off-resonance spectrum. Additional blank spectra were recorded using the same conditions with only protein or only ligand. The absolute STD value was calculated from the ratio between the intensity of a given signal in the off resonance versus the STD spectra. The 100 % value was assigned to the proton with the higher absolute STD value.

For both proteins, the % STD displayed in the ligands' epitope maps were obtained by the ratio of the STD signals in the STD spectrum (I_0 - I_{sat}) and each relative peak intensity of the unsaturated reference spectrum (off-resonance, I_0), at saturation time of 2 s. The highest STD signal was set to 100% and the other signals were normalized to this peak. STD NMR experiments at the same conditions were also performed on a solution of ligand in absence of protein as control.

3.4.2.1 CPMG NMR

The cpmgz pulse program based on spin-echo pulse sequence was used for the acquisition of CPMG NMR experiments. These analyses were carried out to determine T_2 spin-spin relaxations of the ligand in absence and in presence of the protein. The signal relaxation was measured for 100 ms and the experiments were acquired with 256 scans. A pseudo 2D sequence with water suppression using excitation sculpting with gradients was considered. For the calculation of T_2 relaxation time, CPMG experiments at different mixing times were acquired. T_2 signals were analyzed with Dynamic Center 2.7.1 that fitted the data by the following equation (**Equation 3.3**):

$$f(t) = Io * \exp(-t/T).$$

3.5 Biophysical techniques: Isothermal Titration Calorimetry (ITC)

ITC experiments were performed by Prof. Beat Ernst's laboratory (Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland), to study the interactions between high-affinity analogues of sialoglycans and Siglec-8, as had been previously done with other glycomimetics by Conti et al. [9].

ITC data were carried out by a MicroCal PEAQ-ITC Automated instrument (Malvern Panalytical) at 298 K and 10 μcal/s reference power. The protein was dialyzed against assay buffer (100 mM Hepes, and 150 mM Salt, pH 7.4) overnight using Slide-A-Lyzer Dialysis Cassette G2 3 kDa cut-off. The ligands were dissolved in the same protein buffer.

Titrations were performed by titrating 5 mM of the compounds into 50 μM of Siglec-8. In addition, it was used assay buffer as control. Injections were performed every 180 s, for a total of 18 injections (2 μL for each injection), with a 750 r.p.m. stirring speed.

3.6 Computational approaches

3.6.1 Molecular Mechanics (related to Chapter 5)

Molecular mechanics (MM) studies were performed using Maestro software. Adiabatic maps were created for the Gal1-S-1Gal disaccharide, focusing on its glycosidic linkages. This linkage is precisely described by the torsion angles Φ (H1-C1-O-CX') and Ψ (C1-O-CX'-HX'). For the purposes of this calculation, the S atom was used instead of the Se atom.

A MM3* *force field* at a dielectric constant of 80 was carried out for all calculations. Both Φ and Ψ dihedral angles were varied incrementally using a grid step of 18 degrees. For each (Φ, Ψ) coordinate on the map, an energy minimization

was performed utilizing 2000 iterations of the Polak-Ribière (P.R.) algorithm. To determine the Restrained ElectroStatic Potential (RESP) charges, the β-Gal residue featuring a selenium atom at position 1 was modeled in Gaussian 09 [10]. This modeling included a Hartree-Fock (HF) calculation with a 6-31G* basis set. VFFDT [11].

3.6.2 Molecular Docking (related to Chapter 4)

Docking calculations were conducted with CB-Dock2 [12]. Siglec-8 protein structure from *PDB 7QUI* file was prepared using Maestro's Prime protein preparation wizard [13]. Initial models of the Siglec-8/ligand **11** complex (Chapter 4) were generated through a manual docking approach, guided by experimental NMR data. The *PDB 7QUI* crystal structure of the Siglec-8 V-domain bound to compound **2** (6′-O-sulfo NSAα2-3SLac*NA*c) served as the reference pose, onto which the ligand was then superimposed. Following this, an initial docking run was performed with Glide (part of Maestro) [14], which yielded a specific bound pose for ligand **11**. Furthermore, the automated online platform CB-Dock2 was also employed for additional docking calculations [12] and visualizations were created with PyMOL.

3.6.3 Molecular Dynamics (related to Chapter 5)

Molecular dynamics (MD) simulations for analyzing the complexes between Galectin-3 and the ligands, were performed with AMBER 18 software [15] in explicit water. The protein structure was pre-processed first within Maestro Protein Preparation Wizard in Maestro program [16]. The ligands were built by using carbohydrate builder utility of the glycam website [17].

Molecular dynamics calculations were employed with the PMEMD module. Specifically, the force fields to assign atom types and charges of the protein and the saccharide moiety of the ligands were AMBER ff14SB and GLYCAM06j-1, respectively. GAFF2 was used for the aromatic moiety of the ligand SeDG-Bn. For the seleno-glycosidic linkage, a modified glycam force field was developed to incorporate parameters for the Se atom. Charges of molecules were neutralized by adding counterions (Na^+) using the Leap module. The structures were also hydrated considering octahedral boxes with explicit TIP3P water molecules, buffered at 15 Å. Input files were generated using the tleap modules of the AMBER 18 package. The system coordinates were saved for the 100 ns simulations by the PMEMD module in AMBER. To obtain 10,000 dynamic structures for each complex, coordinates were ranked. Trajectories were analyzed using the ptraj module within AMBER18 and visualized by VMD program [18].

The analysis of the clusters with respect to the ligand RMSD was calculated using K-mean algorithm implemented in ptraj module [19]. To illustrate complex interactions, the structure representative of the most frequently occurring cluster was employed. The 3D images were performed with PyMOL [20]. A custom script was utilized to analyze dihedral conformations, allowing us to visualize torsion variations during the MD simulation and generate a histogram of the most frequently observed values.

Section II - References

1. J.M. Propster, F. Yang, S. Rabbani, B. Ernst, F.H.-T. Allain, M. Schubert. "Structural Basis for Sulfation-Dependent Self-Glycan Recognition by the Human Immune-Inhibitory Receptor Siglec-8." *Proc. Natl. Acad. Sci. U. S. A.*, 113 (2016) E4170-E4179.

2. M.I. Singh, V. Jain. "Tagging the expressed protein with 6 histidines: rapid cloning of an amplicon with three options." *PLoS One*, 8 (2013).

3. J. Liu, W. Chang, L. Pan, X. Liu, L. Su, W. Zhang, Q. Li, Y. Zheng. "An Improved Method of Preparing High Efficiency Transformation Escherichia coli with Both Plasmids and Larger DNA Fragments." *Indian J. Microbiol.*, 58 (2018).

4. M. Fathi-Roudsari, A. Akhavian-Tehrani, N. Maghsoudi. "Comparison of Three Escherichia coli Strains in Recombinant Production of Reteplase." *Avicenna J. Med. Biotechnol.*, 8 (2016).

5. L. Pirone, F. Nieto-Fabregat, S. Di Gaetano, D. Capasso, R. Russo, S. Traboni, A. Molinaro, A. Iadonisi, M. Saviano, R. Marchetti, A. Silipo, E. Pedone. "Exploring the Molecular Interactions of Symmetrical and Unsymmetrical Selenoglycosides with Human Galectin-1 and Galectin-3." *Int. J. Mol. Sci.*, 23 (2022) 8273.

6. S. Di Gaetano, L. Pirone, I. Galdadas, S. Traboni, A. Iadonisi, E. Pedone, M. Saviano, F.L. Gervasio, D. Capasso. "Design, Synthesis, and Anticancer Activity of a Selenium-Containing Galectin-3 and Galectin-9N Inhibitor." *Int J Mol Sci.*, 23 (2022) 13589.

7. S.P. Skinner, R.H. Fogh, W. Boucher, T.J. Ragan, L.G. Mureddu, G.W. Vuister. "CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis." *J. Biomol. NMR*, 66 (2016) 111-124.

8. M.P. Williamson. "Using chemical shift perturbation to characterise ligand binding." *Prog. Nucl. Magn. Reson. Spectrosc.*, 73 (2013) 1-16.

9. G. Conti, A. Bärenwaldt, S. Rabbani, T. Mühlethaler, M. Sarcevic, X. Jiang, O. Schwardt, D. Ricklin, R.J. Pieters, H. Läubli, B. Ernst. "Tetra- and Hexavalent Siglec-8 Ligands Modulate Immune Cell Activation." *Angew. Chem. Int. Ed. Engl.*, 62 (2023) e202314280.

10. J. Hiscocks, M.J. Frisch. "Gaussian 09: IOps Reference." In: M. Caricato, M.J. Frisch (Eds.). Wallingford, CT, USA: Gaussian, (2009).

11. B. Kurt, H. Temel. "Parameterization of Boronates Using VFFDT and Paramfit for Molecular Dynamics Simulation." *Molecules*, 25 (2020) 2196.

12. Y. Liu, X. Yang, J. Gan, S. Chen, Z.X. Xiao, Y. Cao. "CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting." *Nucleic Acids Res.*, 50 (2022) W159-W164.

13. G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman. "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *J. Comput. Aided Mol. Des.*, 27 (2013) 221-34.

14. R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin. "Glide: a new approach for rapid, accurate docking

and scoring. 1. Method and assessment of docking accuracy." *J. Med. Chem.*, 47 (2004) 1739-49.

15. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, 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, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, A. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman. *AMBER 2018*, (2018).
16. Schrödinger Release 2022-3, Maestro. New York: Schrödinger, LLC, (2021).
17. www.glycam.com.
18. W. Humphrey, A. Dalke, K. Schulten. "VMD: visual molecular dynamics." *J. Mol. Graph.*, 14 (1996) 33-8.
19. D.R. Roe, T.E. Cheatham. "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data." *J. Chem. Theory Comput.*, 9 (2013) 3084-3095.
20. Schrödinger L.L.C. *The PyMOL Molecular Graphics System, Version 1.8*, (2015).

SECTION III – RESULTS AND DISCUSSIONS

Chapter 4: Novel molecular insight into the interaction of Siglec-8 with sialyl triazoles as glycomimetics

4.1 Introduction

As described in the general introduction, Siglec-8 belonging to a family of CD-33-related receptors is a receptor mainly found on the surface of immune cells, as eosinophils, mast cells and basophils [1-4]. This protein plays a key role in controlling how immune system responds to allergens and inflammation. If Siglec-8 doesn't work correctly, it can lead to problems like apoptosis and ongoing inflammation [2]. Given its role in these processes, Siglec-8 is considered a promising therapeutic target for asthma, chronic sinusitis, persistent urticaria, and hematological disorders [5-7].

It is composed of an extracellular domain (ECD), which features one V-type and two C-type Ig-like domains. This is followed by a single transmembrane domain and an intracellular domain containing an ITIM motif (or ITIM domain) [2]. Ligand binding can induce phosphorylation of this motif by Src family tyrosine kinases, establishing high-affinity binding sites for SHP-1 and SHP-2 (SH2 domain-containing protein tyrosine phosphatases) [8-9]. SHP-1 and SHP-2 can inhibit signaling pathways through the inactivation of key kinases, thereby modulating different immune responses [10-11]. In addition to the ITIM, Siglec-8 possesses a characteristic N-terminal lectin domain on its ECD, enabling it to bind glycans [1, 12].

The 3D structure of V-domain (or lectin domain), determined by NMR [7], consists of a single β-sandwich. This architecture includes two antiparallel β-sheets (specifically, β-strand ABED and C'CFG), which are linked by an intra-sheet disulfide bond. Among different Siglecs, a conserved Arginine residue on β-

strand F is known to form a critical salt bridge, which is crucial for the recognition of sialic acid [7].

Results from glycan array screening showed that 6'-sulfo sialyl Lewis^x (6′S sLeX: Neu5Acα2-3[6S] Gal β1-4[Fucα1–3] [6S]GlcNAc) (**Figure 4.1**, compound **1**) is a preferred ligand for Siglec-8 lectin domain, with a dissociation constant (K_d) of 300 μM [7, 13-14]. The CC' and GG' loops within Siglec-8 were demonstrated to be essential for the recognition of **1** [7] (**Figure 4.2**).

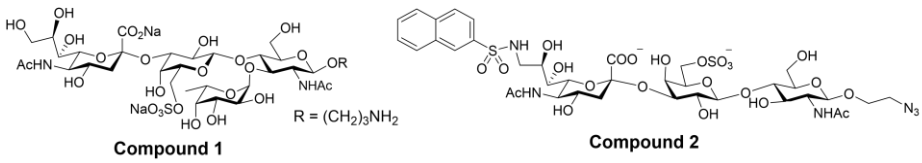


Figure 4.1. The previously identified Siglec-8 inhibitors. The basic 6'-Sulfo-sialyl Lewis^x ligand (**1**) and its derivative: 6′-O-sulfo NSA-α2-3SLacNAc (**2**).

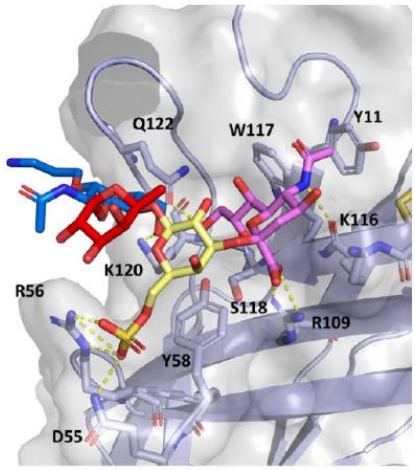
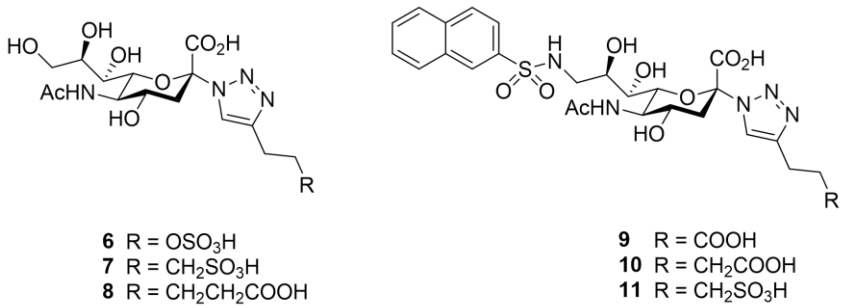


Figure 4.2. The illustration of the interaction between Siglec-8 V-domain and 6'-sulfo sialyl Lewis^x (6′S sLe^x) as described by Propster et al. [7].

Modifying sialic acid synthetically can increment the selectivity and binding strength of Siglec ligands [15]. An expanded collection of sulfonamide-based C-9 modified sialoside analogs was developed for Siglec-8, resulting in the Neu5Acα2-3[6S] Galβ1-4GlcNAc (6′-O-sulfo α2-3SLacNAc) scaffold decorated with the 9-*N*-(2-naphthyl-sulfonyl) moiety at C9 position (6′-O-sulfo NSA-α2-3SlacNAc) (**Figure 4.1**, compound **2**), which showed improved binding affinity [16]. The estimated K_d for **2** is approximately 40 μM by NMR. The prediction of a 15-fold affinity improvement from adding a naphthylsulfonamide group at the 9-position assumes that this modification has an additive effect.

Developing small molecule ligands for Siglec-8 complements the ongoing work on generating antibodies for this receptor, which is an active field of research [17-20]. Consequently, scientists have produced both synthetic multivalent [21-24] and monomeric ligands [15] for Siglec-8. A recent investigation revealed that synthetic multivalent ligands can group Siglec-8 into clusters, known as microdomains. This clustering is vital for triggering an agonistic biological response [24].

The laboratory of Prof. Paul Murphy from University of Galway (Ireland) have designed new monomeric ligands (**Figure 4.3**, compounds **6-11**). These compounds mimic ligand **1** but have much less carbohydrate character compared to **1** and **2**. They specifically looked at using short triazole spacer groups to connect sialic acid to a charged part. They tested carboxylate and sulfonate groups as alternatives to the sulfate group. They also added a 2-naphthyl sulfonamide group at the C-9 position of sialic acid. This process helped them find ligands that bind to Siglec-8 with affinities in the low to mid micromolar range, as measured by ITC.



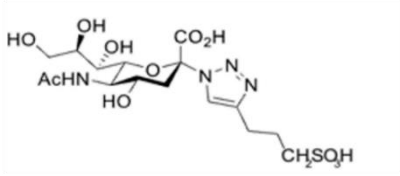
6 R = OSO_3H
7 R = CH_2SO_3H
8 R = CH_2CH_2COOH

9 R = COOH
10 R = CH_2COOH
11 R = CH_2SO_3H

Figure 4.3. Compounds 6-11 synthesized for this work as glycomimetics for Siglec-8. These ligands were provided by Prof. Paul Murphy from University of Galway (Ireland).

Furthermore, in this work the molecular recognition features of the interaction between the carbohydrate recognition V-domain of Siglec-8 with two of these ligands were investigated (**Figure 4.4**, compound **7** and **11**).

Ligand 7



Ligand 11

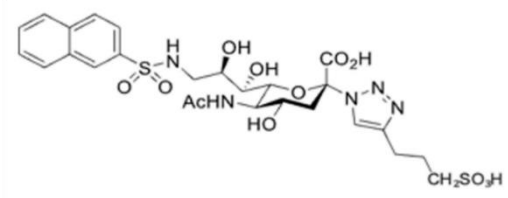


Figure 4.4. Chemical structures of the ligands studied in this work. Sialic acid molecule with a triazole spacer group that connects Sia to a sulfonate group (**7**) and its derivative with 9-*N*-(2-naphthyl-sulfonyl) moiety (**11**).

Two distinct NMR-based approaches were used. First, the protein perspective was examined by observing changes in the lectin's chemical shift perturbation when increasing amounts of the ligands were added. Second, from the ligand's point of view, STD NMR experiments were carried out to elucidate the binding epitopes of the ligands in complex with Siglec-8. Finally, to better understand the structural details of the interaction between Siglec-8 and the ligands, computational studies (docking) were performed to validate the NMR data. This study aims to discover Siglec-8 ligands with higher binding affinity than those previously described, to address the critical need for clinical Siglec-8 blocking agents.

4.2 Generation of the recombinant plasmid pET-43.1(a)-Siglec-8_{d1}

As previously mentioned, the Siglec-8 V-set domain (Siglec-8_{d1}) features three highly conserved cysteine residues; two are utilized in an intra-domain disulfide bond, while the third is typically involved in an inter-domain disulfide bond with the subsequent C2-set domain. To preclude potential misfolding of the recombinant protein due to the formation of incorrect disulfide linkages, the cysteine residue responsible for the inter-domain connection (at position 42) was replaced with a serine residue via site-directed mutagenesis (**Figure 4.5**) [25].

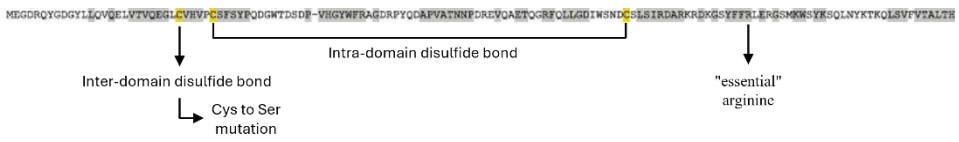


Figure 4.5. Mature protein sequence of the Siglec-8 V-set domain. *Adapted from [25].*

The coding sequence for the Siglec-8_{d1} domain, which incorporates the Cys→Ser point mutation, was subsequently inserted into the pET-43.1(a) expression vector. This insertion utilized the NdeI and XhoI restriction sites. The final construct was designed to include a short two-residue spacer (GlySer) and a thrombin cleavage sequence (LeuValProArg↓GlySer) immediately downstream of the domain [25]. As illustrated in **Figure 4.6**, the resulting fusion protein incorporates a hexahistidine affinity tag (His_6-tag) strategically positioned at the C-terminus. This C-terminal placement was chosen specifically to ensure that the affinity chromatography purification process is restricted solely to fully translated proteins. This engineering choice is critical for preventing the co-purification of any abortive protein products that might result from premature translation termination, thereby guaranteeing the isolation of the full-length target protein [25].

Figure 4.6. Siglec-8 V-set domain expression construct. *Adapted from [25].*

Specifically, both the template plasmid (containing the Siglec-8_{d1} gene) and the pET-43.1(a) vector were enzymatically digested using the NdeI and XhoI restriction endonucleases. Successful completion of the cleavage reaction was confirmed by agarose gel electrophoresis (**Figure 4.7**).

Following the ligation reaction between the DNA insert and the plasmid, the recombinant product pET-43.1(a)-Siglec-8_{d1} was transformed into competent *Escherichia coli* (*E. coli*) TOP10 cells for DNA amplification. Finally, positive

clones were selected and verified through DNA sequencing to confirm the accurate insertion of the desired sequence.

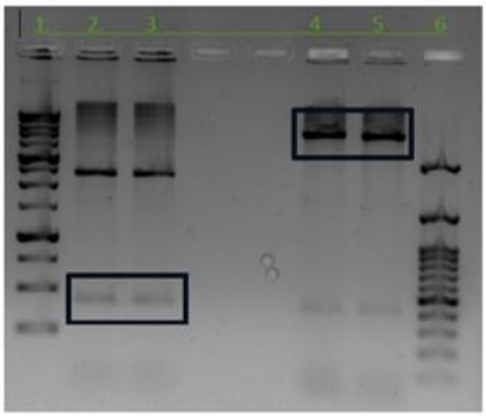


Figure 4.7 Agarose gel electrophoresis of enzymatic digestion. Agarose gel (1%) analysis of the restriction digestion profile of the template plasmid-Siglec8_{d1} construct (lanes 2-3) and the pET-43.1(a) plasmid (lanes 4-5) using the NdeI and XhoI restriction enzymes. The expected bands for the insert (approximately 500 bp) and the pET-43.1(a) vector (approximately 6,000 bp) were obtained. Lanes 1 and 6 contain the 1 Kb DNA Ladder and the 100 bp DNA Ladder markers, respectively.

4.3 Isolation of Siglec-8_{d1}

Based on the methodology outlined by Propster et al. [7], the Siglec-8 carbohydrate recognition domain (V Ig domain) was successfully produced in *Escherichia coli*.

4.3.1 Siglec-8_{d1} purification from *Escherichia coli*

The recombinant plasmid containing Siglec-8_{d1} gene was transformed into *E. coli* Rosetta-gami B (DE3) competent cells (Novagen). The expression of the Siglec-8 lectin domain was tested using *E. coli* Rosetta-gami B (DE3) to find the best expression conditions for Siglec-8. SDS-PAGE electrophoresis was used to analyze whether the protein was in the soluble (supernatant), insoluble (pellet), or both fractions. In this regard, Siglec-8_{d1} was identified in both the portions, as indicated by a 16 kDa band in the SDS-PAGE gel (data not shown). Therefore, because the protein exhibited insolubility, denaturation and refolding protocol from inclusion bodies was required [26-28]. Siglec-8 V Ig domain was then overexpressed and purified using an affinity chromatography for the His-tag located at the C-terminus. The protein was eluted with 500 mM imidazole, concentrated and the monomer was separated from other contaminants and possible aggregates using size exclusion chromatography (Superdex 75, 10/300) (**Figure 4.8**). The monomeric and pure protein eluted as a single peak at mL 17, consistent with a globular protein of 16 kDa. Protein purity was confirmed through SDS-PAGE (15%) analysis (**Figure 4.8**). To confirm proper protein folding, ^{1}H-^{15}N HSQC NMR experiments were employed.

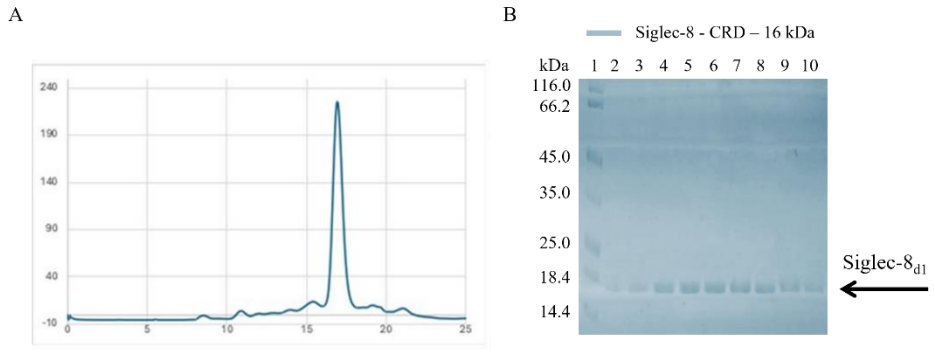


Figure 4.8. Purification of Siglec-8_{d1}. (A) The gel filtration chromatogram (Superdex 75, 10/300) confirms that Siglec-8 is present as a monomer. (B) SDS-

PAGE analysis of the Siglec-8 eluted fractions obtained from the gel filtration column: lane 1 displays the molecular weight marker (MW marker), lane 2-10 represents the protein in the elution peak.

A final protein yield of 2 mg/L was achieved for both unlabeled and ^{15}N-labeled Siglec-8_{d1}. This indicated that the labeling conditions had no impact on either the protein's expression levels or its stability.

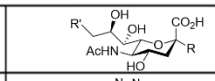
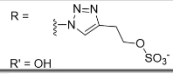
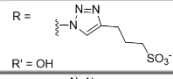
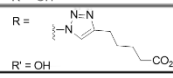
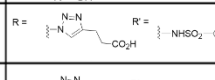
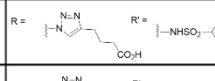
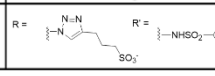
4.4 Binding Analysis of Siglec-8 with Glycomimetic Ligands

4.4.1 ITC study of Siglec-8-ligands binding

ITC was used to determine the binding affinity (K_d) of both ligands (**7** and **11**) with Siglec-8 and identify their thermodynamic fingerprint. These experiments were performed by Prof. Beat Ernst and their collaborators from University of Basel, in Basel (Switzerland) and they performed ITC experiments for all the compound in **Figure 4.3**.

Interestingly, ligand **7**, a sulfonate and triazole derivative, showed no significant difference, with a K_d of 648 μM, compared to the natural ligand 6'-O-sulfo sialyl Lewis^x. A similar pattern emerged with the naphthyl-substituted analogue of **7** (ligand **11**), also analyzed in this work, which showed approximately 2.5-fold higher affinity than the carboxylic acid derivatives **9** and **10** (**Table 4.1**). This confirmed that the sulfonate group enhances affinity compared to carboxylic acids and acts as a chemically stable sulfate surrogate. Sulfates and sulfonates can form a greater number of hydrogen bonds than carboxylates. Additionally, strong hydrogen bonds are known to occur between sulfate oxygen atoms and ammonium ions [29]. Electronic factors likely contributed to stronger interactions with the sulfonate or sulfate group, leading to the observed improvement in binding.

Table 4.1. Thermodynamic parameters for Siglec-8 ligands using ITC.

Compd No.		K_d [μM]	$\Delta G°$ [kJ/mol]	$\Delta H°$ [kJ/mol]	$-T\Delta S°$ [kJ/mol]
6		700 (594 to 836)	-18.0 (-18.4 to -17.6)	-37.5 (-45.5 to -31.6)	19.5 (13.1 to 27.9)
7		648 (457 to 967)	-18.2 (-19.1 to -17.2)	-21.6 (-33.1 to -15.3)	3.4 (-1.9 to 14.1)
8		*n.b.	—	—	—
9		131 (111 to 156)	-22.2 (-22.6 to -21.7)	-15.3 (-17.0 to -13.9)	-6.8 (-7.8 to -5.6)
10		116 (102 to 131)	-22.5 (-22.8 to -22.2)	-19.3 (-20.6 to -18.0)	-3.2 (-4.1 to -2.2)
11		45.5 (43.9 to 47.2)	-24.8 (-24.9 to -24.7)	-26.2 (-26.5 to -25.8)	1.4 (1.1 to 1.6)

95% confidence intervals are given in parentheses; *n.b.: no binding visible under conditions used.

Furthermore, and in line with prior research, the naphthylsulfonamide group also provided a clear benefit, as evidenced by compound **11**'s approximately 14-fold higher affinity compared to compound **7**. Indeed, ligand **11** exhibited a K_d of 45 μM. The improvement in affinity was attributed again to advantageous π-π and CH-π interactions. Conversely, the negative entropic contribution of ligand **7** was due to ligand reorganization in the binding pocket.

Thus, ITC data, consistent with the NMR experiments discussed in the following paragraphs, revealed a lower K_d value for ligand **11** compared to ligand **7**. This result confirmed that the presence of the aromatic group at C9 of Neu5Ac enhances binding to Siglec-8.

Based on these results, the NMR analysis was extended for compound **7** and **11**.

4.4.2 NMR investigation of binding from the ligands point of view

The investigation into molecular recognition started with ligand-based NMR interaction analysis. Initially, the ^{1}H and ^{13}C NMR signals for the corresponding nuclei of the ligands were assigned. This was accomplished using standard 1D and 2D NMR experiments. The complete NMR assignments for ligands **7** and **11** are presented in **Tables 4.2** and **4.3**, respectively.

Table 4.2. ^{1}H and ^{13}C NMR assignments for compound 7.

Ligand	Position/atom	^{1}H (δ, ppm)	^{13}C (δ, ppm)
Sialic acid (B)	$H3_{eq}$	3.17	39.7
	$H3_{ax}$	2.18	
	4	3.56	68.1
	5	3.92	51.6
	6	3.88	68.0
	NAc	1.99	22.2
	7	3.81	71.3
	8	3.86	74.0
	$H9_1$	3.56	62.6
	$H9_2$	3.80	
3-(1H-1, 2, 3-triazol-4-yl) propane-1-sulfonic acid (A)	1	7.96	120.7
	2	2.03	23.9
	3	2.81	23.5
	4	2.87	50.2

Table 4.3. ^{1}H and ^{13}C NMR assignments for compound 11.

Ligand	Position/atom	^{1}H (δ, ppm)	^{13}C (δ, ppm)
Sialic acid (B)	$H3_{eq}$	3.09	39.7
	$H3_{ax}$	2.09	
	4	3.47	69.1
	5	3.82	51.6
	6	3.82	67.9
	NAc	1.96	22.2
	7	3.68	69.8
	8	3.81	74.0
	$H9_1$	3.01	45.3
	$H9_2$	3.35	
Naphthyl group (C)	N1	8.45	128.2
	N3	8.05	129.8
	N4	7.80	121.4
	N5	7.97	127.8
	N6	7.69	129.3
	N7	7.63	127.9
	N8	8.00	129.1
3-(1H-1, 2, 3-triazol-4-yl) propane-1-sulfonic acid (A)	1	7.67	120.3
	2	1.93	23.9
	3	2.65	23.4
	4	2.8	50.1

The STD NMR experiments were employed to disclose the binding epitopes for these two ligands. These studies clearly showed that the terminal Neu5Ac residue produced robust STD NMR signals, identifying it as a primary contact point with Siglec-8. For compound **7**, H4, H5 and H6 showed strong STD response. Furthermore, the triazole hydrogen also exhibited STD effect, suggesting that it's involved in the binding process (**Figure 4.9**).

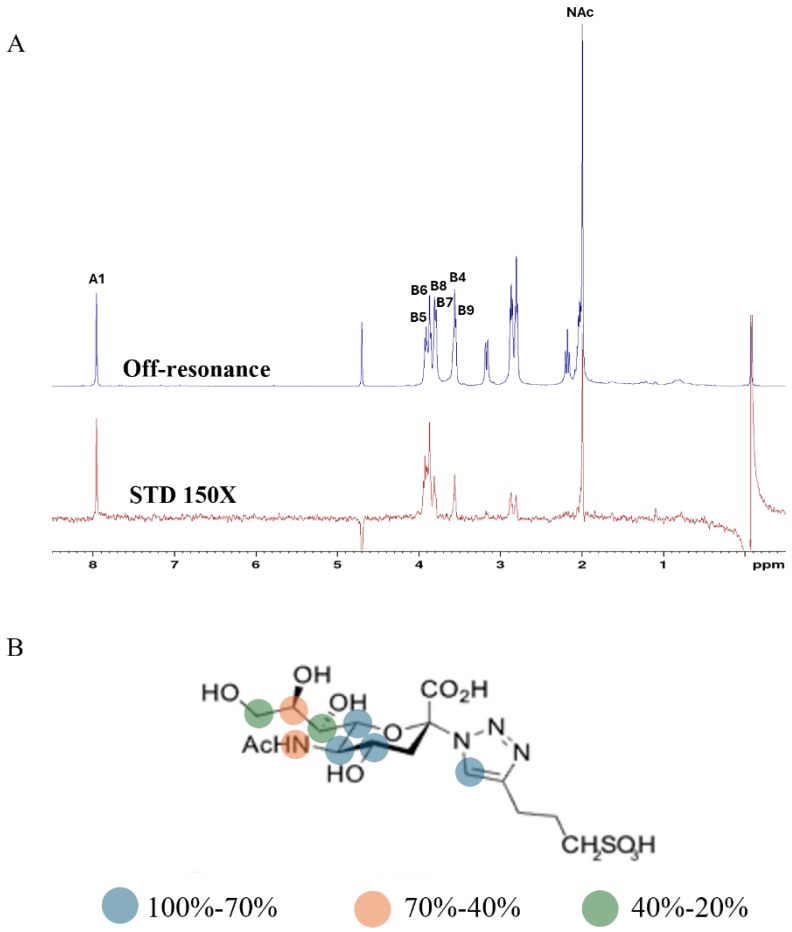


Figure 4.9. STD NMR analysis of the interaction between compound 7 and Siglec-8. (A) Off-resonance (*top*) and STD (*bottom*) NMR spectra illustrating the interaction between ligand **7** and Siglec-8_{d1} (1:50 molar ratio). The inset details the proton assignments. (B) The scheme depicts the binding epitope, as determined by the observed STD intensities.

The STD NMR data for ligand **11**'s sugar protons revealed a binding epitope similar to that of compound **7**. Consistently, the naphthyl group also displayed very high STD signals. These observations powerfully suggested that while the

sialic acid sits in the Siglec-8 binding pocket similarly to ligand **7**, the naphthyl ring also plays a direct role in the interaction (**Figure 4.10**).

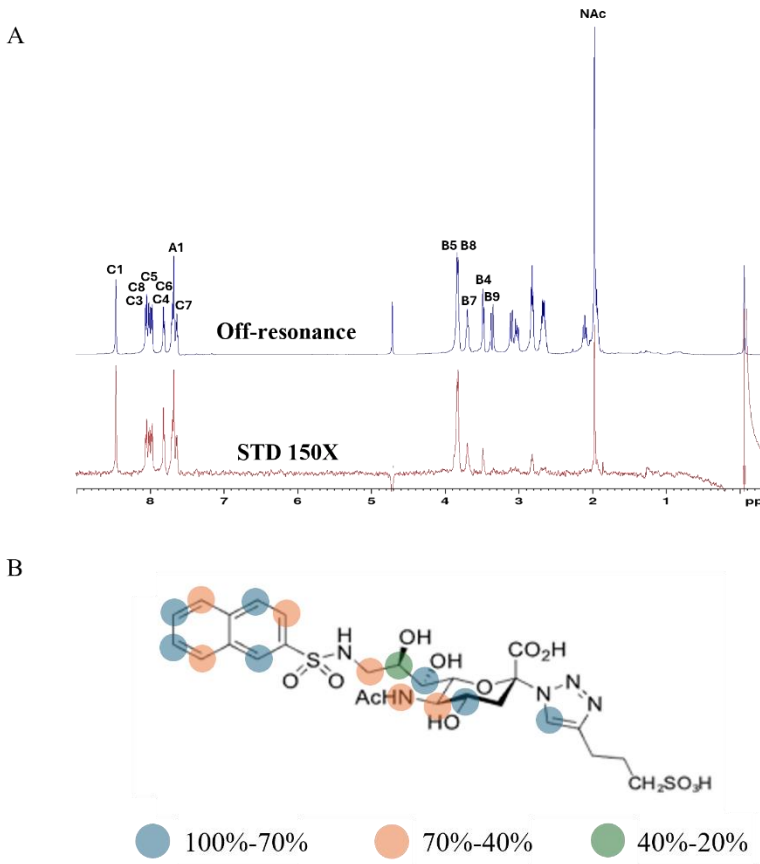


Figure 4.10. STD NMR analysis of the interaction between compound 11 and Siglec-8. (A) Off-resonance (*top*) and STD (*bottom*) NMR spectra illustrating the interaction between ligand **11** and Siglec-8_{d1} (1:50 molar ratio, on-resonance set in the aliphatic region). The inset details the proton assignments. (B) The scheme depicts the binding epitope, as determined by the observed STD intensities. The sugar-binding epitope is essentially the same for both compounds. However, the naphthyl moiety of compound **11** plays a significant role in its interaction with the lectin.

4.4.3 Exploring recognition of the ligands from the protein point of view

Following the ligand-based NMR experiments, the recognition of these two ligands (**7** and **11**) with Siglec-8 was explored from the protein's perspective. Chemical shift perturbation analysis [30], derived from ^{1}H-^{15}N HSQC NMR spectra [31-32], provides valuable information on the location of a ligand's binding site on a protein. The spectra were obtained for the ^{15}N-labelled Siglec-8 V-set domain. In a HSQC spectrum (**Figure 4.11**), each cross-peak corresponds to the NH amide group of a single amino acid (except for proline).

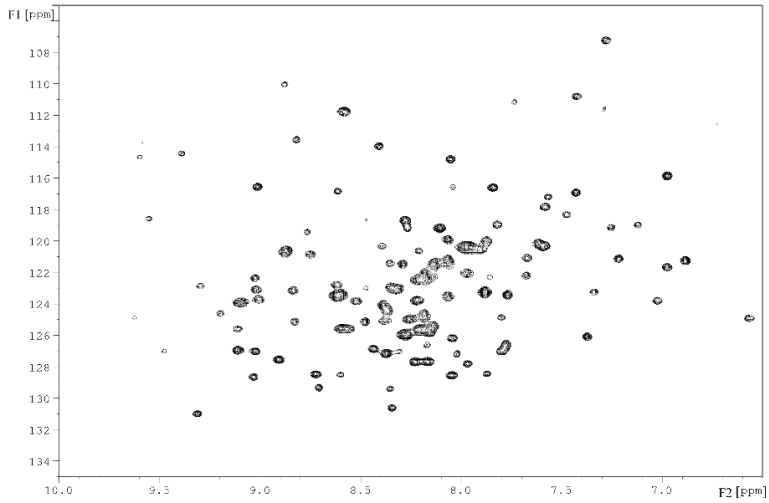


Figure 4.11. ^{1}H-^{15}N HSQC NMR data for Siglec-8 V-set domain (apo form).

The unique chemical environment of each amino acid within the protein's sequence and 3D structure give rise to a characteristic and unique 2D spectrum, which can be thought of as a "protein fingerprint". These cross-peaks were assigned based on the previous work by Propster et al. [7].

A series of ^{1}H-^{15}N HSQC spectra were recorded for the apo form of the protein (**Figure 4.11**) and with increasing concentrations of various ligands. When a ligand binds, it alters the chemical environment of the amino acid residues directly or indirectly participating in the binding event. Induced CSPs directly indicate both the presence of binding, and which residues are most significantly impacted by the ligand's addition. Indeed, by analyzing these changes, it was possible to identify the protein-binding site, determine ligand affinity, and potentially elucidate the structure of the protein-ligand complex.

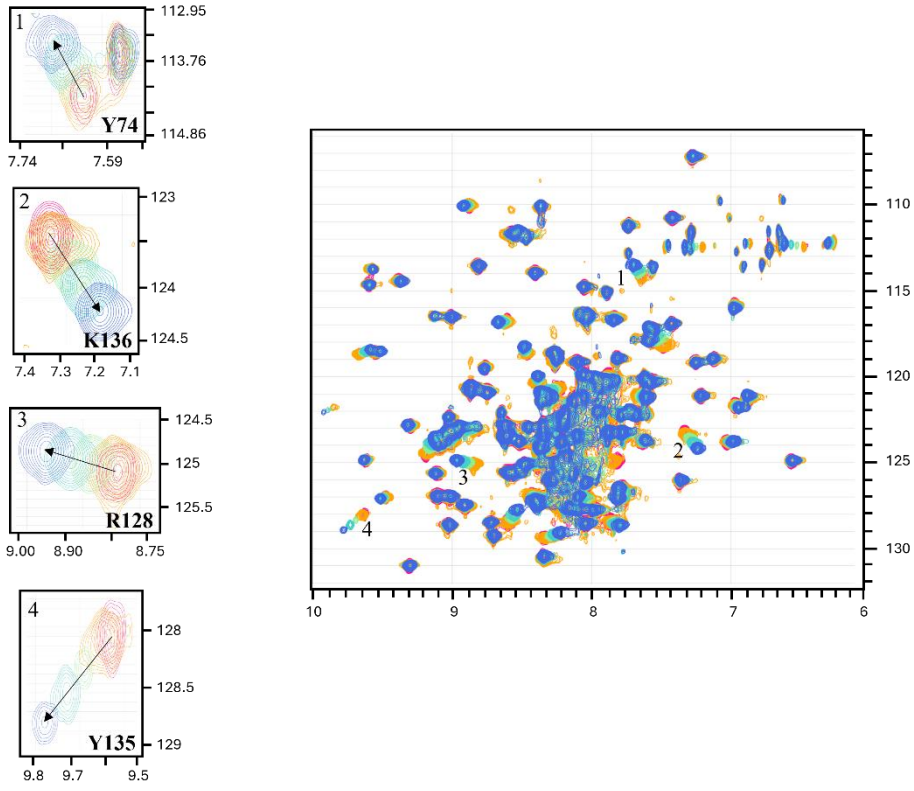


Figure 4.12. CSP analysis of the interaction of Siglec-8 V domain with ligand 7. Detailed views of the specific perturbations in key residues are provided in the insets.

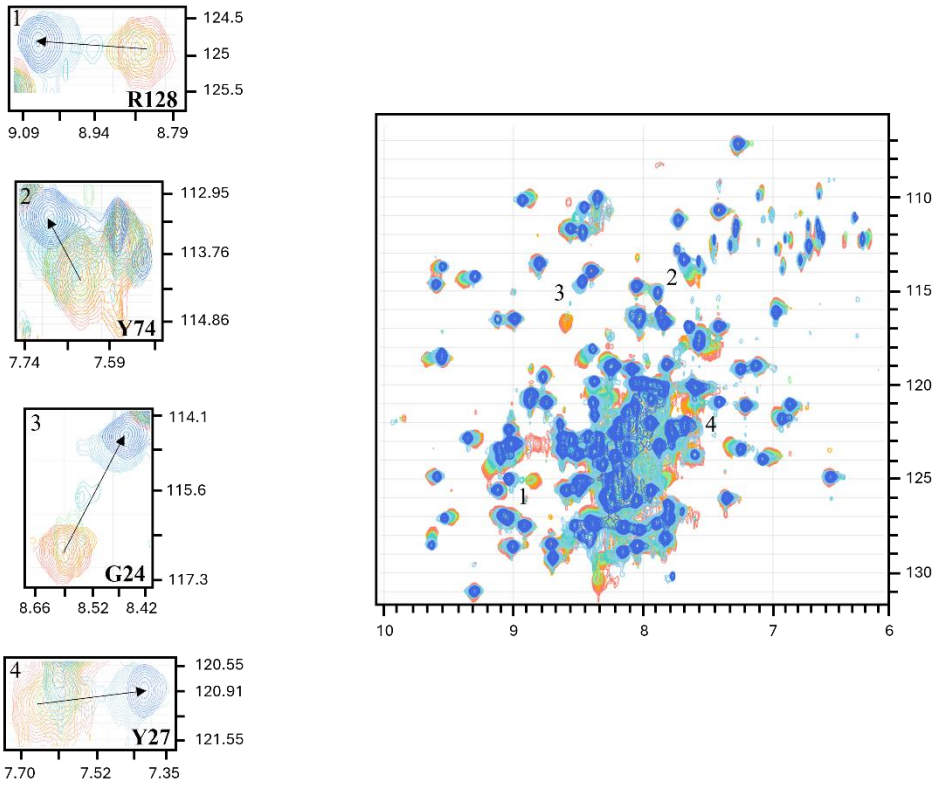


Figure 4.13. CSP analysis of the interaction of Siglec-8 V domain with ligand 11. Detailed views of the specific perturbations in key residues are provided in the insets.

Specifically, residues R125, L126, R128, K136, W133, S134 and Y135, displayed significant CSPs in the binding with ligands **7** and **11** (**Figure 4.12** and **4.13**, respectively), suggesting that these residues are located at the sialic acid binding site of the protein, which include the F and G β-sheets (carbohydrate recognition domain). As reported by literature [7], the canonical arginine, R125, situated within the binding site, forms a salt bridge with the sialic acid's carboxyl group at position 1. Furthermore, Propster et al. also suggested that the remaining polar groups on the sialic acid's pyranose ring form hydrogen bonds with residues K132 and S134. Moreover, CSPs were also observed in the C-C' loop located in front

of the binding site. Indeed, residues such as R72, Y74 and Q75 within the loop C-C' exhibited important CSPs (**Figure 4.12** and **4.13**), suggesting a direct interaction of the sulfate and sulfonate groups of these compounds with Siglec-8 at this specific region, as also previously reported [7, 16]. A fast/intermediate exchange was noted for most cross-peaks during the titration, in agreement with the observations reported in reference [7].

Additionally, with ligand **11**, further CSPs were observed in the N-terminal region and within the G-G′ loop, indicating that the naphthyl group at C9 of the Neu5Ac residue was positioned close to the N-terminus. Significant perturbations for residues S137, Q138, L139, and K142, as well as residue G24, were identified (**Figure 4.13**). This glycomimetic exhibited also a slow-intermediate exchange on the chemical shift timescale for specific residues in the G-G' loop, including K132, S137, Q138, and L139 (**Figure 4.14**).

Interestingly, the analysis of the chemical shift perturbations of the protein cross peaks showed that saturation was achieved with only 8 equivalents of compound **11**, a notable difference compared to ligand **7**, which required 30 equivalents to saturate the lectin (**Figure 4.14**).

In accordance with the observations reported [16], this feature suggested that the introduction of the naphthyl group significantly enhances binding affinity. Moreover, these findings aligned with earlier crystallographic data for compound

2, which illustrated the C-9 naphthyl residues orienting towards the G-G' loop [16].

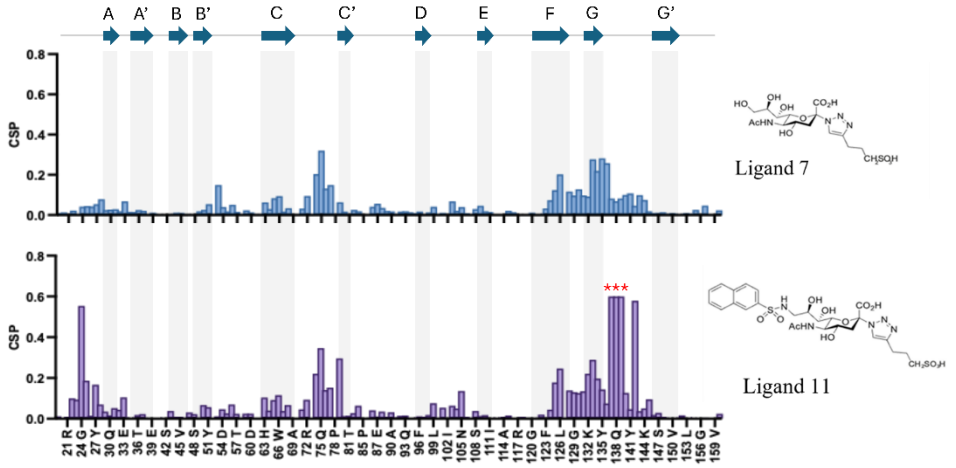


Figure 4.14. HSQC cross-peak CSP (Δδ) plots for Siglec-8 V domain amino acid residues, obtained at saturation with ligands 7 (30 equivalents) and 11 (8 equivalents). At the top of the figure, the secondary structural elements are shown, as determined from the crystal structure of Siglec-8 in an earlier study [16]. Residues displaying prominent signal broadening during the titration are denoted by an asterisk (*).

Taken together, the NMR binding data highlighted that residues in the F and G β-strands within the Siglec-8 binding site play a crucial role in mediating interactions with these two glycomimetics and modulating their binding affinity.

4.4.4 Structural analysis of Siglec-8-ligand complex

The experimental data, analyzed from both the protein and ligand perspectives, was used to construct 3D structure for the complex of Siglec-8 with ligand **11** (**Figure 4.16**). The N-terminal domain of Siglec-8 in complex with sulfonamide sialoside analogue (*PDB 7QUI*) from X-Ray crystallographic structure by Lenza et al. [16] was employed as model.

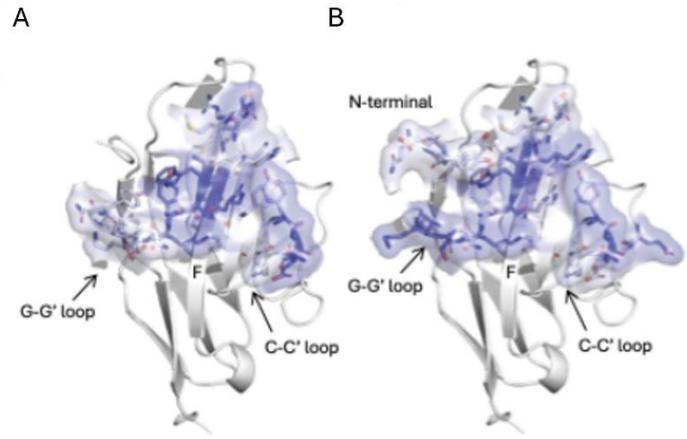


Figure 4.15. 3D views of Siglec-8 structure (*PDB 7QUI*). Amino acids that exhibit stronger (blue) and fewer (light blue) CSP effects upon binding to ligand **7** (A) and **11** (B) are shown as sticks. The CC' and GG' loops within Siglec-8, essential for ligand recognition, are indicated with arrows.

Notably, the findings for the Siglec-8 complexes aligned with the NMR data, including results from both CSP (**Figure 4.15**) and STD experiments. All interactions identified in the docking study of the ligand **11** were consistent with previously reported MD simulation studies and the crystal structure complex of compound **2** with Siglec-8 [16].

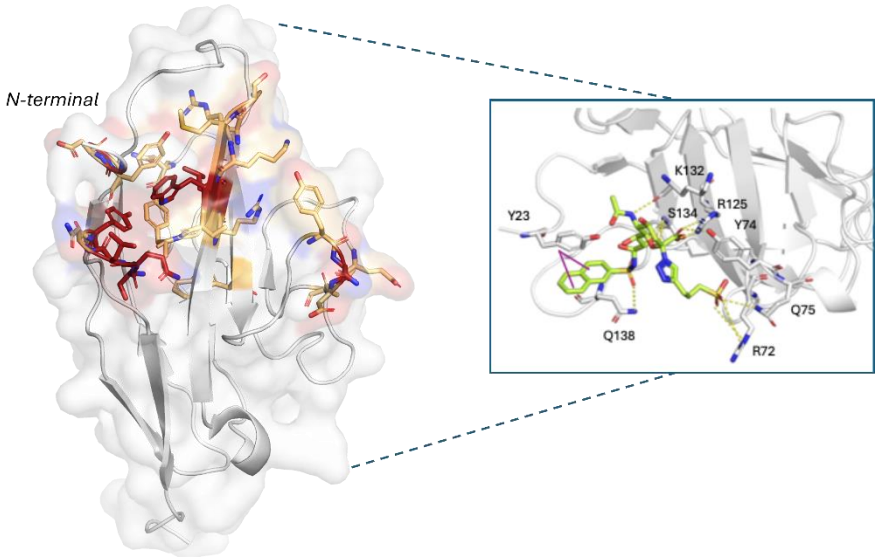


Figure 4.16. Structural insights into Siglec-8-ligand 11 interactions. (*right*) 3D model of Siglec-8 V domain in complex with compound **11** (in green) highlights only amino acid residues that participate in hydrogen bonding (yellow dotted lines) and hydrophobic interactions (solid pink lines) as white sticks. (*left*) Based on the crystallographic structure of Siglec-8 [16], the model shows the observed stronger (red) and fewer (yellow) CSPs resulting from titration with ligand **11**.

In particular, the binding mode of Siglec-8 with compound **11** highlighted that the carboxylate group of Neu5Ac formed a salt bridge with conserved R125 and hydrogen bonding interactions with the hydroxyl group of Y74 (**Figure 4.16**). Moreover, the glycerol chain and the acetamide group were involved in hydrogen bonds with residues K132 and S134, respectively, in the G β-sheet. The sulfonate group engaged in salt-bridge interactions with R72 and formed hydrogen bonds with Q75 and Y74. Additionally, one oxygen atom of sulfonamide formed a hydrogen bond with Q138. Moreover, the naphthyl ring established a π-π interaction with Y23 within a hydrophobic aromatic pocket. Indeed, the naphthyl group at C9 of the Neu5Ac residue was located close to the N-terminus (**Figure 4.16**).

The integration of the experimental data with the structural model provides a coherent description of the interaction between Siglec-8 and the ligand **11**. The CSP analysis highlighted a cluster of perturbed residues located around the canonical sialic acid binding pocket of Siglec-8, which corresponds to the contact region identified in the 3D model. Consistently, STD NMR experiments revealed the strongest STD enhancements for the ligand protons oriented toward the binding cleft in the structural model, supporting the predicted ligand orientation. Moreover, the dissociation constant measured by ITC aligns with the number and type of non-covalent interactions observed in the modeled complex. Altogether, these independent techniques converge toward a unified binding mode in which the ligand engages the canonical Siglec-8 pocket through a combination of electrostatic and hydrophobic contacts, validating the proposed structural model.

4.5 Discussion and Conclusions

Siglec-8 is known to downregulate inflammatory responses driven by eosinophils and mast cells when it binds to specific sialylated glycans. This immune receptor is seen as a compelling target for new anti-inflammatory treatments for conditions like asthma, where uncontrolled or chronic inflammation from these cell types contributes to the disease.

The NMR structure of the Siglec-8 V domain has already been determined, both on its own (apo form) and when bound to its preferred glycan (6′-sulfo sialyl Lewis^x) [7], and its interaction with a glycomimetic (6′-O-sulfo NSAα2-3SLacNAc) has been examined using solution NMR spectroscopy [16].

This work focuses on the synthesis of novel sialylated triazoles as glycomimetics for Siglec-8, in which the 6-O-sulfo-Gal was substituted by a triazole modified with a linker to sulfate, sulfonate and carboxylic acid groups. Sulfonate groups

proved to be effective substitutes for sulfates, demonstrating greater efficacy than carboxylates. Additionally, the integration of naphthylsulfonate groups at C-9 of the sialic acid enhanced binding affinity, a finding that aligns with previous studies [16]. Consistent with ITC measurements, the NMR study validated the binding affinity trends observed for the two sialylated triazoles and provided detailed insights into the interactions between the ligands and Siglec-8, highlighting how the protein adopts a conformational change to accommodate 9-naphthylsulfonate.

The initial step involved successfully expressing and purifying the Siglec-8 V domain from *E. coli* Rosetta-gami B (DE3) cells, in both labeled and unlabeled forms.

$^{1}H,^{15}N$- TROSY experiments were conducted on the ^{15}N-labeled protein, in its unbound state and following the incremental addition of two ligands (**7** and **11**). Notably, both ligands displayed a consistent CSP profile, indicating perturbation of the F-G β-sheet, which constitutes the Siglec-8 binding pocket. Furthermore, the presence of sulfate and sulfonate groups induced perturbations in a previously observed region, the C-C' loop, situated opposite the binding site. An examination of ligand **11**, featuring a naphthylsulfonate group at C9, revealed CSPs in additional residues located within the N-terminal region and the G-G' loop. This observation was significant, as the naphthyl group of the Neu5Ac residue was situated close to the N-terminus. Similarly, these data align with the results previously published by Lenza et al. [16].

Additionally, STD NMR experiments also offered insights into the binding epitope from the ligand's perspective. These studies revealed that the terminal Neu5Ac residue of both ligands produced strong STD NMR signals, establishing it as a primary site of recognition with Siglec-8.

Moreover, based on the experimental data from both CSP and STD, a 3D model for the protein complex with compound **11** was developed.

ITC data highlighted the thermodynamic differences triazoles glycomimetics. The introduction of the 9-substituent into the triazoles led to a similar improvement in affinity. This crucial finding indicates that the Siglec-8 conformational change needed to accommodate the naphthyl groups binding didn't result in an entropy penalty. Moreover, despite carboxylic acid groups sometimes being effective replacements for sulfate/sulfonate inhibitors [33], this study observed no binding for carboxylate **8** compared to its sulfate (**6**) and sulfonate (**7**) counterparts. This strongly suggests that a sulfate or sulfonate group is essential for optimal Siglec-8 interaction, likely due to their superior hydrogen bonding or electrostatic capabilities.

In this context, a comparative analysis between the glycomimetic ligands employed in this study and the natural Siglec-8 ligand (6′-sulfo-sialyl-Lewis^x) further highlights how specific structural features influence lectin selectivity. The native epitope contains a sulfated galactose unit and adopts a well-defined orientation that enables productive interactions within the canonical binding pocket of Siglec-8. In these synthetic triazole-based analogues, the replacement of the 6-O-sulfate with alternative acidic or aromatic substituents preserves the overall geometrical arrangement of the ligand while modulating the strength and nature of the interactions with key residues of the pocket. These modifications provide insight into how subtle changes at the C-6 position of the galactose mimic, or within the triazole scaffold, can either maintain or alter recognition by Siglec-8. Importantly, understanding how such structural variations affect affinity and specificity offers a valuable framework for the rational design of next-generation glycomimetics. This comparative perspective underscores how fine-tuning the

chemical features of the ligand may be strategically exploited to develop more selective and potent Siglec-8 modulators with potential therapeutic relevance.

For these reasons, these findings establish a foundation for the continued development of Siglec-8 inhibitors. Specifically, this work indicates that 9-substitued sialyl triazoles derivatives containing sulfonates, molecules significantly less carbohydrate-like than glycans such as compounds **1** and **2**, could serve as a foundation for a medicinal chemistry initiative. This program aims to identify Siglec-8 ligands with greater affinity than those currently known, to address the clinical need for effective Siglec-8 blockers.

Chapter 5: Elucidating Galectin-3 interaction with a selenoglycomimetic at the structural level

5.1 Introduction

As described in the introduction, Galectins are a family of proteins that bind to carbohydrates, and they play a central role in many physiological and pathological processes. These proteins show a specific attraction to galactosides [34]. Found in various cell types like skin, lung, and gut cells, and fibroblasts, galectins are involved in both intracellular and extracellular activities [35].

Among galectins, Galectin-3 (Gal-3) is especially prominent. It has a single CRD connected to a flexible N-terminal domain rich in proline and glycine, which allows it to form a complex oligomer. Present in the cell's nucleus and cytoplasm, and released outside the cell, Gal-3 controls important cellular interactions, from modulating immune response to regulating cell adhesion, migration, and the formation of the new blood vessels.

Because Galectins can bind to sugar-containing molecules found on cell surfaces, in the spaces between the cells, or inside cell compartments, their dysregulation is linked to various diseases, such as the spread of cancer, autoimmune disorders, obesity, heart disease, allergies and infections [36-37]. For these reasons, there's been a growing interest in targeting galectins for medical treatments over the past few decades.

Gal-3 inhibitors have emerged as a promising class of therapeutic agents, especially for treating neoplastic diseases [38-39]. Given Gal-3's pivotal role in promoting fibrotic and tumorigenic pathways, there has been a concerted effort to identify compounds capable of precisely modulating protein activity.

Since galectins' functions are largely associated with their carbohydrate-binding properties, inhibiting their CRD with antagonists that compete with natural ligands appears to be a viable strategy, which could not only help clarify their exact, often unknown, functions but also lead to the development of new molecules for therapeutic intervention.

A variety of sugar-based inhibitors have been created, with some currently progressing through Phase I or II clinical trials for different cancers [40-41]. One notable example is the monovalent inhibitor TD139 (from Galecto Biotech). It has demonstrated excellent affinity for Gal-3, with binding affinity determined by ITC at 0.068 μM ± 0.01. Structurally, TD139 is an optimized version of thiodigalactoside (TDG), featuring two identical 4-fluorophenyl-triazole groups at positions C3 and C3'. TDG itself was an interesting staring point because replacing the oxygen in its glycosidic bond with sulfur significantly improved its enzymatic and hydrolytic stability. Indeed, TDG not only resists hydrolysis but also shows substantial affinity for various galectins, making it an ideal scaffold for developing further derivatives [42]. Moreover, several studies have confirmed TD139's anti-inflammatory effects, by reducing Galectin-3 expression [43-44]. Both preclinical and clinical investigations have shown TD139 seemed to be remarkably effective in treating pulmonary fibrosis and in modifying the tumor microenvironment by decreasing tumor cell growth, invasion, and metastasis [45-46]. However, its progress was unfortunately hampered by side effects observed during clinical trials [47]. This issue highlights the critical need to explore alternative strategies for inhibiting Galectin-3.

In this context, beyond traditional inhibitors, there's a growing interest in how selenium-containing compounds can affect Galectin-3. Specifically, replacing sulfur with selenium as the bridging atom created a new type of diglycosylated analogue, featuring a unique selenoglycoside bond between two galactose units

(SeDG) [48]. This change proved highly beneficial, for example it improves protease stability along with intrinsic antioxidant and pro-oxidant properties [49-51].

Researchers have investigated the interaction of this compound with human galectins, especially Galectin-1 and Galectin-3 [52]. SeDG was evaluated for its ability to bind and inhibit galectins activity through various biological and biophysical techniques. These results showed that SeDG can strongly interact with Gal-3, suggesting its potential as a therapeutic agent for conditions correlated to Gal-3 overexpression, like fibrosis and cancer.

A new version of SeDG, called benzyl 3,3'-seleno-digalactoside (SeDG-Bn), which incorporates a lipophilic benzyl group at C-3 of both sugar units, was recently found to bind both Gal-3 CRD and Gal-9N CRD [53]. Furthermore, SeDG-Bn demonstrated abilities to slow cell growth and migration in a melanoma cell line, as well as anti-angiogenesis activities. Given its notably high affinity and demonstrated anti-migration and anti-angiogenesis properties, this compound represents a promising lead for the continued development of anti-tumor agents [53].

This chapter investigates how Gal-3 interacts with selenium-containing ligands, focusing on SeDG and SeDG-Bn (**Figure 5.1A** and **B**, respectively). A version of this chapter has been published in *ChemMedChem* as: M.P. Lenza et al. "Structural insights into Galectin-3 recognition of a selenoglycomimetic", *ChemMedChem*, 2025 [54]. Both compounds were synthesized by Professor Iadonisi's group at the University of Naples Federico II, Italy. Interestingly, the binding between Gal-3 and SeDG was analyzed using NMR, only from the protein's perspective. For SeDG-Bn, however, the molecular interaction details with Gal-3 were studied using NMR from the protein and ligand point of view. Furthermore, by combining NMR data with computational studies, it was possible to elucidate the 3D structure

of the complex formed when SeDG-Bn interacts with Gal-3. This comprehensive approach allowed assessment of the recognition and binding process, thereby offering a deeper understanding of the molecular mechanisms behind its inhibitory action.

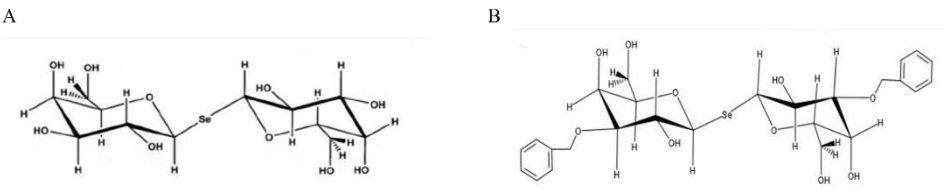


Figure 5.1. Chemical structures of (A) SeDG and (B) SeDG-Bn.

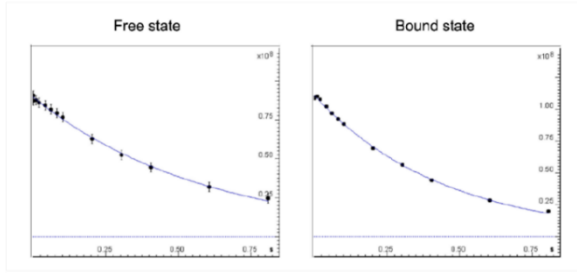
5.2 Galectin-3 interaction studies with selenium-containing compounds

5.2.1 Epitope characterization of SeDG-Bn using ligand-based NMR

Evidence for SeDG-Bn binding to Gal-3 was first seen in a decrease in ligand proton T_2 relaxation times during CPMG experiments (**Figure 5.2**). This finding was further supported by STD NMR experiments, which provided a detailed look into the recognition and binding process and helped map the specific binding epitope (**Figure 5.3**).

The STD NMR results confirmed that Gal-3 CRD recognizes SeDG-Bn, as indicated by the enhancements observed in the mixture Gal-3: SeDG-Bn. The STD NMR analysis demonstrated that Gal-3 primarily interacts with the aromatic protons, although the galactose protons also contribute to the binding.

A



B

T2 free (s)	Error	T2 bound (s)	Error
0.478	0.032	0.393	0.034

Figure 5.2. CPMG NMR analysis of SeDG-Bn in both its free and bound states. (A) The graph shows an example of T_2 decay for the galactose proton at position 4. (B) The T_2 values in the table were calculated by averaging the signal integrations of several different protons. *Adapted from [54].*

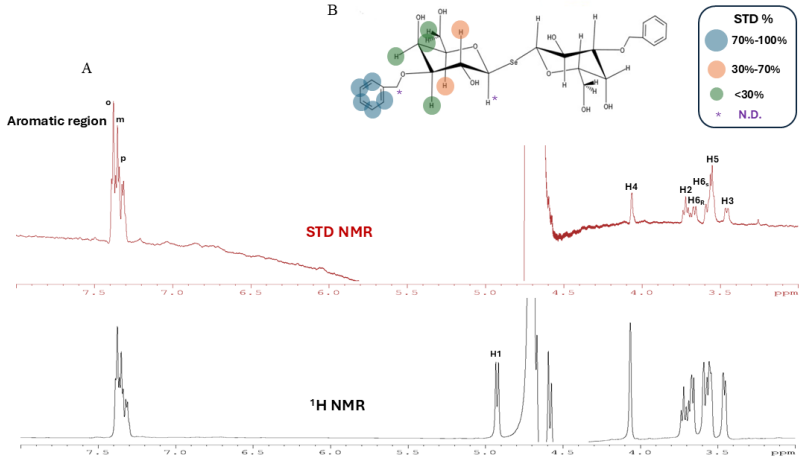


Figure 5.3. STD NMR analysis of the interaction between SeDG-Bn and Gal-3. (A) The STD NMR spectrum (red) and the off-resonance spectrum (black) from a 1:50 mixture of Gal-3 and SeDG-Bn. (B) The binding epitope of SeDG-Bn was mapped by integrating STD NMR intensity data with computational analysis. *Adapted from [54].*

Compared to the reference spectrum, the aromatic ring showed a stronger STD response than the sugar part, with the highest STD effect, as shown in **Table 5.1**. The protons at positions 2 and 5 of the galactose unit revealed moderate STD responses and the protons at positions 3, 4 and 6 exhibited weak STD effects (**Table 5.1**).

Table 5.1. Experimental STD epitope for SeDG-Bn bound to Gal-3 CRD (related to Figure 5.3). The chemical shifts values (^{1}H ppm) for the protons of the SeDG-Bn ligand were assigned based on HSQC NMR experiment. STD epitope (%): the strongest STD signal (the aromatic protons) was assigned to a value of 100%, the other STD intensities were calculated accordingly.

^{1}H	ppm	STD epitope (%)
H2	3.7297	30
H3	3.4802	21
H4	4.0760	25
H5	3.5677	46
H6	3.6812	21
H6’	3.6043	21
Aromatic ring	7.3922	100

The low STD signals for the proton at position 1 and the methylene group of the benzyl substitution were difficult to measure accurately because they were too close to the water signal. However, the symmetric nature of the ligand made it impossible for NMR alone to differentiate between the signals from the two aromatic rings and the two galactose units. These limitations prevented a precise determination of the ligand’s epitope map. To overcome this issue, the STD NMR data was integrated with computational methods to describe the epitope mapping, this latter further discussed below.

5.2.2 Characterization of the Gal-3 binding site interacting with SeDG and SeDG-Bn via protein-based NMR

The binding site of Gal-3 was accurately investigated to understand its interaction with the ligands SeDG and SeDG-Bn. Using 2D ^{1}H-^{15}N HSQC NMR spectroscopy, the chemical shift perturbations [30] of the ^{15}N-labeled CRD of Gal-3 were analyzed. The protein's backbone assignments were based on previous work by Umemoto et al. [55].

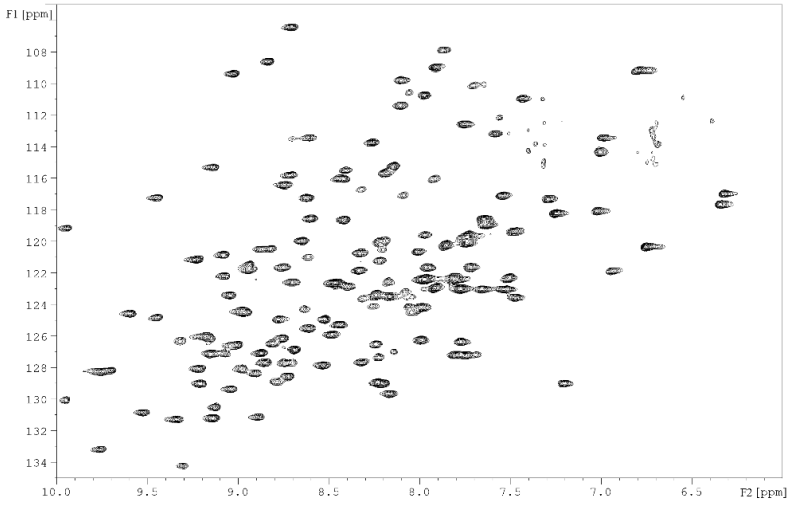


Figure 5.4. ^{1}H-^{15}N TROSY NMR data for Galectin-3 CRD (apo form).

To monitor the binding, CSP experiments were performed by titrating both SeDG and SeDG-Bn into the ^{15}N-labeled Gal-3 sample. Specifically, first the protein in its unbound state was analyzed (**Figure 5.4**), following by increasing concentrations of SeDG (**Figure 5.5**) and SeDG-Bn (**Figure 5.6**).

Changes in the chemical shifts were observed, particularly for amino acid residues located at the binding interface. This method allowed comparison of how the two selenium-containing compounds were accommodated within the binding site.

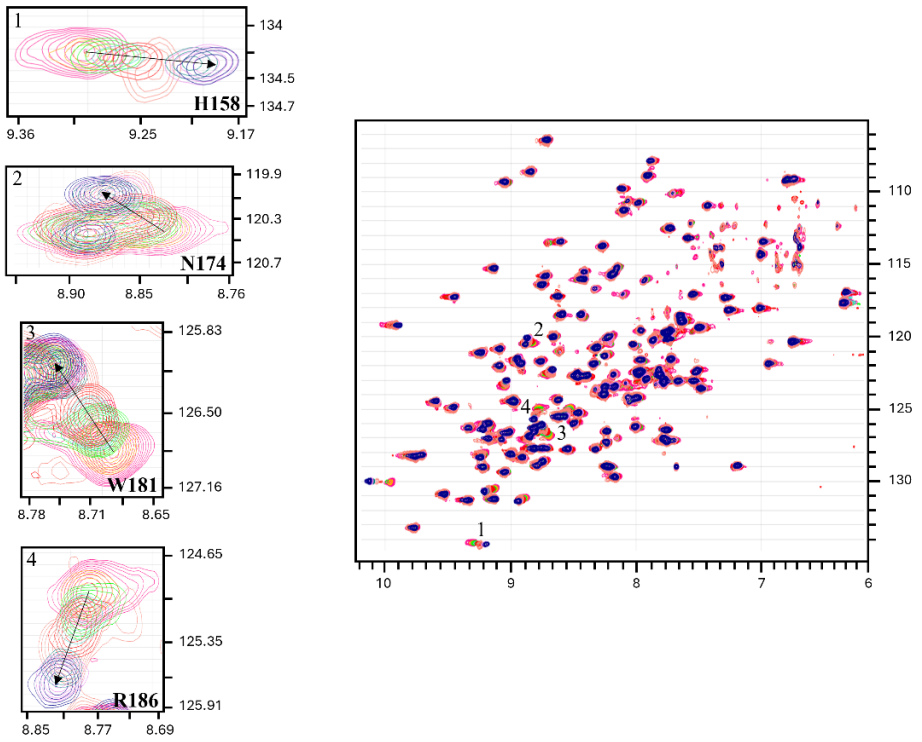


Figure 5.5. CSP analysis of the interaction of Galectin-3 CRD with SeDG. Detailed views of the specific perturbations in key residues are provided in the insets.

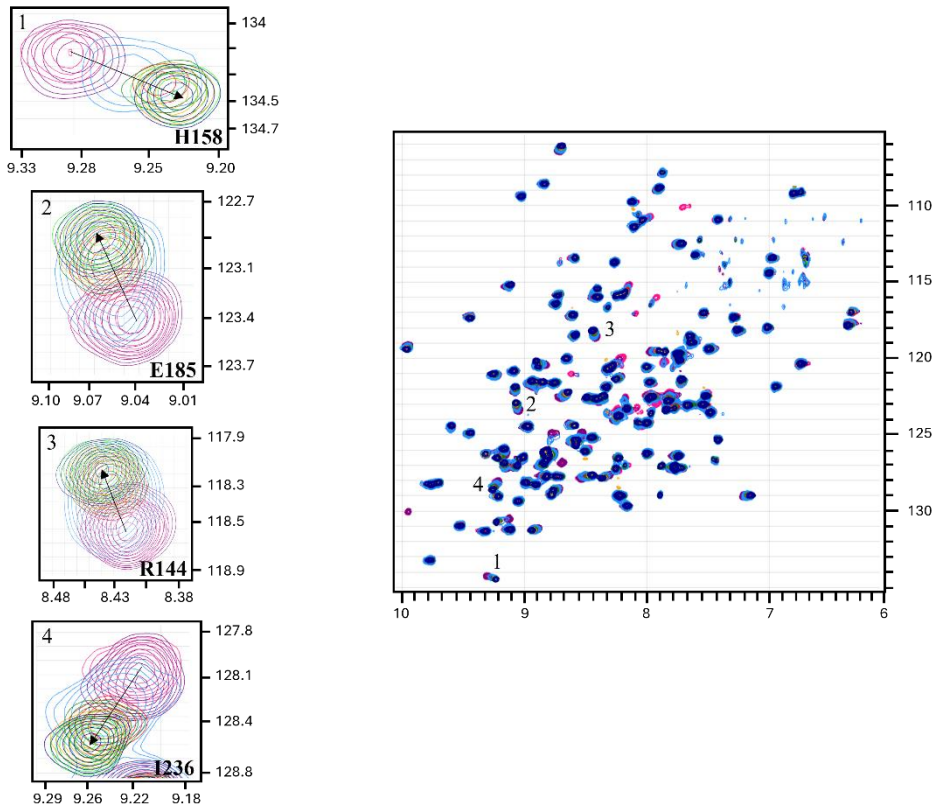


Figure 5.6. CSP analysis of the interaction of Galectin-3 CRD with SeDG-Bn. Detailed views of the specific perturbations in key residues are provided in the insets.

The results showed that SeDG primarily interacts with the canonical binding site, leading to significant CSPs in the β-sheets S4-S5-S6. These perturbations were observed in key residues like H158, R162, N174, and E184, which are known to form hydrogen bonds with the galactose unit (**Figure 5.5**). In addition, a CH-π interaction was identified between the galactose moiety and Trp181.

When the SeDG-Bn CSP was compared with those of SeDG, many similarities were found, suggesting a comparable mode of binding for both ligands. However, crucial differences highlighted the effect of the benzyl group on the protein-ligand

affinity. While the interaction of SeDG was limited to canonical residues, the presence of the benzyl group in SeDG-Bn caused additional perturbations in the S2 and S3 β-sheets, specifically involving residues such as R144, L147, F149 and I236 (**Figure 5.6**). This finding indicated that the benzyl group forms new hydrophobic interactions that extend beyond the traditional binding site.

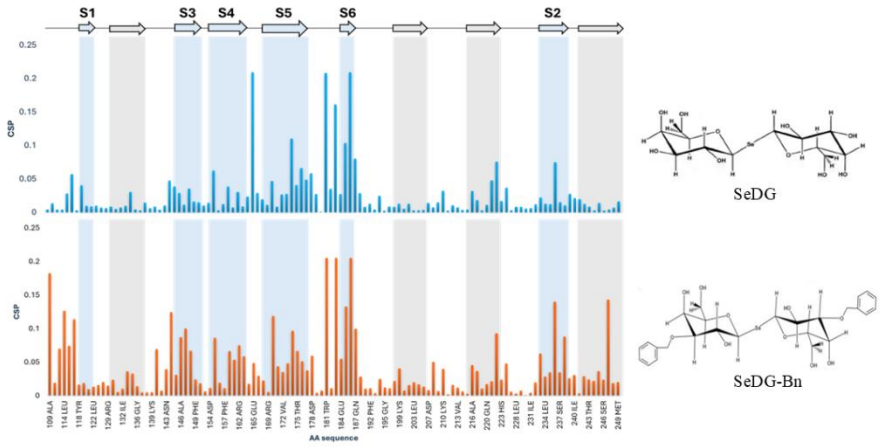


Figure 5.7. TROSY cross-peak CSP (Δδ) plots for Galectin-3 CRD amino acid residues, obtained at saturation with SeDG (20 equivalents) and SeDG-Bn (5 equivalents). The CSP data are plotted against the amino acid sequence of the protein to highlight the specific residues involved in the interaction. *Adapted from [54]*.

Furthermore, another crucial evidence of a stronger binding for SeDG-Bn came from the titration data. Particularly, protein saturation was achieved with only 5 equivalents of SeDG-Bn (**Figure 5.7B**), a significant reduction compared to the 20 equivalents needed for SeDG (**Figure 5.7A**). This result suggested a higher affinity of Gal-3 for the benzyl-substituted compound.

Moreover, the dynamic nature of the binding event was different for the two ligands. In the presence of SeDG-Bn, some cross-peak signals decreased in

intensity, a hallmark of a slow-exchange regime on the NMR timescale. In contrast, SeDG exhibited a fast-exchange regime, as shown by the progressive shifting of cross-peak signals. These results collectively demonstrated that the benzyl group not only enhances the binding affinity but also creates a more stable and "tighter" interaction within the Gal-3 binding pocket.

The CSPs were also used to calculate the equilibrium dissociation constants (K_d) for the interaction between hGal-3 CRD and its two ligands, SeDG and SeDG-Bn (**Figure 5.8A** and **B**, respectively). By fitting the binding curves of the residues with the most significant CSPs, a K_d value of 149 ± 30 µM was determined for the SeDG in complex with the protein. In contrast, the SeDG-Bn complex showed a much lower K_d of 30 ± 12 µM.

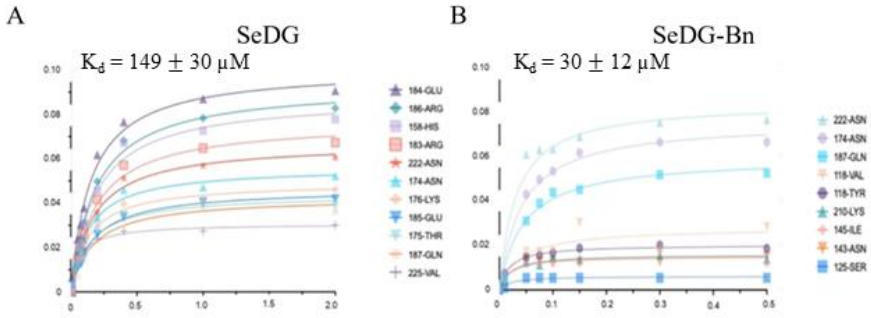


Figure 5.8. The K_d determined by fitting the combined $^{1}H/^{15}N$ chemical shift changes (Δδ) from 2D ^{1}H, ^{15}N-TROSY spectra. The most significant CSPs in the ^{15}N-labeled Gal-3 upon titration with (A) SeDG and (B) SeDG-Bn were used to calculate the respective K_d values. *Adapted from [54]*.

These results are consistent with previous findings [53, 56], which suggested that the addition of a benzyl group at the C3 position of SeDG-Bn significantly enhances its binding affinity compared to SeDG. Therefore, these data confirmed

that the benzyl group is crucial for modulating both the binding specificity and the affinity of the ligand for Gal-3.

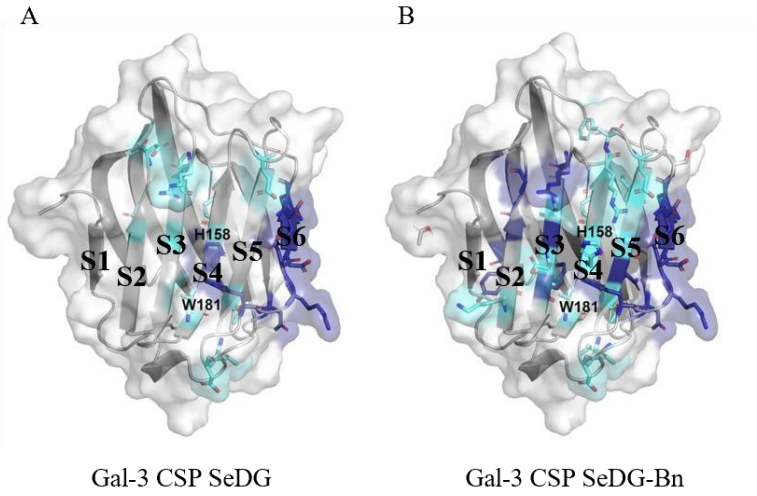


Figure 5.9. Mapping the binding sites of the ligands on Gal-3. NMR CSPs were used to identify the amino acid residues of Gal-3 structure (*PDB: 3ZSJ*) most affected by ligand binding. The protein structure is shown for the SeDG complex (A) and the SeDG-Bn complex (B), with residues exhibiting stronger (blue) and weaker (cyan) CSP effects. *Adapted from [54]*.

The analysis of chemical shift perturbations revealed that the benzyl group plays a key role in the Gal-3 binding process. Specifically, greater perturbations were observed in amino acid residues within the S2 and S3 β-sheets when the protein was bound to SeDG-Bn (**Figure 5.9B**). This effect was attributed to the benzyl moiety engaging with a hydrophobic pocket that includes nonpolar residues like L147, I236, G238, and the hydrophobic residue F149. Additionally, the charged residues R144, D148, and L233 exhibited significant CSPs in the presence of SeDG-Bn. Notably, these same residues within the S2 and S3 β-sheets were largely unaffected by SeDG binding, which lacks the aromatic group (**Figure 5.9A**). This result suggested that the benzyl group binds to a distinct hydrophobic

region, with the benzyl residue maintaining some solvent exposure. The comparable CSPs observed for the S6 β-sheet residues in both SeDG and SeDG-Bn supported this conclusion (**Figure 5.9**).

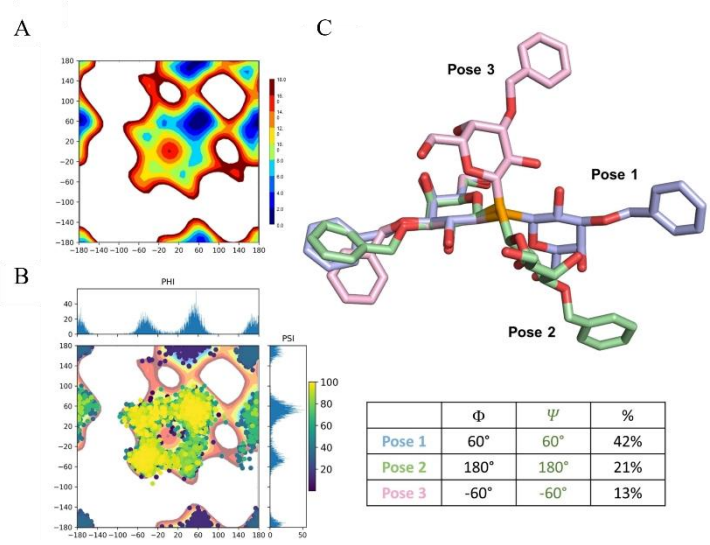
5.2.3 Molecular modeling of the Gal-3/SeDG-Bn complex structure

To investigate the conformational properties of seleno-glycosides, molecular dynamics (MD) simulations were conducted for both the free and Gal-3 bound states [57]. These structural studies were performed by Dr. Ferran Nieto-Fabregat, a postdoctoral researcher in our lab.

Given the unique nature of selenium and the absence of pre-existing parameters for this atom, a crucial parametrization step was first necessary [52]. Specifically, an initial MD simulation focused on the SeDG ligands in its free state to evaluate its stability and possible parametrization errors [52].

The disaccharide backbone of SeDG and SeDG-Bn was performed to identify energetically accessible conformational regions. The resulting adiabatic energy maps, which show the ϕ (H1-C1-Se-C1')/ψ (C1-Se-C1'-H1') glycosidic torsion angles (**Figure 5.10A**), confirmed that the global energy minimum is consistent with the *exo-anomeric* effect. Additionally, the ϕ/ψ glycosidic torsion angles were examined in the free and bound states. In its unbound form, SeDG-Bn showed a high degree of conformational flexibility, existing in various conformations including extended (60°/60°), semi-extended (-60°/-60°) and closed (180°/180°) forms (**Figure 5.10B** and **C**).

Conversely, the bound state exclusively permitted the extended and semi-extended conformations. This observation supports a model where conformer selection occurs upon complex formation (**Figure 5.11**).



	Φ	Ψ	%
Pose 1	60°	60°	42%
Pose 2	180°	180°	21%
Pose 3	-60°	-60°	13%

Figure 5.10. Conformational analysis of SeDG-Bn in free-state using MM and MD simulations. (A) Adiabatic energy maps from molecular mechanics simulations showing the glycosidic torsion angles of SeDG-Bn. (B) Scatter plots of Phi *vs* Psi torsion angles during the MD simulation of free-state SeDG-Bn. Relative histograms depict the most populated energies. (C) A representative three-dimensional model of free-state SeDG-Bn, generated from the 100 ns molecular dynamics simulation. *Adapted from [54].*

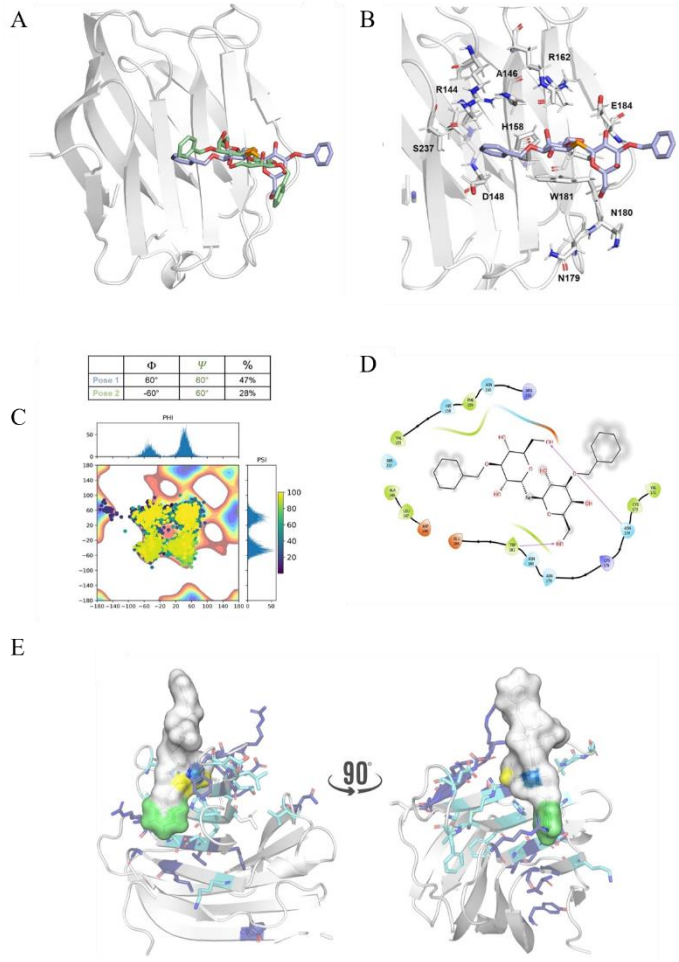


Figure 5.11. A representative three-dimensional model of the complex SeDG-Bn and Gal-3, generated from 100 ns molecular dynamics simulation. (A) The two representative conformations of the SeDG-Bn/Gal-3 complex, with their corresponding ϕ and ψ angle values. (B) A representative three-dimensional model of the SeDG-Bn/Gal-3 complex, derived from the 100 ns MD simulation. The amino acid residues involved in the binding interface are displayed as sticks. (C) Scatter plots of Phi *vs* Psi torsion angles during the MD simulation of SeDG-Bn bound state. Relative histograms depict the most populated energies. (D) A two-dimensional plot providing a schematic overview of the interactions between Gal-3 and SeDG-Bn. (E) The Gal-3/SeDG-Bn complex with Gal-3 colored based on protein-based NMR data and SeDG-Bn ligand colored according to the STD edit code. *Adapted from [54].*

Across all the key binding poses, SeDG-Bn fit into the Gal-3 binding site in a similar manner. In particular, the protein recognized the Bn and Gal parts of the molecule, leaving the Gal' and Bn' groups solvent-exposed due to the linkage's flexibility.

Interestingly, the galactose portions of both the Lac and SeDG ligand were positioned within the Gal-3 canonical binding pocket, a finding consistent with the known binding mode seen in the X-ray structure (*PDB: 3ZSJ*) [58]. This interaction was stabilized by the electropositive face of the galactose ring aligning with the indole ring of the key residue Trp181, which creates a stable CH-π interaction [59]. Specifically, the C-H bonds at positions 3, 4, 5, and 6 of the galactose were located approximately 4 Å from Trp181. Additionally, an internal galactose group created a hydrogen bond with His158.

For SeDG, this binding mode was maintained. The Gal' unit occupied the Gal-3 binding site and established hydrogen bonds with Glu184 and Arg186. However, the addition of the benzyl group in SeDG-Bn altered this interaction by shifting the glycosidic linkage's orientation (**Figure 5.12**). Computational studies revealed that the benzyl group changes the ψ angle from 0° to 60°, displacing Gal' from the binding site and preventing the observed hydrogen bonds.

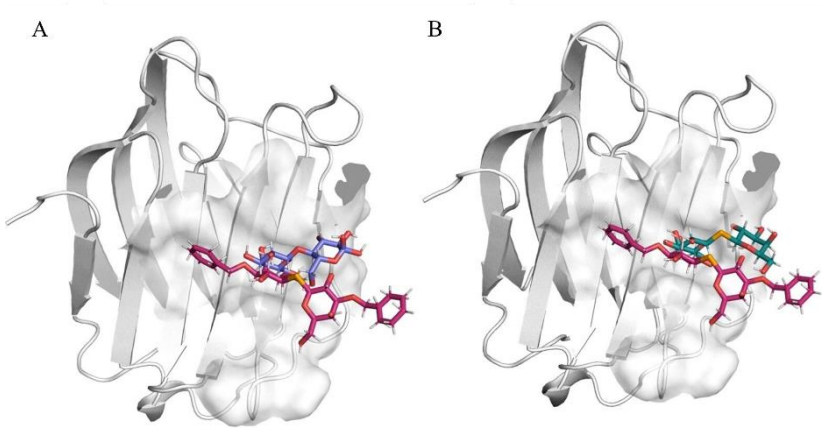


Figure 5.12. Structural superimposition of the complexes formed by Gal-3 with Lac, SeDG, and SeDG-Bn. (A) A full 3D view of the Gal-3 complexes with both Lactose (in purple, *PDB:3ZSJ*) and SeDG-Bn (in pink, from MD simulation). (B) A full 3D view of the Gal-3 complexes with SeDG (in green) and SeDG-Bn (in pink), both obtained from MD simulations. *Adapted from [54].*

Despite this shift, the most stable binding pose of the Gal-3/SeDG-Bn complex retained the key Trp181-galactose CH-π interaction, as shown by cluster analysis of molecular dynamics simulations (**Figure 5.11**, Pose 1). The benzyl group, however, partially displaced both Gal and Gal' compared to the X-ray structure.

Beyond the Trp181 CH-π and van der Waals contacts, other interactions were predicted (**Figure 5.11**). The Gal residue at position 6 formed stable polar interactions with Asn174 and Glu184. The OH at position 4 of Gal formed a hydrogen bond with Arg162. The Bn moiety was stabilized by van der Waals interactions with the S2-S3 β-sheets (Ala146, Asp148, and Ser237). Hydrogen bonds were also predicted between the Gal' residue at position 6 and the peptide backbones of Trp181 and Asn179. Importantly, the Bn' residue was not recognized by Gal-3.

Furthermore, MD simulations were used to study how water molecules stabilize the Gal-3/SeDG-Bn complex (**Figure 5.13**). The results showed that several key water molecules are conserved within the binding pocket, a finding consistent with earlier research [58, 60]. These water molecules act as intermediaries, forming crucial hydrogen bonds between the ligand and specific amino acid residues of Gal-3 like Arg162, Glu184, and Asn174. The simulations also revealed that the benzyl group of SeDG-Bn plays a significant role in displacing some of the temporary water molecules near β-sheets S2 and S3. This displacement enhances the hydrophobic interactions between the ligand and Gal-3 residues such as Leu147 and Phe149.

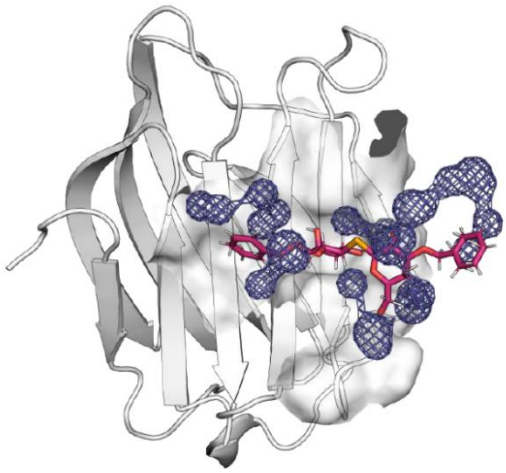


Figure 5.13. The water density occupancy (blue mesh) around the SeDG-Bn ligand (pink) and the binding pocket residues. The data was calculated from the MD trajectory within a 3 Å distance. The Gal-3 protein is shown in white, highlighting how the hydration environment helps stabilize the ligand. *Reproduced from [54].*

These findings highlighted the dual importance of water molecules in this system: they not only help maintain the overall structure of the protein but also actively influence how the ligand binds.

5.3 Discussion and Conclusions

Gal-3, a multifunctional protein, represents a promising therapeutic target due to its involvement in various biological processes and its overexpression in many diseases, including cancer [35, 61]. The CRD of this protein is particularly important because it selectively binds to glycoconjugates containing β-galactose, making it an ideal focus for the development of selective inhibitors [39, 62-63].

Many of the galectin inhibitors have been developed from a disaccharide scaffold, with TDG being a popular choice, as seen with TD139.

In this area, scientific literature also includes binding studies of two selenium-containing Gal-3 inhibitors: SeDG and a related compound. This study analyzes the interaction between SeDG-Bn (a seleno-di-galactoside, modified with a benzyl group) and Gal-3 CRD at the molecular level by NMR and computational studies. In particular, NMR-obtained molecular binding information was complemented with molecular dynamics simulations to obtain a comprehensive 3D model of the SeDG-Bn/Gal-3 complex. This analysis confirmed that a stronger binding affinity for Gal-3 is achieved by adding benzyl groups to the C3 positions of SeDG [53].

NMR experiments provided detailed insights into molecular interactions. STD NMR revealed that the aromatic protons of the benzyl group are selectively recognized by the protein, which contributes to the improved binding affinity [64]. In addition to STD NMR, the CPMG experiment provided further qualitative

evidence of ligand binding. In the presence of Gal-3, SeDG-Bn displayed a pronounced decrease in T_2 values, consistent with partial immobilization of the ligand upon interaction with the protein. This behavior aligns with the expected increase in transverse relaxation experienced by small molecules when engaging in weak-to-moderate affinity binding. Although the CPMG data acquired under these experimental conditions do not allow a quantitative determination of affinity, they corroborate the interaction detected by STD NMR and support the involvement of SeDG-Bn in the Gal-3 binding pocket.

Furthermore, CSP analysis also identified key stabilizing contacts, including hydrogen bonds formed with Arg162, Asn174, and Glu184, as well as CH-π interactions with residue Trp181 [65]. In addition, the presence of the benzyl group in SeDG-Bn caused hydrophobic interactions with Leu147 and Phe149, which extended beyond the canonical binding site of the protein.

Another important point is given by MD simulations which further clarified the binding mechanism. The simulations showed that SeDG-Bn, which is flexible in its free state, adopts a process of conformational selection, becoming more rigid upon binding to Gal-3. This conformational change increases both binding specificity and stability of the protein-ligand complex.

An analysis of the hydration environment of the binding pocket identified key conserved water molecules that stabilize the complex by mediating hydrogen bonds with residues like Arg162 and Glu184. The benzyl group also actively influenced this hydration network by displacing transient water molecules, thereby promoting stronger hydrophobic interactions with residues in β-sheets S2 and S3.

These findings align with recent studies on similar compounds, such as arakyl-thiodigalactosides [66], which also showed that aromatic groups enhance binding

by forming additional hydrophobic and cation-π interactions with key residues like Arg162 and Arg186. The ability of these molecules to stabilize the complex through $\pi - \pi$ stacking interactions with Trp181 and by modulating the local hydration environment further reinforces the observations made for SeDG-Bn.

In conclusion, these results highlight the potential of selenium-containing compounds like SeDG-Bn as a basis for developing new Gal-3 inhibitors, which could be further optimized for treating diseases like cancer and fibrosis.

Section III - References

1. P.R. Crocker, J.C. Paulson, A. Varki. "Siglecs and their roles in the immune system." *Nature Reviews Immunology*, 7 (2007) 255-266.
2. M.S. MacAuley, P.R. Crocker, J.C. Paulson. "Siglec-mediated regulation of immune cell function in disease." *Nat Rev Immunol*, 14 (2014) 653-666.
3. A. Varki, T. Angata. "Siglecs-the major subfamily of I-type lectins." *Glycobiology*, 16 (2006) 1R-27R.
4. H. Floyd, J. Ni, A. L. Cornish, Z. Zeng, D. Liu, K. C. Carter, J. Steel, P. R. Crocker. "Siglec-8: A novel eosinophil-specific member of the immunoglobulin superfamily." *Journal of Biological Chemistry*, 275 (2000) 861-866.
5. T. Kiwamoto, N. Kawasaki, J.C. Paulson, B.S. Bochner. "Siglec-8 as a drugable target to treat eosinophil and mast cell-associated conditions." *Pharmacol Ther*, 135 (2012) 327-336.
6. D.J. Carroll, Y. Cao, B.S. Bochner, J.A. O'Sullivan. "Siglec-8 Signals Through a Non-Canonical Pathway to Cause Human Eosinophil Death In Vitro." *Front Immunol*, 12 (2021) 737988.
7. J.M. Pröpster, F. Yang, S. Rabbani, B. Ernst, F.H.T. Allain, M. Schubert. "Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8." *Proc Natl Acad Sci U S A*, 113 (2016) E4170-E4179.
8. J.V. Ravetch, L.L. Lanier. "Immune inhibitory receptors." *Science (1979)*, 290 (2000) 84-89.
9. D.J. Carroll, J.A. O'Sullivan, D.B. Nix, Y. Cao, M. Tiemeyer, B.S. Bochner. "Siglec-8 is an activating receptor mediating β2 integrin-

dependent function in human eosinophils." *J Allergy Clin Immunol*, 141 (2017) 2196.

10. P.R. Crocker, P. Redelinghuys. "Siglecs as positive and negative regulators of the immune system." *Biochem Soc Trans*, 36 (2008) 1467-1471.
11. J.C. Paulson, M.S. MacAuley, N. Kawasaki. "Siglecs as sensors of self in innate and adaptive immune responses." *Ann N Y Acad Sci*, 1253 (2012) 37-48.
12. M.P. Lenza, U. Atxabal, I. Oyenarte, J. Jiménez-Barbero, J. Ereño-Orbea. "Current Status on Therapeutic Molecules Targeting Siglec Receptors." *Cells*, 9 (2020) 2691.
13. B. S. Bochner, R. A. Alvarez, P. Mehta, N. V. Bovin, O. Blixt, J. R. White, R. L. Schnaar. "Glycan Array Screening Reveals a Candidate Ligand for Siglec-8." *Journal of Biological Chemistry*, 280 (2005) 4307-4312.
14. H. Tateno, P.R. Crocker, J.C. Paulson. "Mouse Siglec-F and human Siglec-8 are functionally convergent paralogs that are selectively expressed on eosinophils and recognize 6′-sulfo-sialyl Lewis X as a preferred glycan ligand." *Glycobiology*, 15 (2005) 1125-1135.
15. B. S. Kroezen, G. Conti, B. Girardi, J. Cramer, X. Jiang, S. Rabbani, J. Müller, M. Kokot, E. Luisoni, D. Ricklin, O. Schwardt, B. Ernst. "A Potent Mimetic of the Siglec-8 Ligand 6'-Sulfo-Sialyl Lewisx." *ChemMedChem*, 15 (2020) 1680-1680.
16. M. P. Lenza, U. Atxabal, C. Nycholat, I. Oyenarte, A. Franconetti, J. I. Quintana, S. Delgado, R. Núñez-Franco, C. T. Garnica Marroquín, H. Coelho, L. Unione, G. Jiménez-Oses, F. Marcelo, M. Schubert, J. C. Paulson, J. Jiménez-Barbero, J. Ereño-Orbea. "Structures of the Inhibitory Receptor Siglec-8 in Complex with a High-Affinity Sialoside Analogue and a Therapeutic Antibody." *JACS Au*, 3 (2023) 204-215.

17. G. Murugesan, B. Weigle, P.R. Crocker. "Siglec and anti-Siglec therapies." *Curr Opin Chem Biol*, 62 (2021) 34-42.
18. P. Ilmarinen, H. Kankaanranta. "Eosinophil Apoptosis as a Therapeutic Target in Allergic Asthma." *Basic Clin Pharmacol Toxicol*, 114 (2014) 109-117.
19. P.V. Murphy, A. Dhara, L.S. Fitzgerald, E. Hever, S. Konda, K. Mandal. "Small lectin ligands as a basis for applications in glycoscience and glycomedicine." *Chem Soc Rev*, 53 (2024) 9428-9445.
20. J. Schanin, S. Gebremeskel, W. Korver, R. Falahati, M. Butuci, T. J. Haw, P. M. Nair, G. Liu, N. G. Hansbro, P. M. Hansbro, E. Evensen, E. C. Brock, A. Xu, A. Wong, J. Leung, C. Bebbington, N. Tomasevic, B. A. Youngblood. "A monoclonal antibody to Siglec-8 suppresses non-allergic airway inflammation and inhibits IgE-independent mast cell activation." *Mucosal Immunol*, 14 (2021) 366-376.
21. E.M. Rapoport, G.V. Pazynina, M.A. Sablina, P.R. Crocker, N.V. Bovin. "Probing sialic acid binding Ig-like lectins (siglecs) with sulfated oligosaccharides." *Biochemistry (Moscow)*, 71 (2006) 496-504.
22. S. Duan, B. M. Arlian, C. M. Nycholat, Y. Wei, H. Tateno, S. A. Smith, M. S. Macauley, Z. Zhu, B. S. Bochner, J. C. Paulson. "Nanoparticles Displaying Allergen and Siglec-8 Ligands Suppress IgE-FcεRI–Mediated Anaphylaxis and Desensitize Mast Cells to Subsequent Antigen Challenge." *The Journal of Immunology*, 206 (2021) 2290-2300.
23. S.A. Hudson, N.V. Bovin, R.L. Schnaar, P.R. Crocker, B.S. Bochner. "Eosinophil-Selective Binding and Proapoptotic Effect in Vitro of a Synthetic Siglec-8 Ligand, Polymeric 6′-Sulfated Sialyl Lewis X." *Journal of Pharmacology and Experimental Therapeutics*, 330 (2009) 608-612.

24. G. Conti, A. Bärenwaldt, S. Rabbani, T. Mühlethaler, M. Sarcevic, X. Jiang, O. Schwardt, D. Ricklin, R. J. Pieters, H. Läubli, B. Ernst. "Tetra- and Hexavalent Siglec-8 Ligands Modulate Immune Cell Activation." *Angewandte Chemie International Edition*, 62 (2023) e202314280.
25. J.M. Pröpster, F. Yang, B. Ernst, F.H. Allain, M. Schubert. "Functional Siglec lectin domains from soluble expression in the cytoplasm of Escherichia coli." *Protein Expr Purif.*, 109 (2015) 14-22.
26. N. Dimasi, L. Moretta, R. Biassoni, R.A. Mariuzza. "Expression, crystallization and preliminary crystallographic analysis of the extracellular IgV-like domain of the human natural killer cell inhibitory receptor p75/AIRM1." *Acta Crystallogr. D Biol. Crystallogr.*, 59 (2003) 1856-1858.
27. M.A. Zhuravleva, K. Trandem, P.D. Sun. "Structural implications of Siglec-5-mediated sialoglycan recognition." *J. Mol. Biol.*, 375 (2008) 437-447.
28. A.W. Barb, X. Wang, J.H. Prestegard. "Refolded recombinant Siglec-5 for NMR investigation of complex carbohydrate binding." *Protein Expr. Purif.*, 88 (2013) 183-189.
29. T.J. Lukianova, V. Kinzhybalo, A. Pietraszko. "Crystal structure of tris-(piperidinium) hydrogen sulfate sulfate." *Acta Crystallogr E Crystallogr Commun.*, 71 (2015) 1444-1446.
30. M.P. Williamson. "Using chemical shift perturbation to characterise ligand binding." *Prog Nucl Magn Reson Spectrosc*, 73 (2013) 1-16.
31. K. Pervushin, R. Riek, G. Wider, K. Wüthrich. "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." *Proc Natl Acad Sci U S A*, 94 (1997) 12366-12371.

32. Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold, B. Brutscher. "BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins." *J Biomol NMR*, 55 (2013) 311-321.
33. A. Macchiarulo, R. Pellicciari. "Exploring the other side of biologically relevant chemical space: Insights into carboxylic, sulfonic and phosphonic acid bioisosteric relationships." *J Mol Graph Model.*, 26 (2007) 728-739.
34. L. Johannes, R. Jacob, H. Leffler. "Galectins at a glance." *J Cell Sci.*, 131 (2018) jcs208884.
35. M.F. Troncoso, M.T. Elola, A.G. Blidner, L. Sarrias, M.V. Espelt, G.A. Rabinovich. "The universe of galectin-binding partners and their functions in health and disease." *J Biol Chem.*, 299 (2023) 105400.
36. R.P.M. Dings, M.C. Miller, R.J. Griffin, K.H. Mayo. "Galectins as molecular targets for therapeutic intervention." *International Journal of Molecular Sciences*, 19 (2018) 905.
37. A. Hara, M. Niwa, K. Noguchi, T. Kanayama, A. Niwa, M. Matsuo, Y. Hatano, H. Tomita. "Galectin-3 as a next-generation biomarker for detecting early stage of various diseases." *Biomolecules*, 10 (2020) 389.
38. L. Song, J.W. Tang, L. Owusu, M.Z. Sun, J. Wu, J. Zhang. "Galectin-3 in cancer." *Clin Chim Acta.*, 431 (2014) 185–191.
39. H. Blanchard, X. Yu, P.M. Collins, K. Bum-Erdene. "Galectin-3 inhibitors: a patent review (2008-present)." *Expert Opin Ther Pat.*, 24 (2014) 1053-1065.
40. P.G. Traber, H. Chou, E. Zomer, F. Hong, A. Klyosov, M.-I. Fiel, S.L. Friedman. "Regression of Fibrosis and Reversal of Cirrhosis in Rats by Galectin Inhibitors in Thioacetamide-Induced Liver Disease." *PLOS ONE*, 8 (2013) e75361.

41. P.G. Traber, E. Zomer. "Therapy of Experimental NASH and Fibrosis with Galectin Inhibitors." *PLOS ONE,* 8 (2013) e83481.
42. T.-J. Hsieh, H.-Y. Lin, Z. Tu, T.-C. Lin, S.-C. Wu, Y.-Y. Tseng, F.-T. Liu, S.-T. D. Hsu, C.-H. Lin. "Dual thio-digalactoside-binding modes of human galectins as the structural basis for the design of potent and selective inhibitors." *Scientific Reports,* 6 (2016) 1-9.
43. Y. Shen, W. Zhang, H. Chang, Z. Li, C. Lin, G. Zhang, L. Mao, N. Liu, H. Lu. "Galectin-3 modulates microglial activation and neuroinflammation in early brain injury after subarachnoid hemorrhage." *Exp Neurol.,* 377 (2024) 114777.
44. J. Xia, Y. Wang, B.R. Qi. "In vitro and in vivo effects of Galectin-3 inhibitor TD139 on inflammation and ERK/JNK/p38 pathway in gestational diabetes mellitus." *Kaohsiung J Med Sci.,* 40 (2024) 916-925.
45. Y. Guo, R. Shen, L. Yu, X. Zheng, R. Cui, Y. Song, D. Wang. "Roles of galectin-3 in the tumor microenvironment and tumor metabolism (Review)." *Oncol Rep.,* 44 (2020) 1799-1809.
46. N. Hirani, A.C. MacKinnon, L. Nicol, P. Ford, H. Schambye, A. Pedersen, U.J. Nilsson, H. Leffler, T. Sethi, S. Tantawi, L. Gravelle, R.J. Slack, R. Mills, U. Karmakar, D. Humphries, F. Zetterberg, L. Keeling, L. Paul, P.L. Molyneaux, F. Li, W. Funston, I.A. Forrest, A.J. Simpson, M.A. Gibbons, T.M. Maher. "Target inhibition of galectin-3 by inhaled TD139 in patients with idiopathic pulmonary fibrosis." *Eur Respir J.,* 57 (2021) 2002559.
47. M.G. Jones, M. Kolb. "Repeat bronchoalveolar lavage in idiopathic pulmonary fibrosis: proceed with caution?" *Eur Respir J.,* 57 (2021) 2100691.
48. A. Iadonisi, S. Traboni, D. Capasso, E. Bedini, S. Cuomo, S. Di Gaetano, G. Vessella. "Switchable synthesis of glycosyl selenides or diselenides

with direct use of selenium as the selenating agent." *Org Chem Front.*, 8 (2021) 1823-1829.

49. A. Razaghi, M. Poorebrahim, D. Sarhan, M. Björnstedt. "Selenium stimulates the antitumour immunity: Insights to future research." *Eur J Cancer.*, 155 (2021) 256-267.
50. L.H. Duntas. "Selenium and inflammation: underlying anti-inflammatory mechanisms." *Horm Metab Res.*, 41 (2009) 443-447.
51. R. Zhang, X. Xie, J. Liu, R. Pan, Y. Huang, Y. Du. "A novel selenoglycoside compound GlcSeCys alleviates diets-induced obesity and metabolic dysfunctions with the modulation of Galectin-1 and selenoproteins." *Life Sci.*, 359 (2024) 123259.
52. L. Pirone, F. Nieto-Fabregat, S. Di Gaetano, D. Capasso, R. Russo, S. Traboni, A. Molinaro, A. Iadonisi, M. Saviano, R. Marchetti, A. Silipo, E. Pedone. "Exploring the Molecular Interactions of Symmetrical and Unsymmetrical Selenoglycosides with Human Galectin-1 and Galectin-3." *Int J Mol Sci.*, 23 (2022) 8273.
53. S. Di Gaetano, L. Pirone, I. Galdadas, S. Traboni, A. Iadonisi, E. Pedone, M. Saviano, F.L. Gervasio, D. Capasso. "Design, Synthesis, and Anticancer Activity of a Selenium-Containing Galectin-3 and Galectin-9N Inhibitor." *Int J Mol Sci.*, 23 (2022).
54. M.P. Lenza, C. Di Carluccio, F. Nieto-Fabregat, L. Pirone, R. Russo, S. Di Gaetano, D. Capasso, A. Stornaiuolo, A. Iadonisi, M. Saviano, R. Marchetti, E. Pedone, A. Silipo. "Structural Insights into Galectin-3 Recognition of a Selenoglycomimetic." *ChemMedChem.*, 20 (2025) e202401005.
55. K. Umemoto, H. Leffler. "Assignment of 1H, 15N and 13C resonances of the carbohydrate recognition domain of human galectin-3." *J. Biomol. NMR*, 20 (2001) 91-92.

56. M. Raics, Á.K. Balogh, C. Kishor, I. Timári, F.J. Medrano, A. Romero, R.M. Go, H. Blanchard, L. Szilágyi, K.E. Kövér. "Investigation of the Molecular Details of the Interactions of Selenoglycosides and Human Galectin-3." *International Journal of Molecular Sciences*, 23 (2022) 2494.
57. F. Nieto-Fabregat, M.P. Lenza, A. Marseglia, C. Di Carluccio, A. Molinaro, A. Silipo, R. Marchetti. "Computational toolbox for the analysis of protein-glycan interactions." *Beilstein J. Org. Chem.*, 20 (2024) 2084-2107.
58. K. Saraboji, M. Håkansson, S. Genheden, C. Diehl, J. Qvist, U. Weininger, U.J. Nilsson, H. Leffler, U. Ryde, M. Akke, D.T. Logan. "The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics." *Biochemistry*, 51 (2012) 296-306.
59. Y.-C. Chan, H.-Y. Lin, Z. Tu, Y.-H. Kuo, S.-T. D. Hsu, C.-H. Lin. "Dissecting the Structure-Activity Relationship of Galectin-Ligand Interactions." *Int J Mol Sci.*, 19 (2018) 392.
60. J. Su, T. Zhang, P. Wang, F. Liu, G. Tai, Y. Zhou. "The water network in galectin-3 ligand binding site guides inhibitor design." *Acta Biochimica et Biophysica Sinica*, 47 (2015) 192-198.
61. L.C. Soares, O. Al-Dalahmah, J. Hillis, C.C. Young, I. Asbed, M. Sakaguchi, E. O'Neill, F.G. Szele. "Novel Galectin-3 Roles in Neurogenesis, Inflammation and Neurological Diseases." *Cells*, 10 (2021) 3047.
62. K. Bum-Erdene, I.A. Gagarinov, P.M. Collins, M. Winger, A.G. Pearson, J.C. Wilson, H. Leffler, U.J. Nilsson, I.D. Grice, H. Blanchard. "Investigation into the feasibility of thiodigataloside as a novel scaffold for galectin-3-specific inhibitors." *Chembiochem*, 14 (2013) 1331-1342.

63. J. Stegmayr, F. Zetterberg, M.C. Carlsson, X. Huang, G. Sharma, B. Kahl-Knutson, H. Schambye, U.J. Nilsson, S. Oredsson, H. Leffler. "Extracellular and intracellular small-molecule galectin-3 inhibitors." *Sci Rep*, 9 (2019) 1-12.
64. F. Mangiavacchi, I.F. Coelho Dias, I. Di Lorenzo, P. Grzes, M. Palomba, O. Rosati, L. Bagnoli, F. Marini, C. Santi, E.J. Lenardao, L. Sancineto. "Sweet Selenium: Synthesis and Properties of Selenium-Containing Sugars and Derivatives." *Pharmaceuticals*, 13 (2020) 211.
65. R.C. Diehl, R.S. Chorghade, A.M. Keys, M.M. Alam, S.A. Early, A.E. Dugan, M. Krupkin, K. Ribbeck, H.J. Kulik, L.L. Kiessling. "CH-π Interactions Are Required for Human Galectin-3 Function." *JACS Au*, 4 (2024) 3028-3037.
66. F. Hőgye, L.B. Farkas, Á.K. Balogh, L. Szilágyi, S. Alnukari, I. Bajza, A. Borbás, K. Fehér, T.Z. Illyés, I. Timári. "Saturation Transfer Difference NMR and Molecular Docking Interaction Study of Aralkyl-Thiodigalactosides as Potential Inhibitors of the Human-Galectin-3 Protein." *Int J Mol Sci*, 25 (2024) 1742.

SECTION IV – KEY CONCLUSIONS

The role of glycans as recognition markers for glycan-binding proteins (GBPs) represents a central and rapidly advancing field in current scientific research. Understanding the molecular determinants governing the interactions between sialic acids and Siglecs, and between β-galactosides and Galectins, is not only essential for clarifying fundamental *self*/*non-self* recognition mechanisms but also for informing the design of next-generation therapeutics targeting inflammatory, autoimmune, and neoplastic diseases.

Within this framework, the present PhD thesis provides molecular and structural insights into lectin-glycan interactions. The results contribute to the broader understanding of how GBPs decode glycan information, ultimately defining new strategies for the rational design of glycomimetic ligands.

For Siglec-8, the development and characterization of novel sialylated triazoles reveal that modulating the electronic environment and hydrogen-bonding network around the C-9 position of sialic acid is a generalizable control point in Siglec recognition. The observation that sulfonate and, in particular, naphthylsulfonate substituents markedly enhance affinity compared with canonical sulfates or carboxylates aligns with and expands current structure-activity models of Siglec-glycan recognition. These findings suggest a broadly applicable design principle for engineering monomeric Siglec inhibitors, potentially extendable to family members such as Siglec-7 and Siglec-9, which share conserved features in their sialic acid binding pockets. Overall, this work establishes an accessible synthetic route toward high-affinity Siglec-8 ligands based on sialyl azide intermediates while contributing to the general rules that govern selective Siglec binding.

The study of the interaction between Galectin-3 and selenium-containing disaccharides similarly advances the current understanding of β-galactoside

recognition. The incorporation of a selenium bridge and a C-3 benzyl substituent not only yields exceptional selectivity for Gal-3 but also highlights the plasticity of the binding site. In addition, the results underscore the importance of hydrophobic contacts, alongside the canonical polar interactions, as key contributors to affinity and selectivity. More broadly, the demonstration that selective chemical perturbation and heteroatom introduction within glycan scaffolds can tune galectin recognition provides a conceptual framework that may extend to other GBP families that bind β-galactosides.

These achievements were made possible through an integrated methodological strategy. The combined use of advanced NMR spectroscopy (from both ligand- and protein-based perspectives), biophysical methods, and computational approaches allowed for a detailed understanding of the molecular interaction process. The definition of epitope mapping, bioactive conformations, and the catalog of non-covalent interactions stabilizing the complexes (hydrogen bonds, π–π stacking, electrostatic forces, van der Waals interactions) contributes to a more general understanding of how glycans engage lectins at the molecular level.

In summary, this thesis not only identifies promising glycomimetic candidates for Siglec-8 and Galectin-3 but also provides conceptual and structural foundations for glycan medicinal chemistry. The results demonstrate that rational chemical modification of glycan motifs can overcome the low affinity and broad specificity of native glycan-lectin interactions. By integrating experimental and computational evidence, this work contributes to defining generalizable design principles for developing targeted therapeutic agents aimed at restoring immune homeostasis and treating complex diseases, such as chronic inflammation, fibrosis, and cancer, in which Siglecs and Galectins serve as critical molecular checkpoints.