

**University of Naples “Federico II”
Polytechnic and Basic Sciences School**

Department of Chemical Sciences



Ph.D. School in Chemical Sciences – Cycle XXXIV

**NMR and computational studies of the molecular
recognition of eukaryotic glycans by receptor proteins**

Cristina Di Carluccio

Supervisors:

Prof. Alba Silipo

Prof. Antonio Molinaro

Examiner:

Prof. Angela Lombardi

Project dedicated to the memory of Giulio Regeni

Progetto dedicato alla memoria di Giulio Regeni

Ogni anno una delle borse di dottorato in Scienze Chimiche viene dedicata alla memoria di Giulio Regeni, esempio di coraggio, perseveranza e sacrificio, un uomo violato del diritto di pensiero ed un dottorando privato della libertà della sua ricerca.

È un onore per me portare avanti il ricordo di Giulio, adesso più che mai a sei anni dalla sua scomparsa.

È per questo, dunque, che dedico la mia tesi di dottorato a chi avrebbe dovuto condividere la gioia di vedere il proprio figlio e fratello diventare Dottore: la sua famiglia.

Ecco il suo modo di “fare cose”!

Abstract

All living cells are covered by a layer of glycans at the interface between the environment and the cell membrane, capable of mediating cellular behavior, including critical mechanisms in immunoregulation and pathological processes. The molecular recognition of glycoconjugates from several proteins triggers a plethora of biological functions, especially in the infection process, immune response, and inflammation. Within this frame, interactions between glycans and their binding partners at molecular level have been studied, using a multidisciplinary approach of advanced NMR techniques, including ligand- and protein-based approaches, in combination with biophysical and computational methodologies, such as docking and molecular dynamic simulations.

Siglecs (Sialic acid-binding immunoglobulin-type lectins) exploit a major application in the immune system regulation, recognizing glycans containing sialic acid. Indeed, in their cytoplasmic region, Siglecs contain one or multiple tyrosine-based signaling motifs that trigger cellular signaling, inhibiting the immune cell activation. In this thesis, the molecular binding of different inhibitory Siglecs, in particular Siglec-2 and -7, containing cytosolic immunoreceptor tyrosine-based inhibition motifs (ITIMs), have been investigated with several glycoconjugates. Although the inhibition of immune system plays a fundamental role in some aberrant events, such as the over-reaction of response against self-molecules that often leads to produce autoimmune diseases, it is worth knowing that many pathogens have evolved the ability to cover their surfaces of sialic acids, subverting the immune system and dampening the host immune recognition. Thus, Siglecs have been studied as attractive targets for the design of therapeutic agents, such as antibodies or glycomimetics, for the treatment of inflammatory, autoimmune, and infectious diseases. In the case of Siglec-2, or CD22, the binding mode with complex-type *N*-glycans has been assessed, showing the possibility to form CD22 homo-oligomers on the B-cell surface, favoring the *cis* interactions on the same cell. As for Siglec-7, mainly located on NK cells, novel structural insights have been provided on its binding to sialylated lipopolysaccharides

on different strains of the oncogenic pathogen *F. nucleatum*, with the aim to develop therapies for the modulation of both Siglec-7 activity and host-pathogen binding.

On the other hand, bacterial adhesins, also implicated in the biology of infection, as in the bacterial pathogenesis, have been studied in interaction with their cognate ligands. In particular, Siglec-like adhesins, similar to Siglecs in the V-set *N*-terminal domain of sialoglycan recognition, are serine-rich repeat glycoproteins involved in the pathogenesis of infective endocarditis. In this context, the binding site of Siglec-like adhesins expressed on different strains of *Streptococci* has been investigated in interaction with a variety of sialylated *N*- and *O*-glycans.

Partially related systems have been also investigated during the PhD, involving the study of the interactions between monoclonal antibodies (mAb) against bacterial glycoconjugates (and mimetics).

Therefore, the molecular details of different glycans recognized by mammalian and bacterial proteins, as well as monoclonal antibodies, that play roles in health and disease, or host-pathogen interactions have been unveiled to provide a tool for the design of glycomimetics for therapeutic targets of human diseases.

TABLE OF CONTENTS

I. INTRODUCTION **1**

1.1 THE RELEVANCE OF GLYCANS IN BIOLOGICAL FUNCTIONS	1
1.1.1 THE STRUCTURE OF GLYCOCONJUGATES	2
1.1.2 SIALIC ACID	5
1.2 GLYCOCONJUGATES ON BACTERIAL CELL WALL	7
1.2.1 LIPOPOLYSACCHARIDES (LPSS)	8
1.2.2 TEICHOIC ACID GLYCOPOLYMERS	9
1.3 STRUCTURE AND FUNCTION OF ANTIBODIES	13
1.3.1 CLASSES OF ANTIBODIES	15
1.3.2 MONOCLONAL ANTIBODIES	18
1.4 I-TYPE LECTINS: SIGLECS	19
1.4.1 STRUCTURAL FEATURES OF SIGLECS	20
1.4.2 THE SIGLECS-SIALOGLYCANS INTERACTION	23
1.4.3 SIGLEC-2	24
1.4.4 SIGLEC-7	25
1.4.5 SIGLEC-10	26
1.5 BACTERIAL ADHESINS	27
1.5.1 SIGLEC-LIKE ADHESINS	30
1.6 OBJECTIVES	33

II. UNVEILING THE PROTEIN-LIGAND MOLECULAR BINDING **35**

2.1 NUCLEAR MAGNETIC RESONANCE NMR SPECTROSCOPY	35
2.1.1 LIGAND-BASED NMR APPROACHES	37
2.1.2 TRANSFERRED-NOESY	39
2.1.3 SATURATION TRANSFER DIFFERENCE	41
2.1.4 WATERLOGSY	44
2.1.5 OTHER NMR TECHNIQUES	45
2.2 PROTEIN-BASED NMR APPROACHES	48
2.2.1 CHEMICAL SHIFT PERTURBATION	50
2.3 COMPUTATIONAL METHODS	51
2.3.1 DOCKING	52

2.3.2 MOLECULAR MECHANICS AND DYNAMICS SIMULATION	54
2.3.3 CORCEMA-ST PROGRAM	58

III. UNVEILING THE MURINE AND HUMAN CD22 RECOGNITION TOWARDS SIALOGLYCANS **61**

3.1 INTRODUCTION	61
3.2 INSIGHTS INTO THE AFFINITY OF THE INTERACTION BETWEEN H-CD22 AND ACETYLATED <i>N</i>-GLYCANS	62
3.2.1 BIOPHYSICAL TECHNIQUES FOR DETECTING THE BINDING BETWEEN H-CD22 AND <i>N</i>-GLYCANS	62
3.2.2 MOLECULAR BINDING BETWEEN H-CD22 AND ACETYLATED 6'SLN	65
3.2.3 MOLECULAR BINDING BETWEEN H-CD22 AND COMPLEX-TYPE <i>N</i>-GLYCAN	74
3.3 COMPARISON OF HUMAN AND MURINE CD22 INTERACTIONS WITH <i>N</i>-GLYCANS	80
3.3.1 DETERMINATION OF THE BINDING AFFINITIES OF H- AND M-CD22 TO ACETYLATED AND GLYCOLYLATED 6'SLN	80
3.3.2 COMPARISON OF HUMAN AND MURINE CD22 INTERACTIONS WITH ACETYLATED <i>N</i>-GLYCANS BY NMR	81
3.3.3 COMPARISON OF HUMAN AND MURINE CD22 INTERACTIONS WITH GLYCOLYLATED <i>N</i>-GLYCANS	82
3.3.4 COMPARISON OF ACETYLATED AND GLYCOLYLATED 6'SLN INTO THE H-CD22 AND M-CD22 BINDING SITES	90
3.4 DISCUSSION	91

V. MOLECULAR DETAILS OF SIALOGLYCANS RECOGNITION BY *STREPTOCOCCUS GORDONII* SIGLEC-LIKE ADHESINS **122**

5.1 INTRODUCTION	122
5.2 STEADY-STATE FLUORESCENCE ANALYSIS	123
5.3 MOLECULAR BINDING OF STA-THR TO SLBR-H	123
5.4 MOLECULAR BINDING OF STA-THR TO SLBR-B	128
5.5 MOLECULAR BINDING OF 3'SLN TO SLBR-H	133
5.6 THE GANGLIOSIDE GM1B AS NOVEL LIGAND FOR SLBR-H AND SLBR-B	138

5.7 DISCUSSION	140
-----------------------	------------

VII. MOLECULAR BINDING OF DIFFERENT *STAPHYLOCOCCUS AUREUS* WALL TEICHOIC ACID GLYCOFORMS RECOGNIZED BY SPECIFIC ANTIBODIES **170**

7.1 INTRODUCTION	170
7.2 MOLECULAR INVESTIGATION OF IGG MAB 4461 WITH A-GLCNAC MODIFIED WALL TEICHOIC ACIDS	171
7.2.1 IGG MAB 4461 BINDING TO A- WTA-TRIMER	171
7.2.2 IGG MAB 4461 BINDING TO A- WTA-HEXAMER	174
7.3 MOLECULAR INVESTIGATION OF IGG MAB 4497 WITH B-GLCNAC MODIFIED WALL TEICHOIC ACIDS	177
7.3.1 IGG MAB 4497 BINDING TO B-(1,4)-WTA-TRIMER	177
7.3.2 IGG MAB 4497 BINDING TO B-(1,3)-WTA-TRIMER	180
7.4 DISCUSSION	183

GENERAL CONCLUSIONS **185**

VIII. MATERIAL AND METHODS **186**

8.1 PRODUCTION OF RECOMBINANT PROTEINS	186
8.1.1 SIGLEC-2 (RELATED TO CHAPTER III)	186
8.1.2 SIGLEC-7 (RELATED TO CHAPTER IV)	186
8.1.2 RECOMBINANT EXPRESSION OF SIGLEC-LIKE ADHESINS SLBR-B, -H AND -N IN <i>E. COLI</i> (RELATED TO CHAPTERS V AND VI)	190
8.2 NMR ANALYSIS	191
8.2.1 SLBR-N PROTEIN ASSIGNMENT AND TITRATION (RELATED TO CHAPTER VI)	191
8.2.2 NMR LIGAND-BASED TECHNIQUES	191
8.3 BIOPHYSICAL TECHNIQUES	194
8.3.1 FLUORESCENCE ANALYSIS (RELATED TO CHAPTERS III-V)	194
8.3.2 ALPHA SCREEN ASSAY (RELATED TO CHAPTER III)	195
8.3.3 NANO DIFFERENTIAL SCANNING FLUORIMETRY (RELATED TO CHAPTER IV)	196

8.3.4 DYNAMIC LIGHT SCATTERING (RELATED TO CHAPTER IV)	196
8.3.5 ANALYTICAL ULTRACENTRIFUGATION (CHAPTER IV)	197
8.4 <i>IN SILICO</i> ANALYSIS	197
8.4.1 HOMOLGY MODELING (RELATED TO CHAPTER III)	197
8.4.2 MOLECULAR MECHANICS	198
8.4.3 MOLECULAR DOCKING	198
8.4.4 MOLECULAR DYNAMICS (CHAPTERS IV-VII)	198
8.4.5 CORCEMA-ST (RELATED TO CHAPTER III)	200

APPENDIX **201**

**MOLECULAR BINDING OF MAB 2C7 RECOGNIZING A MIMOTOPE OF
NEISSERIA GONORRHOEAE LOS** **201**

1.1 INTRODUCTION	201
1.2 BINDING ANALYSIS BETWEEN MAB 2C7 AND TMCP2	204
1.3 BINDING ANALYSIS BETWEEN MAB 2C7 AND CP2	205
1.4 DISCUSSION	209

BIBLIOGRAPHY

Chapter I:
Introduction

I. Introduction

1.1 The relevance of glycans in biological functions

Glycocalyx is a matrix of complex glycans and glycoconjugates (glycolipids and glycoproteins) at the interface between plasma membranes and the surrounding environment present in all mammalian cells surface.¹ The huge complexity of glycans results from the many combinations in which monosaccharides, the sugar building blocks, can be connected. Indeed, contrary to the linear organization of amino acids and nucleotides, building blocks of proteins and nucleic acids, glycans are highly heterogeneous, can be linear or branched, made up of different monosaccharides with various glycosylation positions and anomeric configurations, substituted with aglycon groups (as acetyl, sulfate) or with proteins or lipids.^{2,3,4}

Among the biological roles, glycans are key actors in modulatory activity, including cell-adhesion and compartmentalization, molecular trafficking, protein folding and stability, and play a crucial role in recognition of proteins involved in the regulation of immune system.⁵ Indeed, several glycan-protein interactions are involved in the critical balance between immune tolerance and generation of a strong immune response. In this context, during the evolution the glycocalyx has developed the ability to distinguish between endogenous (*self*) and exogenous (*non-self*) components. The innate non-specific response is the first defense against *non-self* foreign microorganisms. Here, proteins expressed by cells of the innate immune system (*e.g.* macrophages, dendritic cells, neutrophils, monocytes and epithelial cells), called Pattern Recognition Receptors (PRRs), are activated against microbial organisms, including microbe associated molecular patterns (MAMPs), mostly composed of bacterial glycoconjugates, danger-associated molecular patterns (DAMPs) and cell death-associated molecular patterns (CDAMs). These interactions lead to the activation of cellular mechanisms which trigger proinflammatory cytokines and promote the stimulation of adaptive immunity.⁶

Beyond the role in the regulation of immune responses, glycans can be also implicated in tumor progression and metastasis.⁷ Indeed, peculiar glycosylation patterns (e.g., truncated *O*-glycans, branching and bisecting *N*-glycans, fucosylation or sialylation) are often related to the presence of malignant cells.⁸

1.1.1 The structure of glycoconjugates

Glycans linked to lipids (glycolipids) or to proteins (glycoproteins) are the most abundant glycoconjugates found in fluids and at the extracellular surface of the plasma membrane (figure 1.1). Due to their amphiphilic nature, glycolipids form stable micelles in aqueous solution. The major classes of glycolipids in mammals are: i) glycosphingolipids, where a glycan is linked to ceramide (lipid composed of sphingosine and fatty acids) and mainly involved in cell-cell interactions or cell adhesion events, and ii) glyco glycerolipid, composed of glycerol, lipids and carbohydrates, that exhibit structural functions like membrane bilayer stability, serve as precursors for the formation of complex membrane components, mediate the anchorage of proteins to membrane cells. The presence of substituents, such as phosphate and sulphate groups, confers the negative charge to glycolipids. An example is the ganglioside, a glycosphingolipid mainly found in the brain that shows negative charge due to the presence of sialic acid.⁹

Glycosylation on proteins is an important post-translational modification occurring in mammals. The saccharide portion can be attached through a *N*-glycosidic linkage to an asparagine residue, to form *N*-glycans or via *O*-linkage to a serine or threonine residue, resulting in the formation of *O*-glycans (figure 1.2).¹⁰

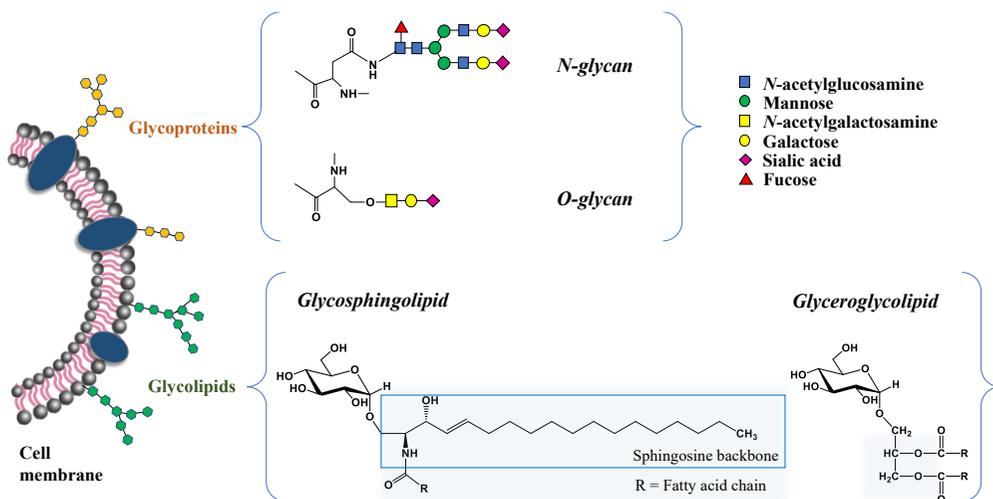


Figure 1.1. Representation of glycoconjugates. Glycans can be found attached to lipids or proteins, to give glycolipids and glycoproteins, respectively. Glycoproteins can be synthesized as *N*-glycans and/or *O*-glycans. The main glycolipids in mammals are glycosphingolipids and glyceroglycolipids.

The biosynthesis of glycoproteins occurs between the endoplasmic reticulum (ER) and the Golgi apparatus. A lipid dolichol pyrophosphate on the ER membrane is the starting point of the co-translational modification process from which the oligosaccharide chain arises. The first phase of the *N*-glycosylation is the anchorage of two GlcNAc and five mannose residues to the lipid dolichol deriving from specific nucleotide sugars, UDP-GlcNAc and GDP-Man, in the cytoplasmic region of ER. The early oligosaccharide chain continues to growth in the lumen, where other monosaccharides are added until a chain composed by $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ is generated. After glycosyltransferases have assembled 14 sugar residues of the glycan, an oligosaccharyl-transferase moves the oligosaccharide chain from the dolichol pyrophosphate to an asparagine, located in the lumen of ER, belonging to consensus sequence Asn-X-Ser/Thr (where X is any amino acid except proline). In a second stage, called trimming, the core region of *N*-glycan is subjected to modifications based on the removal of three glucose residues and one mannose from the chain by specific

glycosidases, to translocate the glycoprotein to Golgi apparatus. In this region, the glycoprotein undergoes further modifications by specific glycosidases and glycosyltransferases depending on forthcoming biological functions they are going to conduct. These result in a significant heterogeneity of *N*-glycans structures.

N-linked glycans are classified in three groups, sharing a common region, the chitobiose core, constituted of three mannoses and two N-acetylglucosamines (GlcNAc), one of which is covalently linked the asparagine residue: high-mannose, complex-type and hybrid (figure 1.2 A).¹¹

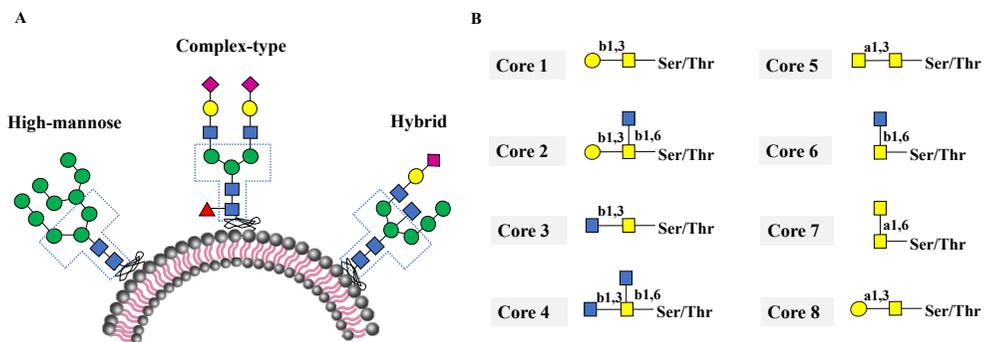


Figure 1.2. Classification of *N*-glycan structures and *O*-glycans core. A) During the biosynthesis of *N*-glycans, three structures can be formed, all sharing the chitobiose core (dashed): high-mannose-type, complex-type and hybrid glycans. High-mannose-type glycans, enriched by the mannose residues in the entire structure, are often expressed on tumor cells.¹² Complex glycans are branched structures, usually ending with α 2,6 or α 2,3 sialic acid linked to a galactose, involved in the recognition by lectins. A mixture of the high-mannose and complex-type *N*-glycans constitutes hybrid structures, associated with several diseases and implied in the regulation of the immune system.^{13,14} B) The core structures of mucin-type glycans. The common structural feature is the presence of GalNAc residue linked to Ser/Thr amino acid.

O-linked glycans are also highly heterogeneous structures. Their biosynthesis does not require a precursor to transfer the oligosaccharide chain to protein; instead, various enzymes can connect the first GalNAc monosaccharide to a serine or threonine residue that does not belong to a consensus O-glycosylation sequence, differently from the oligosaccharyltransferase of *N*-glycans. Furthermore, the O-glycosylation

pattern is considered a post-transductional modification because, even if the synthesis of protein occurs in the ER, the single sugars, starting from the GalNAc residue, are added step by step in the Golgi apparatus.

The most common O-linked proteins are mucins, glycoproteins representing the main components of mucus. Mostly linear, the mucin-type O-glycans contain a GalNAc unit covalently linked to a serine/threonine amino acid (figure 1.2 B). O-linked glycans are implicated in inflammation processes and cell signaling events. 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.

1.1.2 Sialic acid

Sialic acids are a family of more than 50 nine-carbon ulosonic acids decorating the outermost part of N-glycans, O-glycans and glycosphingolipids (Figure 1.3).¹⁵ The anomeric C-2 position is generally α -(2,3)- or α -(2,6)-linked to a hydroxyl group of a galactose or N-acetylgalactosamine residue. Nevertheless, sialyltransferases can also form disialyl core structures, common in gangliosides, where two sialic acid residues are connected by α -(2,3), α -(2,6) or α -(2,8) glycosidic linkages. The other positions of the backbone can be subjected to modifications (*e.g.*, acetylation, sulfation, methylation, and phosphorylation) to originate a variety of structures (figure 1.3).

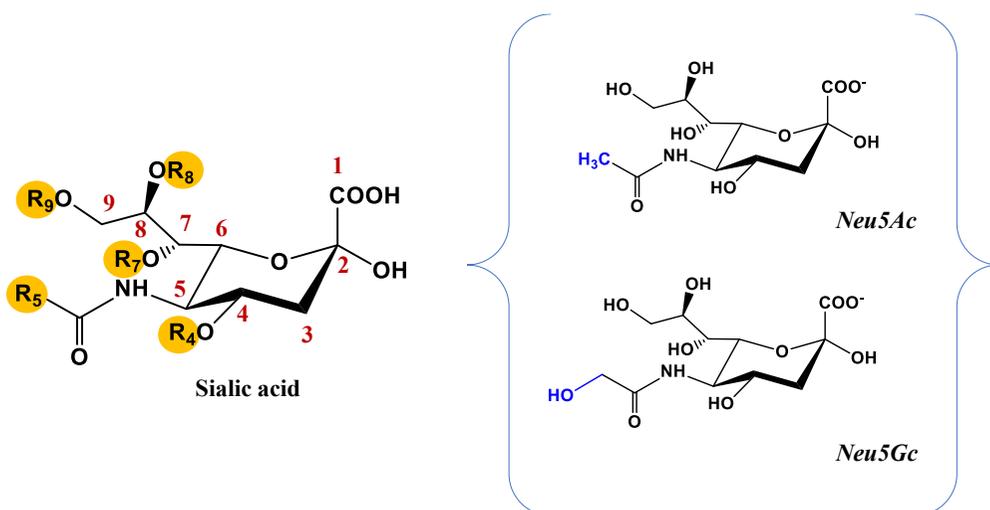


Figure 1.3. Diversity in the Sialic acids. The nine-carbon backbone in the α configuration is shown, with the most common structures in mammals: the acetylated (Neu5Ac) and glycolylated (Neu5Gc) neuraminic acids. Some variations can occur as indicated: at physiological pH, the carboxylate gives the negative charge of Sia; R1 can form lactones with hydroxyl groups on the same molecule or on other glycans or lactams with a free amino group at C-5. R2 = alpha-linkage to Gal (3/4/6), GalNAc (6), GlcNAc (4/6), Sia (8/9), or 5-O-Neu5Gc. R4 = H; -acetyl; Fuc; Gal. R5 = Amino; *N*-acetyl; *N*-glycolyl; hydroxyl; *N*-glycolyl-*O*-acetyl; *N*-glycolyl-*O*-methyl. R7 = H; -acetyl. R8 = H; -acetyl; -methyl; -sulfate; Sia; Glc. R9 = -H; -acetyl; -lactyl; -phosphate; -sulfate; Sia.

The most common sialic acid in mammals is the 5-acetamido-2-keto-3,5-dideoxy-*D*-glycero-*D*-galactonononic acid (*N*-acetylneuraminic acid, Neu5Ac). The presence of CMAH (Cytidine monophospho-*N*-acetylneuraminic acid hydroxylase) gene in mammals, like mice and chimpanzee, is responsible for the expression of the *N*-glycolylneuraminic acid (Neu5Gc), that catalyzes the biosynthesis of a hydroxyl group to *N*-acetyl moiety at 5th position of Neu5Ac. Although the inability of humans to produce Neu5Gc, since CMAH is inactivated, it can be metabolically assumed by diet¹⁶ and is often associated with carcinoma.^{15,17,18,19}

Given their variability, sialic acids are essential in many biological functions,^{20,21,22} including cell signaling and modulation of immune responses.²³ Interestingly, the

ipersialylation (overexpression of sialic acids on cell surface) can be an indication of the presence of cancer.²⁴ Thus, targeting sialylated glycans could be important in novel therapeutic approaches against cancer and autoimmune diseases.

1.2 Glycoconjugates on bacterial cell wall

Cell wall microbial glycoconjugates play fundamental roles in the dynamic host-guest recognition with implications in normal and pathological processes.²⁵

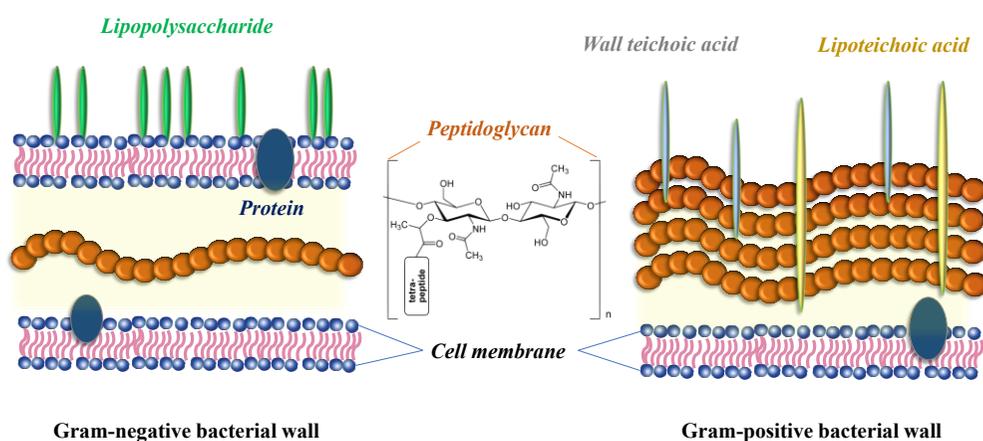


Figure 1.4. Structural organization of cell membranes of Gram-negative (left) and Gram-positive (right) bacteria. The main difference among these bacteria is the thickness of peptidoglycans, which repetitive unit is shown.

Glycoconjugates found on the bacterial cell wall, such as lipopolysaccharides, peptidoglycans, teichoic acids, can act as virulence factors called PAMPs (Pathogen Associated Molecular Patterns) due to their ability to activate host immune response and their inflammatory potential, with implications in immunoevasion and immunosuppression. The peptidoglycan (PG or murein) is the main component of the bacterial cell wall (figure 1.4), to which provides rigidity and structure. It is composed of a network of glycan strands of repeating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) β -1,4-linked units. Alternating L and D amino acids

(4 or 5 residues) are linked to the lactyl group of the muramic acid. The peptidoglycan composition depends on bacterial species. In Gram positive bacteria, it forms a thick layer, constituting 40-80% of the bacterial wall (shell of 30–100 nm), while peptidoglycan in Gram negative species constitutes a thin layer covered by an outer membrane, mainly composed of lipopolysaccharides and lipoproteins.

1.2.1 Lipopolysaccharides (LPSs)

Lipopolysaccharides (LPS) are the major outer surface membrane components of Gram-negative bacteria. They play important roles in the integrity of the outer-membrane permeability, serving as a barrier for bacterial protection, and act as stimulators of innate or natural immunity, due to their extensive participation in host–pathogen interplay. LPS structure varies depending on bacterial strains and defines the immunopotential function. The LPS is an amphiphilic macromolecule composed of three defined regions (figure 1.5): the lipid A, the core oligosaccharide (core OS) and a polysaccharide portion (O-chain). The complete form of LPS is named smooth-type LPS (S-LPS), but some others can terminate with the core oligosaccharide, and they are called rough-type LPS (R-LPS) or lipooligosaccharide (LOS). The hydrophobic lipid A allows the anchoring to the outer bacterial membrane and represents the endotoxin portion of LPS, responsible of its immunostimulatory potential. Indeed, it can trigger the release of proinflammatory mediators via Toll-like receptor 4 (TLR4)-dependent signaling on macrophages and endothelial cells, developing a variety of biological effects up to sepsis and septic shock.²⁶ Lipid A is composed of two β -(1→6)-linked D-glucosamine units variously acylated by fatty acids of different length and phosphorylated at anomeric position of the reducing (GlcNI) unit and at 4' position of the non-reducing (GlcNII) residue (figure 1.5). The inner core OS contains characteristic monosaccharides such as heptoses (L-glycero-D-manno heptose or D-glycero-D-manno heptose) and Kdo (3-deoxy-D-manno-

octulosonic acid), the latter representing a marker of all Gram-negative bacteria and covalently linked to the GlcN II of the Lipid A. The outer core OS is usually composed of hexoses and deoxyhexoses, uronic acids and aminosugars. The external portion of LPS is constituted by the hydrophilic O-polysaccharide, a polymer of saccharide repeating units (up to eight sugars) which can be repeated up to 50 times. The O-chain is highly variable and heterogeneous; indeed, it can be linear or branched, homo- or heteropolymeric, and can be composed by non-carbohydrate substituents.

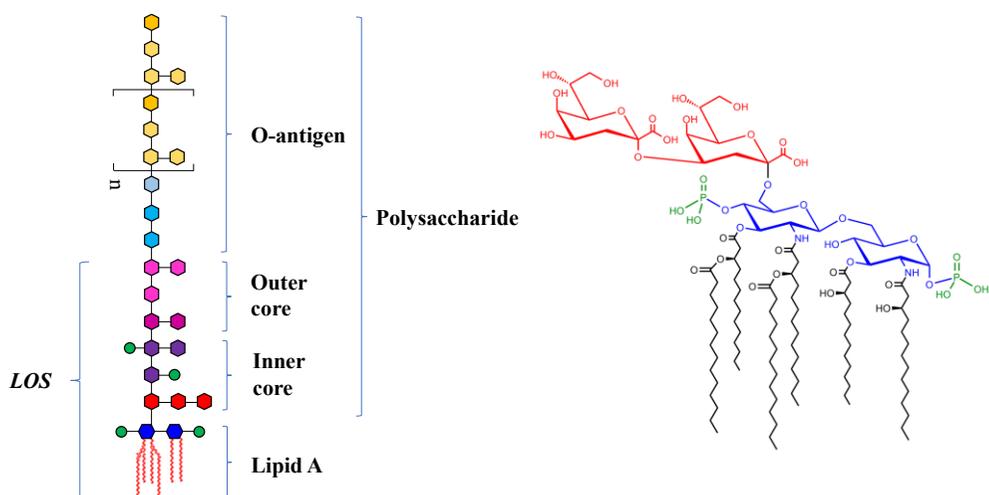


Figure 1.5. Representation of LPS structure on Gram-negative membrane. The main components of LPS are: lipid A, core and O-antigen. Structures lacking O-antigen consist of LOS. On the right, an example of Kdo- α -(2→4)-Kdo (in red) connected to β -(1→6)-linked D-glucosamine residues (GlcN I and II, in blue) of the LPS of *E. coli*.

1.2.2 Teichoic acid glycopolymers

Cell envelope components attached to PG in Gram-positive bacteria can be capsular polysaccharides (CP) and polyanionic polymers called teichoic acids (TAs), promising targets for anti-infective therapies and vaccines due to their key role in host-cell interactions. There are two different kinds of TAs: lipoteichoic acids (LTA) that

are tethered to the membrane via glycolipids, and cell wall teichoic acids (WTAs), instead covalently anchored to peptidoglycan in the cell wall. The chemical structure of both glycopolymers varies among Gram-positive organisms.

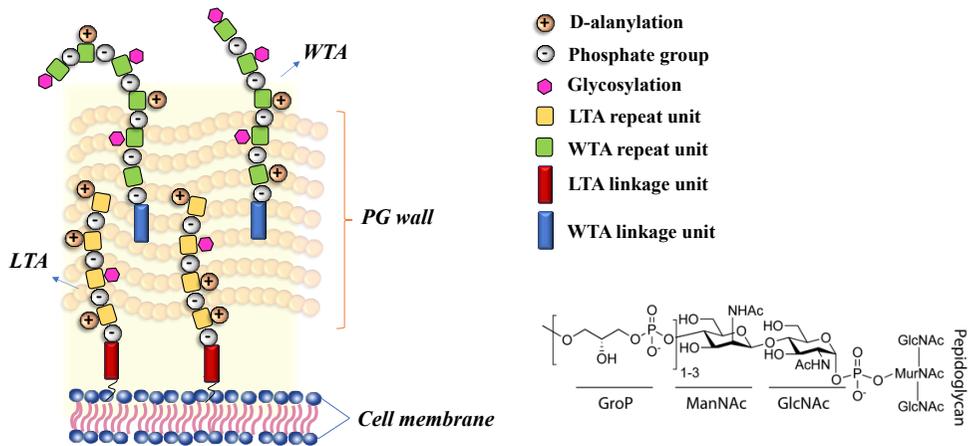


Figure 1.6. LTA and WTA structures on Gram-positive cell membrane. The common part of a WTA linkage unit consists of a GroP-ManNAc-GlcNAc-phosphate covalently attached to peptidoglycan (right bottom).

LTA structure is generally composed of a polyglycerol-phosphate chain linked to a glycolipid in the membrane. WTAs are typically composed of a conserved *N*-acetylmannosamine ($\beta 1 \rightarrow 4$)-*N*-acetylglucosamine-1-phosphate (ManNAc-($\beta 1 \rightarrow 4$)-GlcNAc-1P) disaccharide unit. The reducing end is connected to PG by a phosphodiester bond, while the non-reducing sugar is linked to a polymeric backbone of phosphodiester-linked polyol repeat units,²⁷ with the best-characterized containing repeat units of 1,5-D-ribitol-phosphate (RboP) or 1,3-L- α -glycerol-phosphate (GroP). The presence of phosphate groups imparts the anionic charges to the glycopolymers. The main chain polymer in TAs can be enriched by sugar moieties, depending on bacterial species. Moreover, glycopolymers are often decorated with D-alanine esters occurring at C2 position of the backbone, whose positive charges neutralize the negative charges of phosphate groups. These sugar and alanine modifications on TAs

have been implicated in different functions in cell physiology and infection.²⁸ It is well known, for example, the TAs glycans recognition by some lectins, such as langerin and macrophage galactose-type lectin (MGL) on immune cells.²⁹ Due the involvement of TAs in many biological functions, including cell adhesion, host colonization, virulence, and pathogenesis, they are considered attractive targets for therapeutics against the antibiotic-resistant infections.³⁰

1.2.2.1 Wall teichoic acids in *Staphylococcus aureus*

Staphylococcus aureus is a Gram-positive bacterium that cause severe infections, including bacteremia, staphylococcal toxic shock syndrome, endocarditis and osteomyelitis.^{31,32} The matrix of peptidoglycans (PG) constituting *Staphylococcus aureus* is functionalized up to 60% with WTAs, which are composed of up to 40 RboP subunits,^{33,34} modified with D-alanine and N-acetylglucosamine (GlcNAc).

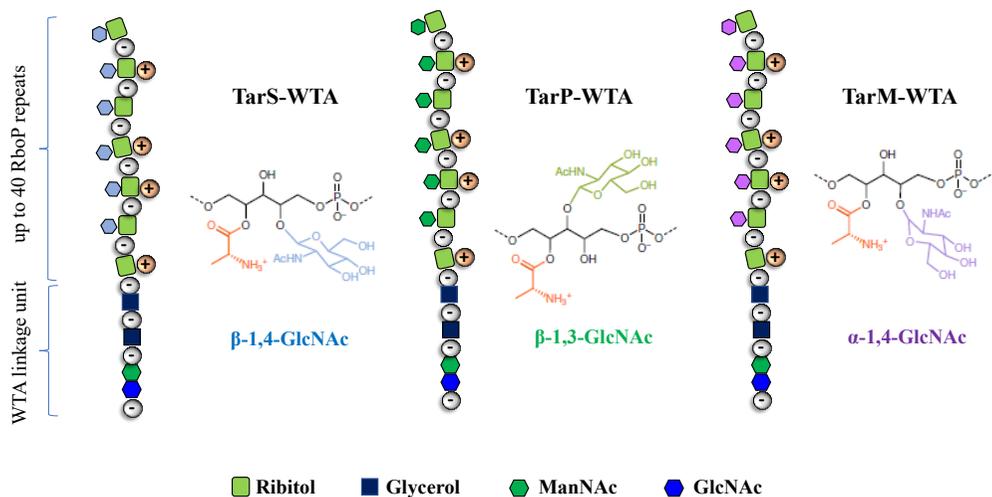


Figure 1.7. *S. aureus* WTA structures and their variation by different GlcNAc transferases. The three identified RboP WTA variants generated by the glycosyltransferases TarS, TarP and TarM are shown.

The sugar modification on *S. aureus* is mediated by specific Tar (teichoic acid ribitol) glycosyltransferases, whose activity plays crucial roles in cell shape formation, regulation of cell division and other crucial aspects of Gram-positive bacterial physiology.^{35,36} The glycosylation can occur at position 3 or 4 of RboP backbone, with GlcNAc residue linked in α - or β -configuration (figure 1.7).

In particular, TarM and TarS catalyze the α -1,4-GlcNAc and β -1,4-GlcNAc,²⁷ while TarP modifies RboP by attaching a β -GlcNAc at position 3 of the backbone (β -1,3-GlcNAc),³⁷ leading to a less immunogenic WTA polymer.³⁵ These sugar modifications are differently recognized by both innate and adaptive immune system, thus impacting the capacity of host-mediated immune detection and clearance.³²

S. aureus WTAs have developed the β -lactam antibiotic resistance, causing the infection difficult to treat. Therefore, non-antibiotic therapeutic based strategies are urgently needed. For this reason, vaccines and therapeutic antibodies, some of them

currently in clinical trials, are promising treatments to overcome this bacterial infection.

1.3 Structure and function of antibodies

Antibodies or immunoglobulins are Y-shaped globular proteins involved in the recognition of *non-self* molecules, called antigens. Structurally, antibodies contain a larger subunit (50 kDa), called heavy chain and indicated as **H**, and a smaller portion (23 kDa), called light chain, known as **L**, associated to another identical heterodimer, all linked through disulfide bonds (figure 1.8).³⁸ Digestion with papain cleaves the antibody in three parts: two identical **Fab** arms (fragments of antigen-binding) at the N-terminal part of H and L chains containing the antigen-binding determinants, thus important in the selectivity of the antibody, and the **Fc** stem (fragment crystallizable), not involved in the antigen recognition, but important to define the biological functions of the immunoglobulin. In some antibodies, the arms are connected to the stem by flexible hinge regions. Each heavy and light chain contains a large constant (**C**) region with amino acids sequences in the C-terminal domain and a smaller but similar-sized variable (**V**) region. The variability in the V region is given by the so-called hypervariable regions or complementarity-determining regions (CRDs), each containing approximately 10 amino acid residues, that form a cleft between H and L chains and define the antigen-binding sites. In particular, three CRDs are located in each V domain, denoted as CRD-H1, CRD-H2 and CRD-H3 in the heavy and CRD-L1, CRD-L2 and CRD-L3 in the light chain. Among the CRDs sequences, conserved amino acids, named framework segments, represent about 85% of the V region.

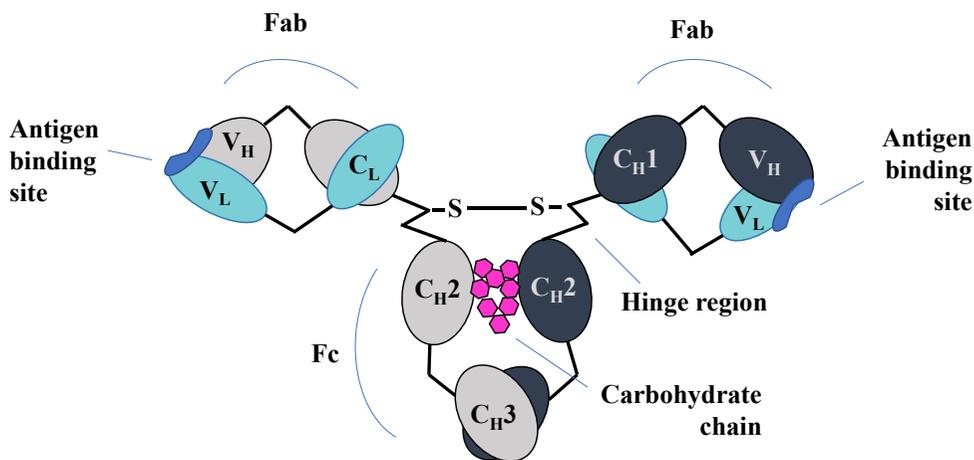


Figure 1.8. Representation of IgG antibody. The variable domains of the heavy and light chains are indicated as V_H and V_L , respectively, and contain the complementarity-determining regions for the antigen recognition. Constant domains for each chain are present in the Fab (C_H1 and C_L), whereas C_H2 and C_H3 belong only to the Fc of the heavy chain.

The constant and variable regions of the antibody are defined as domains, all having roughly 110 amino acids in length, folded in anti-parallel β -sheets to form a compact and globular structure. Depending on the β -strands organization in the β -sheets between the domains, the V region results less compact with longer loops with respect to the C domain.³⁹ Although the domains of different antibodies are folded in similar manner, changes in amino acid residues at position of the cavity between H and L chains defined by the CRD regions change the shape and the specificity of the entire antibody. Among the regions, CRD-H3 shows high diversity in amino acids sequence and conformation, indeed it is mainly involved in the antigen recognition. In this context, the antigenic determinants of the antibodies, which induce the immune system activation, show three levels of variability, classifying these molecules as: isotypes, allotypes and idiotypes.

Isotype antibodies are species-dependent since all members of given species inherit constant regions genes in a normal individual and define the antibody classes and

subclasses of heavy chains and subclasses of light chains. Humans express five classes of antibody, known as IgG, IgA, IgM, IgD and IgE, which show different chemical and serological properties. The heavy chain determinants of these isotypes are in turn divided in subclasses indicated with Greek letters (γ , α , μ , δ , ϵ). Regarding the light chains, there are only two different isotypes, called κ and λ , in 60:40 ratio in humans, and 95:5 in mice. Fab regions containing λ light chains display more flexibility with respect to those having κ light chain, because, in the first case, the angle between the V and C domains can assume a wider range of values. Isotype antibodies are important in the measurement of Ig levels, in the identification of B cell tumors or in the detection of an immunodeficiency.

Allotypic antigenic determinants are found in constant regions of heavy and light chains and depend on the allelic form of a given antibody gene. Indeed, it is possible that members of a species inherit the same set of gene with a modification in one or multiple alleles. These allotypes are observed for example during pregnancy or blood transfusion and are important in forensic application, paternity testing or to monitor bone marrow grafts.

Idiotypic determinants are individual-specific. They are found in the V region because determine the antigen binding specificity. In this regard, it is known their import role in the treatment of B cell tumors and vaccines.

1.3.1 Classes of antibodies

The human serum is mainly composed of IgG antibodies (80%), important molecules involved in coating antigens and enhancing their phagocytosis by macrophages and neutrophils. This immunoglobulin is a dimer of 150 kDa with two heavy chains with four γ subclasses and two light chains (κ or λ), where all the polypeptide chains are held by disulfide bonds and non-covalent interactions. In particular, the single variable domain of each heavy chain (V_H) is coupled to the corresponding variable domain of

the light chain (V_L). Since the constant portion of the IgG antibody has one domain in the light chain and three domains in the heavy chain, S-S bonds occur between the respective C_L and C_{H1} domains in the Fab fragment and among each C_{H2} and C_{H3} in the Fc fragment (figure 1.9 A).

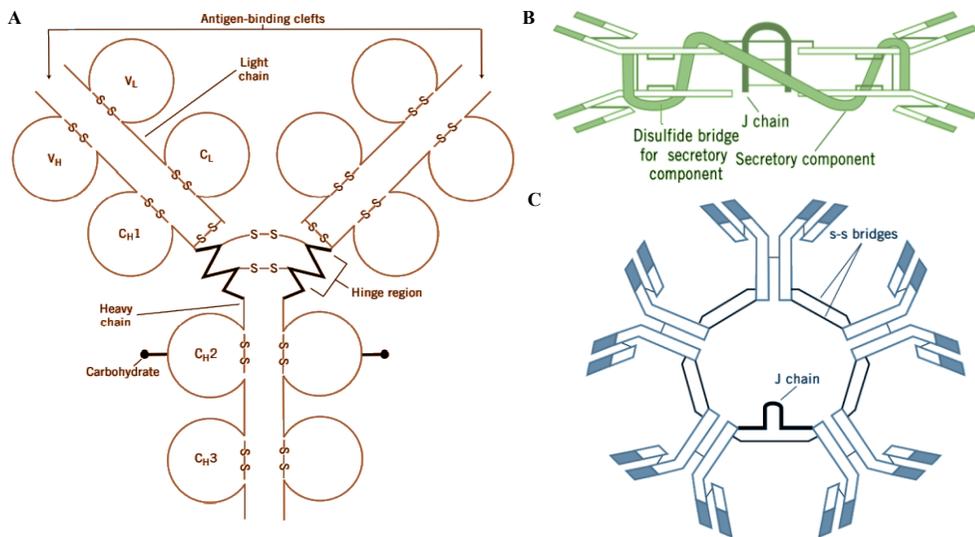


Figure 1.9. Examples of antibodies structure: A) IgG, B) IgA (dimeric form), C) IgM (pentameric form).

Additionally, hydrogen bonds can occur at the C_{H2} - C_{H2} interface, due to the presence of a carbohydrate residue covalently bound to Asn297 of each Fc fragment. Moreover, the hinge region of IgG antibodies connects Fabs and Fcs and is located in the middle of C_{H1} and C_{H2} domains, helping both Fab to interact to multiple targets and Fc to independently communicate with other elements of the immune system.

Among the classes of antibodies, only IgG, IgA and IgD have the flexible hinge regions. Furthermore, additional amino acids, called tail pieces, are present on the C-terminal of the C_{H3} domain of the heavy chain of IgM, IgA and IgD. These residues allow antibodies to interact with other molecules to form multimeric structures, also

stabilized by their disulfide interactions with the joining (J) chain polypeptide found on IgM and IgA.

IgA is another class of human antibodies, that has two subclasses, called IgA1 and IgA2. Generally, IgA is found as dimeric form of 390 kDa (figure 1.9 B), with the two monomers linked by a 15 kDa chain in the Fc. Only 13% of serum contains IgA, but it is prevalently found in extravascular secretions (respiratory, gastrointestinal and urogenital tracts), indeed this antibody is also known as secretory IgA, and plays a fundamental role in the immune function of mucous membranes.

The largest antibody is the IgM, a 950 kDa pentameric polypeptide found for 8% in the serum (figure 1.9 C). The Fab fragments are oriented outward, while Fc tails are cross-linked by disulfide bonds that can be connected by the J chains. Due to the high avidity to bind different antigens at the same time, IgM represents the first immunoglobulin appearing in the immune response and a crucial activator of complement. IgM is also the first antibody formed by a developing fetus.

The monomeric forms of IgD (175 kDa) and IgE (190 kDa) are the lowest components of the serum, with percentages below 1% and 0.003% respectively. Co-expressed with IgM, IgG antibody is found on B cells surface, thus taking part in the activation of the immune system through antigen internalization. It is also known the ability of IgD to bind to and activate basophils and mast cells to produce antimicrobial factors in respiratory tract defense.⁴⁰ IgE is an antibody found only in mammals and triggers the symptoms of allergies.

Antibodies play crucial roles in protecting the body from bacteria and virus entry and activating a long-lasting immune response. The action is mediated by the recognition (neutralization) of foreign molecules from the V region of the antibody that determines the specificity toward the antigen and that can activate complement. Then, the immunological activities can be triggered by the interactions between Fc domains and specific Fc receptors (FcR). However, in aberrant situations immune system may

react against the “self” molecules, producing antibodies (autoantibodies) to attack them, causing autoinflammatory diseases, such as rheumatoid arthritis and multiple sclerosis.⁴¹ A pharmacological tool for the treatment of autoimmune diseases, firstly introduced for cancer therapy, is the use of monoclonal antibodies.⁴²

1.3.2 Monoclonal antibodies

A monoclonal antibody (mAb) is an artificial antibody with a single antigenic determinant that specifically targets a certain antigen. In 1975, the immunologists Georges Kohler and Cesar Milstein fused antigen-specific B cells from the spleen of an immunized mouse with myeloma cells to form a *hybridoma*, a cell with the specificity of lymphocyte’s antibody and the immortality of the tumor cell. Hybridoma cultures can be an easy source of monoclonal antibodies that, alone or joined to other molecules (drugs or radioactive isotopes), recognize a single antigenic site on almost any molecule, avoiding to involve other sites. For this reason, monoclonal antibodies are often used for diagnostic purposes, as in the identification of tumor cells, and therapeutic goals, for example against inflammatory and immune diseases. A method to humanize the monoclonal antibody is the transfection of the hybridoma cells, that is the integration of a DNA into a cell’s chromosomes. This recombinant DNA technique is based on the isolation of the gene encoding for the antigen specificity of the antibody and subsequent fusion with a human DNA encoding for an antibody. The hybrid antibody is then grown in bacterial media. Another way to produce human mAb is the phage display technique, that uses bacteriophages to produce fusion proteins on phage surface which leads to a combinatorial library.^{43,44}

There are different kinds of mAb: 1) *murine*, that are produced from mouse proteins (drugs end in -omab); 2) *chimeric*, that are a combination of mouse and human proteins (treatments end in -ximab); 3) *humanized*, non-human species from small parts of mouse proteins whose sequences have been modified to increase the similarity

to human antibody (end in -zumab) and 4) *human*, fully human proteins (drug treatments end in -umab). An example was the adalimumab, the first human mAb approved in 2002 for rheumatoid arthritis therapy.

It is not excluded that mAb can be trigger allergic reactions or immune responses in the body that recognizes them as “non-self” molecules. Although these side effects for some patients, the potentiality of the monoclonal antibody-based method in cancer, infections and immune diseases is constantly developing.

1.4 I-type Lectins: Siglecs

Glycan-binding proteins (GBPs) bind carbohydrates exposed on cell surfaces and their interaction play significant roles in several cellular mechanisms, including cell-cell communication, immunomodulation and inflammation processes.^{25,45,46,47} Among GBPs, lectins are ubiquitous macromolecules that modulate immune responses to pathogens and interact with carbohydrates to mediate adhesion or signaling events. The major lectin families include C-type lectins, I-type lectins, P-type lectins, and or S-type lectins or galectins (figure 1.10), classified according to their structure, specificity for carbohydrates and species location.

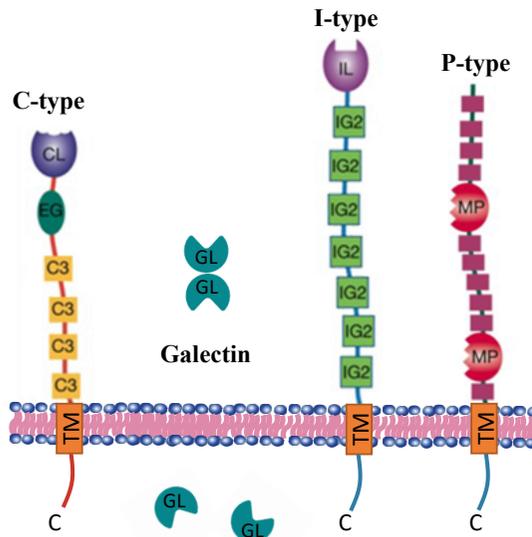


Figure 1.10. Classification of lectins. C-type lectins bind sialoglycans in a calcium-dependent mechanism; I-type lectins contain an immunoglobulin-like carbohydrate recognition domain (CRD); P-type lectins are specific to glycoproteins containing mannose 6-phosphate; galectins are thiol-dependent soluble proteins and specific to β -galactosides. C-type lectin CRD (CL), galectin CRD (GL), P-type lectin CRD (MP), I-type lectin CRD (IL), EFG-like domain (EG), immunoglobulin C2-set domain (IG2), complement regulatory repeat (C3), transmembrane region (TM).

Among the I-type lectins, Siglecs have attracted a lot of interest in the innate and adaptative immune system.^{10,48,49,50,51}

1.4.1 Structural features of Siglecs

Siglecs are transmembrane I-type receptors that vary in their length and specificity for sialic acid-containing ligands.^{52,53} To date, 15 Siglecs have been identified in humans, and 9 in murine species (figure 1.11), all containing an extracellular N-terminal V-set Ig (Ig-V) domain, responsible for the binding of sialoside ligands, connected to 1–16 C2-set Ig domains by a disulphide bridge.^{10,53,58} In the cytoplasmic region, Siglecs

contain one or multiple tyrosine-based signaling motifs, variable cytosolic tails that trigger cellular signaling.

The extracellular *N*-terminal Ig-like domain has nine β -strands (named A-G) assembled with a distinct topology, has close sequence homology to Ig V regions and is the region where sialic acid containing ligands are selectively accommodated.⁵⁴ A critical arginine on the F strand plays a key role in the sialoglycans recognition, forming a bidentate salt bridge with the ionized carboxylic group of sialic acid.⁴⁹ The formation of CH- π contacts between aromatic side chains of the protein and C-H bonds of the carbohydrate's hydrophobic faces is another determinant recurring in the Siglecs recognition.⁵⁵ Indeed, a disulfide bond occurs between B and E β -strands, allowing separation between the β -sheets and consequent exposure of aromatic residues on the A and G strands,⁵⁸ establishing CH- π interactions with lateral glycerol chain and *N*-acetyl group of sialic acid.⁵³ A conserved loop (CC' loop) between F and G strands also contributes to the ligand recognition, especially in the interaction with longer glycan chains,⁵⁸ and sometimes undergoes a conformational change upon sialic acid binding, as observed for Siglec-7 in complex with the GTb1 ganglioside containing Neu5Ac- α -(2-8)-Neu5Ac.⁵⁶

Most of Siglecs contain cytosolic immunoreceptor tyrosine-based inhibition motifs (ITIMs) that can function in inhibitory capacities.⁵⁷ In this case, the interaction between Siglecs and sialylated ligands drives the ITIM domain to execute the signal to the downstream receptor, inhibiting the immune cell activation. The mechanism involves the binding and activation of phosphatases, such as Src homology region 2 domain-containing phosphatase-1 (SHP-1) and Src-homology 2-containing inositol 5' phosphatase (SHIP).^{58,59,60} Such signaling pathway is initiated by the phosphorylation of the tyrosine residues on ITIMs by the Src family kinases.^{48,59} Thus, inhibitory Siglecs control immune reactions serving as negative regulators of immune cells to limit an excessive inflammation state in the host and prevent autoimmune diseases.⁵³

On the other hand, few Siglecs, such as Siglec-14, Siglec-15 and Siglec-16 contain immunoreceptor tyrosine-based activation motifs (ITAMs). Here, the signaling is activated via Syk family tyrosine kinase, through the association of ITAMs to adapter proteins such as DAP12 (DNAX-activating protein of molecular mass 12 kDa).^{48,53,61,62}

Other Siglecs, such as Siglec-1 and Siglec-4, feature neutral transmembrane domains, without signaling cytosolic motifs, and their function is only related to sialic acid-binding, for example for cell adhesion.^{19,63,64}

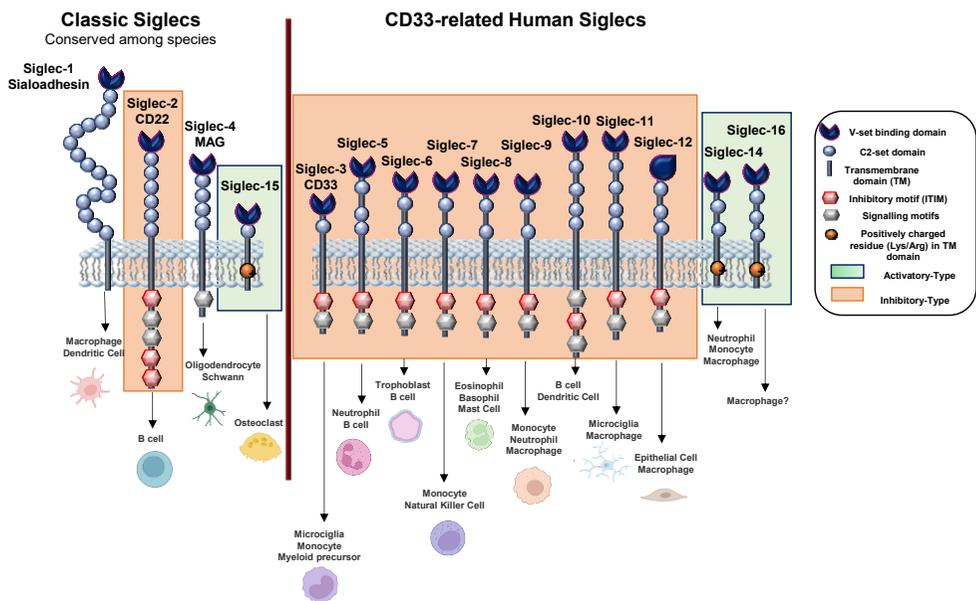


Figure 1.11. Scheme of human Siglecs. Siglecs family is divided in “evolutionary conserved”⁶⁵ (Siglecs -1, -2, -4 and -15) and “CD33-related”^{62,66,67} (Siglecs -3 and from -5 to -16) categories, depending on their sequence similarity and conservation across the orthologs.⁴⁹ The biological functions of Siglecs depend on the nature of the cytoplasmic tails and the transmembrane domain features that distinguish between activatory and inhibitory proteins.

Therefore, Siglecs are proteins expressed on the surface of immune cells and their binding to sialoglycans allows cell–cell communication and regulation of innate and adaptive immune system,^{10,48,62} including tolerance in B-lymphocytes, modulation of T-cell activation, homeostasis and inflammation.^{19,48,62,68}

1.4.2 The Siglecs-sialoglycans interaction

Siglecs can bind sialylated ligands present on the same cell that expresses the receptor (*cis* interaction) or with sialylated structures found on different cell or proteins (*trans* interaction) (figure 1.12).⁶⁹ This is a dynamic competition that depends on the ligand affinity and accessibility. *Cis* interactions commonly occur because of the high local concentration of sialoglycans present on immune cell surfaces.

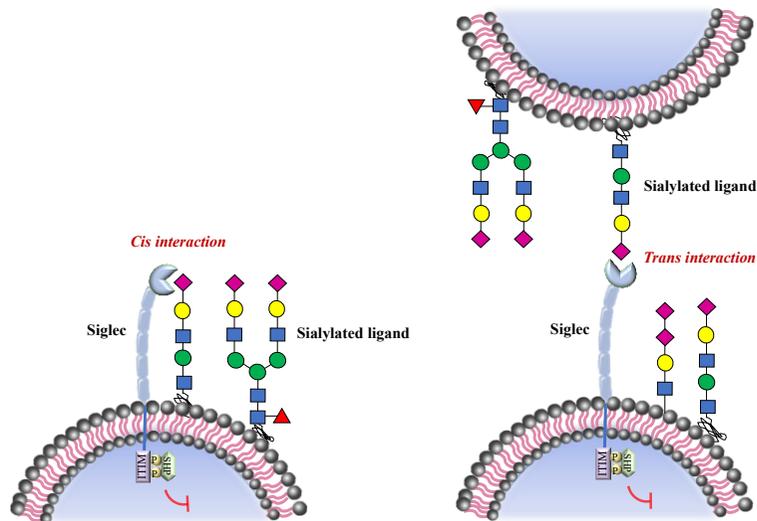


Figure 1.12. Siglec-sialoglycan binding mode. Siglecs can interact with a ligand expressed on the same (*cis*) or a different (*trans*) cell surface.

Here, Siglec, expressed on the same cell of ligands, results 'masked' by *cis* interactions, thereby preventing non-specific cell-cell interactions that could

trigger unnecessary signaling.⁶⁷ Thus, Siglec-sialoglycan binding on the same cell is essential for the modulation of the signaling.

On the other hand, since Siglecs show relatively low affinity for their endogenous ligands, when higher density of ligands is found in proximity of another cell, such as a pathogen, *trans* interactions prevail. In this case, the biological function of the Siglec is activated.⁷⁰

The modulation of immune response from Siglecs is correlated to their ability to discriminate between “self” (endogenous) and “non-self” (exogenous) molecules. Worthy, some human pathogens, including group B *streptococci* (GBS), *Neisseria* species, *Campylobacter jejuni*, have evolved the ability to subvert the host immune response. Indeed, these exogenous molecules mimic SAMPs (self-associated molecular patterns) structures, for example coated themselves of sialylated capsular polysaccharides (CPS) or lipooligosaccharides (LOS), and result mistakenly recognized as “self” molecules, avoiding the activation of immune response and promoting the host colonization.^{19,71,72,73}

Moreover, aberrant glycosylation can also occur on malignant cells. An over-expression of sialic acids is typical of tumor cells and is strictly related to immune suppression.^{74,75} Thus, Siglecs have been studied as attractive targets for the design of therapeutic agents, such as antibodies or glycomimetics, for the treatment of inflammatory, autoimmune, and infectious diseases and for the reduction of cancer progression.^{69,76}

1.4.3 Siglec-2

Siglec-2 or CD22 is an evolutionary conserved inhibitory Siglec expressed on B cells involved in the inhibition of the B cell antigen receptor BCR signals and inducing tolerance to *self*-antigens to prevent autoimmune diseases.^{49,53,77,78} The N terminal V set domain of Siglec-2 selectively binds sialic acids α -(2-6)-linked to a galactose

residue on endogenous glycoproteins of mammalian cells. The crystal structure of human CD22 in complex with α -2,6 sialyllactose and further conformational studies of complex-type *N*-glycans show that the terminal Neu5Ac- α -(2-6)-Gal disaccharide is the only portion recognized the receptor.^{78,79}

In resting B cells, CD22 binds adjacent self sialylated glycans via *cis* interactions, forming CD22 homo-oligomers.⁸⁰ When *trans* interactions occur, the presence of ITIMs in the cytosolic tails of CD22 triggers the activation of phosphatases which dephosphorylate positive components of the B-cell antigen receptor (BCR) signaling cascade. This provokes the disruption of CD22 oligomers, increasing CD22-BCR association and enhancing Ca²⁺ inhibition upon anti-IgM stimulation, consequently leading to suppression of immune response.

The correlation of Siglec-2 to the modulation of B cell tolerance and to many autoimmune diseases, such as rheumatoid arthritis, Systemic Lupus Erythematosus (SLE) and hairy cell leukemia in humans,^{81,82} makes this inhibitory receptor a candidate target in immunomodulation therapies.^{62,66,83,84}

1.4.4 Siglec-7

Siglec-7 is an inhibitory receptor belonging to the CD33 related Siglecs family. This protein is mainly expressed on innate lymphoid natural killer NK cells, but is also found on T cells, eosinophils, monocytes and dendritic cells. The extracellular domain is characterized by the presence of two C2 Ig spacers and a N-terminal V set domain that preferentially binds α -(2,8)-linked disialylated ligands, generally found as terminal portions of various gangliosides. The crystal structure of the Siglec-7 V-set domain was the first to be solved among CD33-related Siglecs and it has been widely studied in complex with different sialylated ligands, containing the key Arg124 residue that establishes a conserved contact with sialic acid. As mentioned before, Siglec-7 undergoes a significant conformational change of CC' (R67-W78) loop upon

binding to GT1b ganglioside, containing the Neu5Ac- α -(2,8)-Neu5Ac- α linkage, the favorite epitope of Siglec-7.⁵⁶ Worthy, a second binding site has been recently discovered, comprising Arg67 in addition to Arg124, suggesting that the two ligand-binding sites are potentially controlled by each other due to the flexible conformation of the CC' loop of Siglec-7.⁸⁵

Trans interactions between Siglec-7 and cognate ligands lead to the phosphorylation of ITIM sites recruiting phosphatases SHP1/2 which impede the NK cell activating pathways. Thus, the immune system allows the evasion of tumor cell and consequent migration within the circulatory system. Therefore, as a negative regulator of NK cell-mediated functions, crucial within tumor immunosurveillance, Siglec-7 has recently emerged as target molecule for cancer immunotherapy.

Although the central role in cancer, Siglec-7 is also involved in other diseases and pathologies, such as HIV-1, obesity, hepatitis, as emerged over the last years.^{86,87} Interestingly, GQ1b-like epitopes containing LOS of *Campylobacter jejuni*, involving in Guillain-Barré syndrome (GBS), can be recognized by Siglec-7, leading to the modulation of host-pathogen binding.⁸⁸ Moreover, the presence of sialylated lipopolysaccharide on certain *Fusobacterium nucleatum* strains, oncogenic pathogen in different human tissues, may induce the activation of Siglec-7, causing immunosuppression that may promote its carcinogenic behavior.⁸⁹

1.4.5 Siglec-10

The presence of one ITIM domain in the cytosolic tail of Siglec-10 allows to define it as an inhibitory protein belonging to CD33-related Siglecs family.⁹⁰ As CD22, Siglec-10 is expressed on B cells surface, but it can be also found on myeloid and dendritic cells and on subsets of human leukocytes, such as neutrophils and macrophages.^{91,92,93,94} However, while CD22 is highly specific for α -(2,6) sialoglycans, Siglec-10 can recognize both α -(2,6) and α -(2,3) sialylated ligands. The

crystal structure of this inhibitory Siglec has not been solved yet, but our recent homology modeling studies have suggested a 3D structure of human Siglec-10.⁹⁵ In particular, differently from CD33-related Siglecs, we observed that the CC' loop conformation of Siglec-10 points outward to the binding residues, allowing to accommodate sialoglycans with different shapes and lengths.

Siglec-10 is associated to several patho-physiological processes, for example, it is known that its binding to CD24 cells promotes the tumor immune evasion.⁹⁶ Worthy, Siglec-10 is also able to bind sialic acid analogues, such as pseudaminic acid on *Campylobacter jejuni* flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor promoting an anti-inflammatory response.⁹⁷

1.5 Bacterial adhesins

Adhesins are virulence factors involved in bacteria attachment to host cells. These bacterial proteins play important roles in some cell signaling processes, in mediating cell–cell and cell–extracellular matrix interactions and in the infection process. Adherence is indeed the crucial step in bacterial pathogenesis.⁹⁸

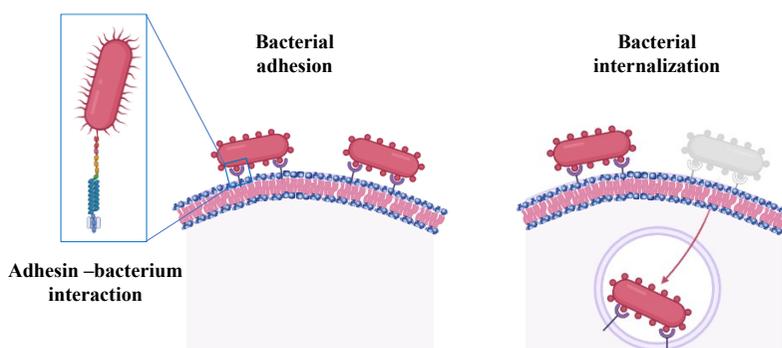


Figure 1.13. Bacterial adhesion (left) on host cell and subsequent internalization in internal vesicle (right).

Host-pathogen interactions are typically required for bacterial colonization or internalization and are mediated by adhesins on the microbial surface (figure 1.13). The binding event may involve and trigger a complex signal transduction cascade in the host cell that can lead to the activation of innate host defenses or the subversion of cellular processes facilitating bacterial colonization or invasion. Depending on the biochemical role, the nature of the adhesion can vary: the binding may be weak and nonspecific, with hydrophobic interactions establishing to the host surface; other adhesins can make highly specific interactions giving rise to high-affinity and stable interactions.⁹⁹

Most pathogens possess more than one adhesin on their surface, often acting in cooperative manner, thus the interaction between the pathogen and the host will depend on which receptor or sequential combination of receptors is engaged.¹⁰⁰ Bacterial adhesins are attached to thin thread-like structures, called *pili* or *fimbriae*, that extend outward from the bacterial cell surface. These protein appendages vary in lengths (generally one-micron long) and diameters (2–10 nm) and consist of several hundred major subunits tipped or interspersed with minor subunits, one or more of which carries the adhesive function.¹⁰¹ *Fimbriae* are classified depending on the host cell receptor with which they interact. The adherence and the colonization of some Gram-positive bacteria are usually mediated by the presence of surface adhesins that interact with host proteins found on the surface of damaged valves. The adhesins of Gram-positive bacteria are attached on the surface by different mechanisms. One includes the anchoring of the adhesin through covalent linkage via its LPXTG motif to the cell wall peptidoglycan. Another mechanism is the association of adhesins with surface proteins, as for lipoteichoic acid (LTA) that forms LTA-binding proteins complexes (e.g., M protein), which together bind the streptococci to fibronectin on the animal cell surface. The lectin-carbohydrate recognition is the type of adhesion shared by most bacterial pathogens. Indeed, many bacterial adhesins are lectins, a family of sugar-binding proteins that recognize carbohydrate moieties of glycolipids or

glycoproteins on the mammalian host cell. The opposite case can also happen, where polysaccharides on either the capsule or the outer membrane lipopolysaccharides of bacteria bind to cognate lectins on the host cell surface. Other kinds of interactions can involve a bacterium surface protein to a complementary protein on the mucosal cell surface, and, the last characterized, the binding interaction between hydrophobic moieties of proteins with lipids on the other cell.¹⁰²

Streptococcus and *Staphylococcus* species produce head-stalk-type adhesins, for example serine-rich repeat proteins (SSRP)¹⁰³ that are anchored into the cell wall and bind to sialylated glycoconjugates.¹⁰⁴ In *S. aureus*, in particular, the adhesins can be covalently bound to cell-wall peptidoglycans, and are known as MSCRAMMs (Microbial Surface Component Reacting with Adhesive Matrix Molecules),¹⁰⁵ or can be secreted and rebound to the bacterial surface, known as SERAMs (secretable expanded repertoire adhesive molecules).¹⁰⁶

Due to the role of adhesins in the pathogenesis, several vaccines have been developed. The activity of anti-adhesin antibodies can indeed disrupt the interaction between bacterium and host cell, rendering it non-pathogenic. For example, the anti-adhesin vaccine to enterotoxigenic *E. coli*, that usually attaches to upper intestinal mucosa in humans, leading to diarrhea and infection, has been proven highly effective. However, some issues must be addressed, including the variety and the large number of bacterial adhesins and the fact that they depend on the local environment.

In the next paragraph the serine-rich repeat proteins (SSRP) found on different Streptococcal bacteria strains, called Siglec-like adhesins, have been described, highlighting their role in the colonization of heart valves and in the pathogenesis of the infective endocarditis.

1.5.1 Siglec-like adhesins

Streptococcal species involved in the pathogenicity of the infective endocarditis (IE) can contain serine-rich repeat glycoproteins (SRRPs),¹⁰⁷ named Siglec-like adhesins, or, as recently found, the so-called AsaA proteins (associated with sialic acid adhesion A) present in IE-isolates lacking SRRPs.¹⁰⁸ A novel sialic acid-binding adhesin present in multiple species contributes to the pathogenesis of IE.¹⁰⁸ The pathogenesis and etiology of IE have been partially defined and typically originate when commensal bacteria transit into the bloodstream.¹⁰⁹ Whereas some species such as *Staphylococcus aureus* may infect native or prosthetic valves and cause acute disease, the Mitis group of oral streptococci tend to infect damaged valves and cause more chronic, sub-acute disease.^{110,111} Pieces of evidence suggest that the adherence of oral streptococci to platelets represents a crucial step in the pathogenesis of IE and this process is mediated in part by the presence of serine-rich repeat (SRR) proteins anchored to the bacterial cell wall (figure 1.14).

Depending on the organism to which the SRR adhesins belong, the BRs can vary in amino acid length and sequence, and in secondary structure and folding, and these characteristics define the ligand specificity for different bacterial strains.

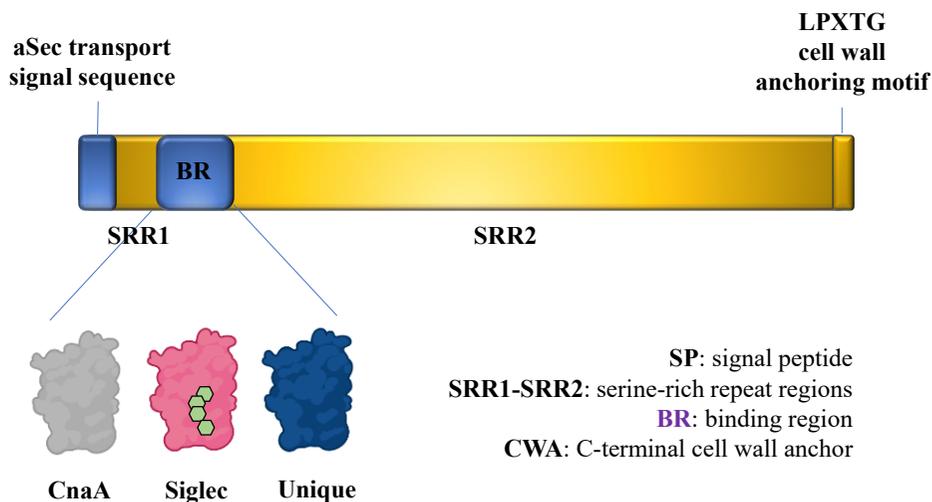


Figure 1.14. Representation of serine-rich repeat glycoproteins (Siglec-like adhesins). SRR adhesins are organized with an N-terminal around 90 amino acid signal peptide (SP), followed by a short serine-rich region (SRR1), a ligand binding region (BR), a long serine-rich repeat region (SRR2), and a C-terminal cell wall anchor (CWA). The “Siglec” and “Unique” domains are involved in the sialoglycans (in green) interaction.

Generally, BR is composed of two conserved domains important for sialoglycan binding: a V-set Ig fold Siglec subdomain, highly similar to that found in mammalian Siglecs in terms of topology and strand inserts (hence the name “Siglec-like” adhesins), and the Unique domain, not directly involved in the interaction with carbohydrates, though possibly modulating the conformation of the nearby Siglec domain. A third domain, called CnaA, can also be present in the serine-rich repeat adhesins (e.g., GspB in *S. gordonii* M99 strain), but it does not contribute to glycan binding.¹¹² The Unique and Siglec subdomains of SRR adhesins play key roles in mediating bacterial recognition of host sialoglycans.^{113,114} In particular, a YTRY consensus sequence, further refined to a Φ TRX motif in the broader family of SLBRs, is present on the F strand of the Siglec domain (figure 1.15), establishing crucial contacts with Neu5Ac- α -(2,3)-Gal containing ligands.

In this thesis, the Siglec-like adhesins SLBR-B and SLBR-H expressed by *S. gordonii* strains M99 and DL1, and SLBR-N from NCTC10712 strain of *S. mitis* have been investigated. Previous analysis on the protein structures (PDB: 5IUC, 6EFD and 6EFF) and chimeragenesis experiments showed that the sialoglycans binding is strongly affected by the protein loops.¹¹⁵

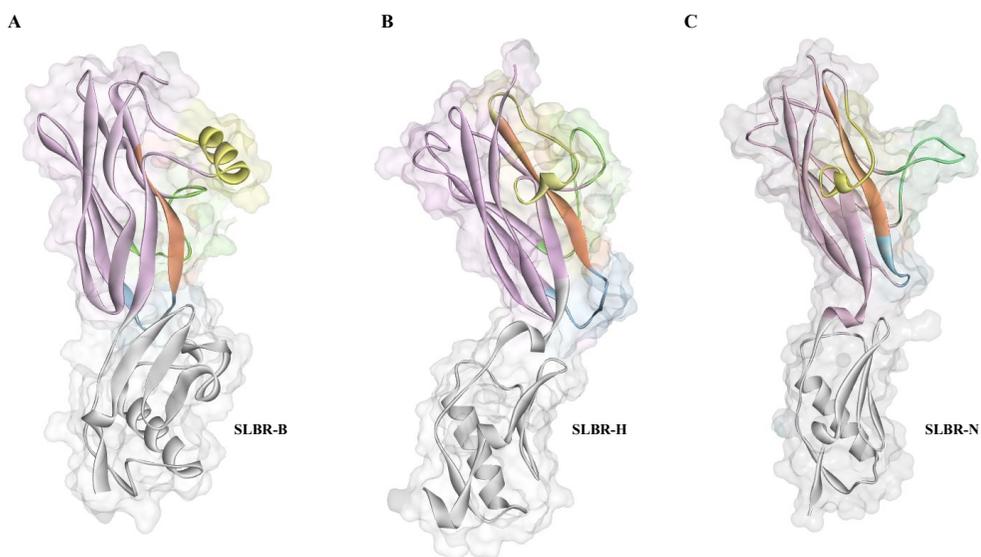


Figure 1.15. 3D structure of the Siglec-like adhesins studied in the thesis: A) SLBR-B, B) SLBR-H, C) SLBR-N. The Siglec and Unique domains were colored in pink and grey, respectively. CD, EF and FG loops were colored in green, blue and yellow, respectively. The F-strand containing YTRY consensus sequence was highlighted in orange.

In particular, CD and FG loops have a fundamental role in the ligand selectivity, while EF loop adjusts the ligand orientation to promote the interactions with the protein. The role of the Siglec-like adhesins in the pathogenesis of infective endocarditis has been widely demonstrated *in vitro* and *in vivo*. Among bacteria found in the oral cavity, *Streptococcus gordonii* and *mitis* are associated to the bloodstream infection and to the progression of IE. Although the moderate sequence identity of the proteins studied, especially between SLBR-H and SLBR-N (80%), the selectivity of their BRs

toward sialoglycan structures is different;¹¹⁶ indeed, SLBR-B strictly recognizes sialyl-T-antigen (sTa), while SLBR-H and SLBR-N bind a repertoire of glycans with different shapes and topologies, including sTa, 3'-sialylactosamine (3'-SLn) and related structures.¹¹⁷ Interestingly, SLBR-N seems to prefer disialylated structures.¹¹⁸ Furthermore, the impact of the Siglec-like adhesins on the virulence of these streptococcal pathogens differs with respect to the bound sialoglycan: for example, strains that bind sialyl-T-antigen are more virulent compared with a strain that binds core 2 *O*-glycans.¹¹² This emphasizes the need for selective inhibition of binding to the former *O*-glycan structure.

1.6 Objectives

Given the fundamental roles of glycan-protein interactions in various biological functions, including cell-adhesion, modulation of immune responses, development of diseases and tumor progression, the investigation of the binding of different human and bacterial proteins to their cognate ligands has been carried out in this thesis by a combination of several techniques.

In the context of the immune regulation, Siglecs have been revealed as key actors for the treatment of inflammatory, autoimmune, and infectious diseases.

The inhibitory *Siglec-2*, expressed on B-cells and involved in tolerance and prevention of autoimmunity, was investigated with Neu5Ac- α -(2,6)-Gal containing carbohydrates, including complex-type *N*-glycans typically found on cell surfaces (Chapter III). The outcomes revealed the structural features for potential design and development of high-affinity ligands to mediate the *Siglec-2* biological functions.

Another inhibitory lectin, the *Siglec-7*, mainly found on NK cells, was investigated in interactions with certain strains from *Fusobacterium nucleatum*, oncogenic pathogen involved in the development of colorectal cancer (Chapter IV). The expression of

glycosylated forms of Siglec-7 in human embryonic kidney (HEK293S) cells and the molecular interactions with *Fusobacterium nucleatum* OPS led the basis for the development of cancer therapeutic approaches targeting *F. nucleatum*-Siglec-7 interaction.

Since adhesins are implicated in the biology of infection, some Siglec-like adhesins, serine-rich repeat glycoproteins expressed on several streptococcal strains and involved in the pathogenesis of infective endocarditis (IE), were studied in interaction with different *N*- and *O*-glycans (Chapters V-VI). The aim was the description of the binding modes at molecular level together with the dynamic range of conformations adopted by the SLBR–sialoglycan complexes. A comparison between Siglecs/Siglec-like adhesins binding sites was also explored. The outcomes provided the basis for the identification of novel therapeutics to prevent or treat IE disease, such as the development of specific inhibitors that do not interfere with Siglecs interactions.

Partially related projects (Chapters VII, Appendix), also investigated during the PhD, involved the study of the interactions between monoclonal antibodies (mAb) against bacterial glycoconjugates (and mimetics). On one hand, the wall teichoic acids (WTA) decorating the cell surface of the Gram-positive *Staphylococcus aureus*; on the other hand, the peptidomimetic of the lipooligosaccharide (LOS) exposed on Gram-negative *Neisseria gonorrhoeae*. Since these bacteria have developed resistance to antibiotic drugs (methicillin for *S. aureus* and ceftriaxone for *N. gonorrhoeae*), different therapies, as the development of vaccines, are urgent for the prevention and treatment of the diseases. The results allowed to define the ligand epitopes crucial for the mAb recognition in order to give the basis for the synthesis of specific and effective targets.

Chapter II:

Unveiling the protein-ligand molecular binding

II. Unveiling the protein-ligand molecular binding

The architecture of protein-ligand 3D complexes in biological systems is a prerequisite in structure-based drug design processes, for developing new therapeutic approaches and strategies. Specific non-covalent interactions in solution are the fundamental basis of molecular recognition processes, characterize protein-ligand interface and contribute to the complex formation. Due to the presence of hydroxyl groups, hydrogen bonds dominate binding forces in carbohydrates. In some cases, OH groups can also act simultaneously as a hydrogen-bond donor and acceptor, resulting in cooperative hydrogen bonding, often found in glycans-lectins complexes. In this regard, a further important polar interaction is the ionic bond between charged residues (salt bridge), as occurs with Siglecs recognizing the carboxylate moiety of sialic acids. On the other hand, carbohydrates are also composed of non-polar patches. Indeed, aliphatic protons of the sugar ring and the presence of glycerol moiety, as in sialic acids, are usually packed against the face of π -electron cloud of aromatic amino acid residues, forming so-called stacking interactions.

Thus, an ensemble of heterogeneous techniques, including biophysical, spectroscopic and computational methods, is required representing powerful tools to unveil the complex interactions occurring at molecular level.¹¹⁹ In this thesis the molecular binding of protein-glycoconjugates has been investigated by means of NMR spectroscopy in combination with biophysical approaches and *in silico* methods, such as docking and molecular dynamics.

2.1 Nuclear Magnetic Resonance NMR spectroscopy

NMR spectroscopy is a useful technique to dissect recognition and binding events, as those occurring in case of protein-glycoconjugate interactions.¹²⁰

The equilibrium of a small-molecule ligand (L) that binds to a large receptor protein (P) to form a complex (PL) follows a bimolecular association reaction with second-order kinetics:



With k_{on} and k_{off} the association and dissociation constants respectively, whose ratio ($k_{\text{off}}/k_{\text{on}}$) determines the rate constant K_D , defined as follows:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad [2.1]$$

Two limiting cases can occur, depending on the K_D value: slow and fast chemical exchange between free and bound forms. In the slow exchange regime (K_D around 10^{-5} M), the lifetime of the complex is longer than the chemical shift difference between the free and bound states, resulting in two NMR signals. Conversely, in the fast exchange regime (K_D around 10^{-8} M), the process is fast compared to the time scale of the chemical shift difference of the two states, and the signals collapse in one single peak that represents the average of the chemical shifts in the free and bound forms. Since some of the NMR techniques depend on the protein-ligand exchange regime, the selection of the appropriate NMR experiment to study a protein-ligand interaction is crucial.¹²¹ Table 2.1 shows the applicability of some NMR techniques here used and described in the following paragraphs.

Table 2.1. Typical range of applicability of the main NMR techniques for the study of protein-ligand interactions.

	K_D [M]	Target MW [kDa]	Typical protein/ligand ratio	Labeled target required	Target binding site	Ligand epitope mapping	Ligand selectivity in a mixture
Tr-NOE	$10^{-6} - 10^{-3}$	No limit	1:5/1:50	No			✓
STD NMR	$10^{-6} - 10^{-3}$	> 15	1:50/1:200	No		✓	✓
WaterLOGSY	$10^{-6} - 10^{-3}$	No limit	1.5/1:50	No		✓	✓
Diffusion experiments	$10^{-6} - 10^{-3}$	No limit	1:1/1:20	No		✓	✓
CSP	$10^{-9} - 10^{-3}$	< 100	1:1/1:10	Yes	✓		

Most of the screening methods based on nuclear magnetic resonance (NMR) are effective in the identification of small molecules interacting with macromolecular receptors. Screening may proceed by ligand- or protein-based methods.

2.1.1 Ligand-based NMR approaches

Besides the ligand-based NMR techniques used for protein-ligand interactions, the ligand NMR assignment is required and can be achieved combining mono- and bidimensional NMR spectra. ^1H and ^{13}C NMR experiments are useful to define the monosaccharide composition, the α or β anomeric configuration, the substitution pattern, the nature of non-glycidic substituents. In particular, ^1H NMR experiments give information about signal *multiplicity*, scalar interactions between vicinal and geminal nuclei propagated through the bond electrons, measured by the coupling constant (J), a value used to gain important structural information.

Table 2.2. Typical ^1H and ^{13}C chemical shift values of sugar compounds.

δ (ppm)	^1H
8.5 – 7.5	Amide 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 the 6-deoxy sugars and of the acetyl groups
δ (ppm)	^{13}C
160 – 180	Carbonyl carbons
95 – 105	Anomeric carbons
60 – 80	Sugar ring carbons
45 – 60	Nitrogen bearing carbon signals
~ 30	Aliphatic methylene carbons of deoxy sugars
20 – 17	Methyl carbons of deoxy sugars, acetyl groups

For example, a $^3J_{H_1,H_2}$ above 8 Hz is indicative of a β -configured pyranose ring with *gluco*- or *galacto*-configuration; $^3J_{H_1,H_2}$ below 3 Hz is diagnostic of an α -configured sugar. The magnitude of $^1J_{C_1,H_1}$ is diagnostic of the anomeric configuration and is a method applicable to a plethora of sugar moieties (below 170 Hz indicative of a β -anomer and above 170 Hz of an α -anomer). The 1H and ^{13}C typical chemical shift (δ) regions are listed in the table below (table 2.2).

To build the sugar sequence, a series of 2D NMR experiments is required. Homonuclear COSY (Correlation Spectroscopy) and TOCSY (Total Correlation Spectroscopy) correlate chemical shifts (resonances) of 1H nuclei with geminal and vicinal couplings and those belonging to the entire spin network, respectively. NOESY (Nuclear Overhauser Effect Spectroscopy) measures the spins cross-relaxation rates and reveals which protons are close to each other in space; the corresponding NOESY measured under spin-locked conditions is called ROESY (Rotating-frame Overhauser Effect Spectroscopy), that always gives a positive signal, and is often used when NOE is close to zero (e.g., for small oligosaccharides, as in many trisaccharides). Regarding the carbon-proton correlation, HSQC (Heteronuclear Single Quantum Correlation) correlates directly couple ^{13}C and 1H , HMBC (Heteronuclear Multiple Bond Correlation) utilizes multiple-bond couplings over two or three bonds ($J=2-15$ Hz) for determining long-range 1H - ^{13}C connectivity.

Following ligand assignment, ligand-based NMR methods are performed. It is worth knowing that this approach renders the molecular weight of the receptor molecule irrelevant. However, ligand-based NMR experiments rely on the exchange-mediated transfer of bound state information to the free state. Thus, the requisite to perform ligand-based methods is the fast protein-ligand exchange regime, with $K_D \geq 100 \mu M$ in the medium–low affinity range, dissociation rate constant in the range $1000 < k_{off} < 100\,000 s^{-1}$ and the use of large ligand molar excesses.¹²²

2.1.2 Transferred-NOESY

NOE effects (NOEs) are awfully practical for the determination of 3D structure of molecules in solution. The transferred NOE effect is based on the different behavior of a ligand in the free and bound states and allows to detect and characterize ligands' binding. Low-medium molecular weight molecules (e.g., ligands lower than 2 KDa) exhibit fast tumbling in solution and short correlation times τ_c and can assume positive NOE, no NOE or small negative NOE, depending on MW, experimental conditions and strength of magnetic field. Conversely, high molecular weight molecules, such as proteins, have slow tumbling in solution and long correlation times τ_c , showing negative NOE (Figure 2.1).

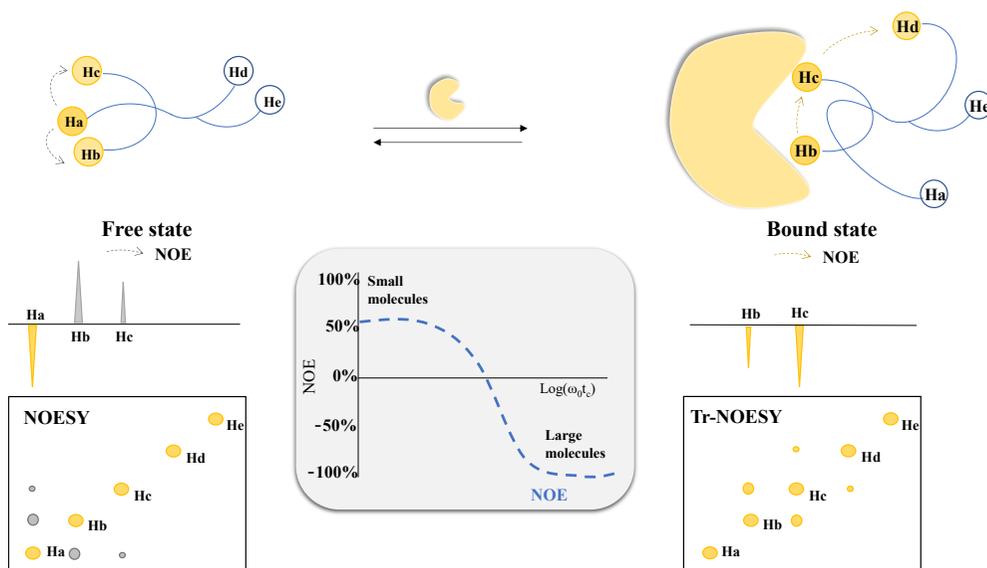


Figure 2.1. Schematic representation of NOE effects. In the free state, small molecules exhibit positive NOEs (cross-peaks with opposite sign to the diagonal peaks in the NOESY spectrum); in the bound state, small molecules adopt negative NOEs, behaving as the large protein, as shown in the tr-NOESY spectrum.

Interestingly, a small molecule that binds to a receptor protein behaves as part of the macromolecule, adopting the corresponding NOE, called transferred NOE (tr-NOE). The ligand, freely and rapidly exchanging between the bound and free forms, retains the NMR properties of the protein and stores information on the bound state.¹²³

The discrimination of NOEs between the free and bound states can be also achieved by the build-up rate, that is the time required to achieve maximum intensity of the NOE. The maximum NOE is a function of molecular tumbling rates, defined by $\omega_0\tau_c$, with ω_0 being the spectrometer observation frequency and τ_c the rotational correlation time, connected to the molecular size. The maximum enhancement for tr-NOEs is observed at significantly shorter mixing times (τ_{mix} in range of 50 to 100 ms) than for unbound ligands (four- to ten-times as longer).

The analysis of NOE-derived *inter*-protons distances allows to detect conformational changes of the ligand upon binding and to obtain the so-called *bioactive conformation* (the conformation adopted by ligand in the bound state).¹²⁴

The construction of NOE build-up curves accurately determines ¹H-¹H nuclear distances. Upon integration of the NOEs of spectra acquired at different mixing times, the build-up curves are fitted to a double exponential function:

$$f = a(e^{-ct})(1 - e^{-bt}) \quad [2.2]$$

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

$$f(0) = a \times b \quad [2.3]$$

From the initial slope, the *inter*-proton distances are derived by using the isolated spin pair approximation:

$$r_{ij} = r_{ref} \sqrt[6]{\frac{\sigma_{ref}}{\sigma_{ij}}} \quad [2.4]$$

where r_{ij} is the distance to calculate, r_{ref} is a known distance used as reference, σ_{ref} is the cross-relaxation rate, and σ_{ij} is the cross-relaxation time that gives the desired distances.

2.1.3 Saturation Transfer Difference

Saturation transfer difference (STD) NMR spectroscopy is a powerful NMR technique for the detection and characterization of transient receptor–ligand interactions in solution.^{125,126} Based on the magnetization transfer from the protein to the ligand protons by spin diffusion and intermolecular NOE, STD NMR allows to verify the occurrence of protein-ligand interaction and investigate the binding process at molecular level, by deriving the ligand epitope map. STD NMR is acquired as a pseudo-2D experiment, resulting from the subtraction of two mono-dimensional NMR spectra: 1) the *off-resonance*, where a region far from the protein and ligand signals (usually at around 40 ppm) is irradiated to detect the reference spectrum, and 2) the *on-resonance*, where the protein signals are saturated (usually in the aromatic or aliphatic spectral region) by applying a selective low power radio frequency-pulse train in the range of seconds (*saturation time*). The saturation is therefore transferred from the receptor to the interacting ligand during its residence time in the protein binding pocket by intermolecular saturation transfer and fast chemical exchange. In the on-resonance, the enthalpic relaxation (R_1) of the free ligand is slower than the k_{off} , an accumulation of saturated free ligand in the bulk solution occurs, and protons that receive magnetization from the receptor decrease their signals intensities. Since the STD is the subtraction of the off- and on-resonance spectra, ligand protons closer to the protein binding site will receive a higher degree of magnetization by intermolecular ^1H - ^1H cross relaxation pathways, showing higher STD signals; on the other hand, ligand protons farer from the protein will receive little or no saturation,

resulting in lower or in absence on STD signals (figure 2.2). The STD intensities (I_{STD}) are calculated as follows:

$$I_{STD} = \frac{I_0 - I_{sat}}{I_0} \quad [2.5]$$

where I_0 is the intensity of a signal in the off-resonance experiment, I_{sat} is the intensity of the signal in the on-resonance experiment. Depending on the degree of saturation of the protons, it is possible to map the interacting epitope of the ligand.¹²⁷ Once set the most intense STD signal as 100%, all the other protons are normalized to obtain STD% values.

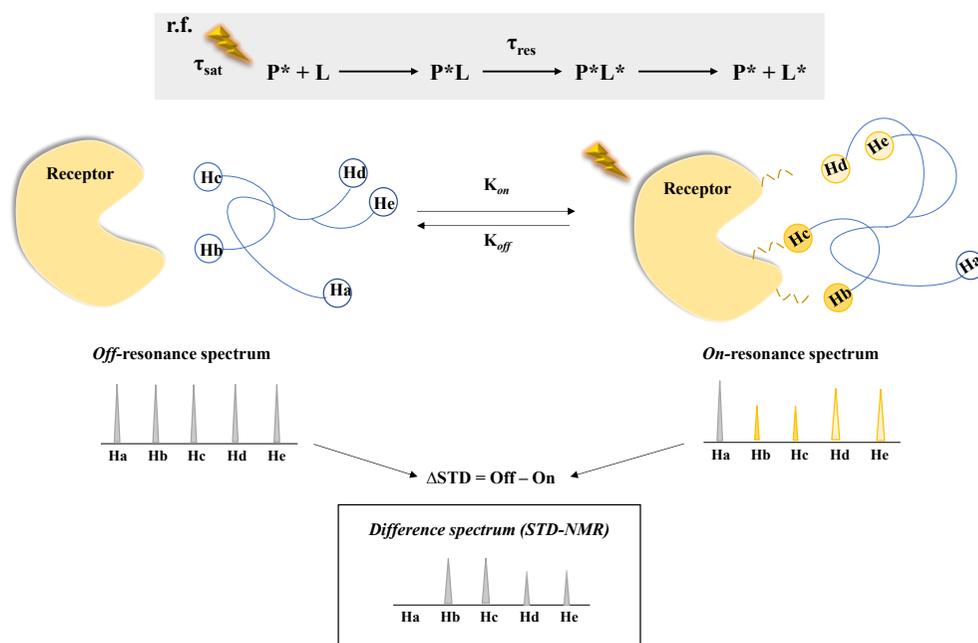


Figure 2.2. Schematic representation of STD NMR technique. In the off-resonance spectrum, acquired by irradiating far from both protein and ligand signals, ligand signals (Ha-Hb-Hc-Hd-He) do not show a decrease in their intensities. In the on-resonance spectrum, the receptor is selectively saturated with RF pulses and the magnetization is transferred to the ligand

protons by intermolecular NOE. Only the ligand signals involved in the binding process will show STD signals in the spectrum.

To overcome possible artifacts correlated to the relaxation times, STD NMR spectra can be recorded at different saturation times (t_{sat} is usually from 0.5 s to 5 s) to obtain the STD build-up curves.

The initial growth rates are obtained deriving the STD intensities close at the limit of zero saturation time, where no ligand re-binding or relaxation takes place. STD build-up curves are calculated from the STD amplification factor (A_{STD}) following the equation:

$$A_{\text{STD}} = \frac{I_0 - I_{\text{sat}}}{I_0} * \frac{[L]_0}{[P]_0} = \frac{I_{\text{STD}}}{I_0} * \frac{[L]_0}{[P]_0} \quad [2.6]$$

where I_{STD}/I_0 is the relative STD effect at total ligand ($[L]_0$) and protein $[P]_0$ concentrations. A_{STD} is calculated for each proton involved in the interaction at each saturation time. Data are fitted to the following mono-exponential function:¹²⁸

$$\text{STD}(t_{\text{sat}}) = \text{STD}_{\text{max}} * (1 - \exp(-k_{\text{sat}} * t_{\text{sat}})) \quad [2.7]$$

where $\text{STD}(t_{\text{sat}})$ is the observed STD intensity, STD_{max} is the asymptotic maximum of the build-up curve, t_{sat} is the saturation time, and k_{sat} is the rate constant related to the relaxation properties of a given proton that measures the speed of the STD build-up. The parameter STD_{fit} represents the slope of the STD build-up curves when saturation time is 0 and depends on the proximity of the ligand to the protein. To obtain the epitope map of the ligand, the STD_{fit} values are normalized with respect to the highest one, set to 100%.

Therefore, STD NMR is a useful method to get information on the binding epitope of the ligand in interaction with a macromolecule, of primary importance, for example, in the development of drugs and/or mimetics. Worthy, sample containing a low concentration of macromolecule (in the μM range) is sufficient, but a large molar

excess of ligand is required (typically from 1:10 up to 1:1000). As mentioned above, STD NMR technique is only applicable to protein-ligand systems in fast exchange, with a medium-weak affinity to the receptors, exhibiting a dissociation constant generally in the millimolar to micromolar range (K_D 10^{-6} – 10^{-3} M).

2.1.4 WaterLOGSY

Water molecules play as mediators of hydrogen bonds between sugar and protein, especially in lectin-carbohydrate complex structures. Water sites are defined as space regions close to the receptor surface where the probability to find a water molecule is significantly higher than in the bulk solvent.¹²⁹ Water molecules can form hydrogen bonds with both protein and ligand, sometimes remaining at their interface as fixed structural elements. Therefore, important information can be obtained performing hydration NMR experiments, by using the WaterLOGSY (Water-Ligand Observation with Gradient Spectroscopy) technique, where the source magnetization originates from bulk solvent (H_2O) protons instead of target receptor (figure 2.3). Being a NOE-based experiment, the fast-exchange regime is required to describe protein-ligand binding events via bulk water.¹³⁰

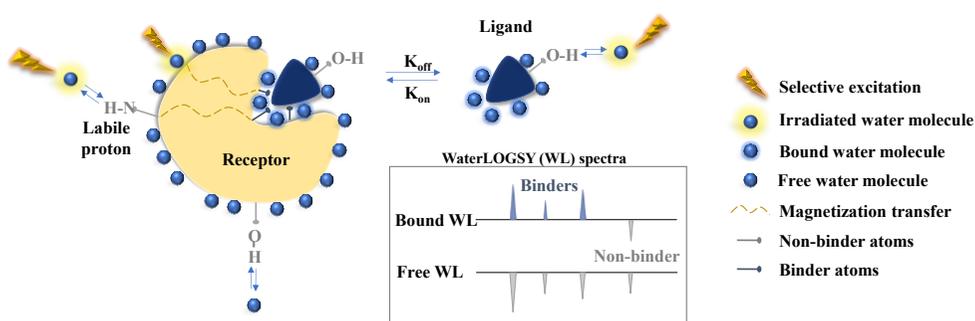


Figure 2.3. Schematic representation of WaterLOGSY principle. Water molecules, surrounding the receptor surface or in chemically exchange with the protein groups, are selectively irradiated; then, the magnetization is transferred from the receptor to the ligand, which interacting protons will assume the NOE properties of the macromolecule. The water

suppression is carried out by either excitation sculpting with gradient suppression scheme or WATERGATE, hard pulse water suppression by gradient tailored excitation.

The WaterLOGSY is conventionally acquired as 1D NOE-ePHOGSY pulse sequence, based on three steps: 1) the selective excitation of the bulk water with the inversion of resonance, 2) the magnetization transfer via NOE and spin diffusion (NOE mixing time) and 3) the water suppression.¹³¹ Conventionally signals in the WaterLOGSY spectra are “upside down” phased: ligand in the free state or non-binder molecules show negative signals, while binder resonances will appear as positive in the spectrum. However, depending on the protein-ligand exchange regime, binders can remain negative in the WaterLOGSY spectrum, only decreasing their intensity. Thus, to avoid artifacts is necessary the acquisition of a WaterLOGSY spectrum of the ligand in absence of the protein (as reference) and is recommendable to perform the experiments by using a mixing time in the range of 1-3 s. Comparing the ligand signals in the free and bound states, resonances that undergo changes in sign or relative intensities are indicative of presence of resident water molecules in the protein-ligand complex.

2.1.5 Other NMR techniques

Relaxation and diffusion can also provide further information on the protein-ligand complexes in solution. Several conditions influence these parameters, such as molecular size and shape as well as temperature, viscosity of the solvent and strength of the NMR magnetic field.

A relaxation-based NMR method studied in this thesis is the Carr–Purcell–Meiboom–Gill (CPMG) experiment.¹³² When a small ligand interacts with a macromolecule, a decrease of T_2 relaxation time occurs, resulting in an

observation of a line broadening and lower signal intensities in the CPMG spectrum. Thus, this phenomenon is indicative of the complex formation (figure 2.4).

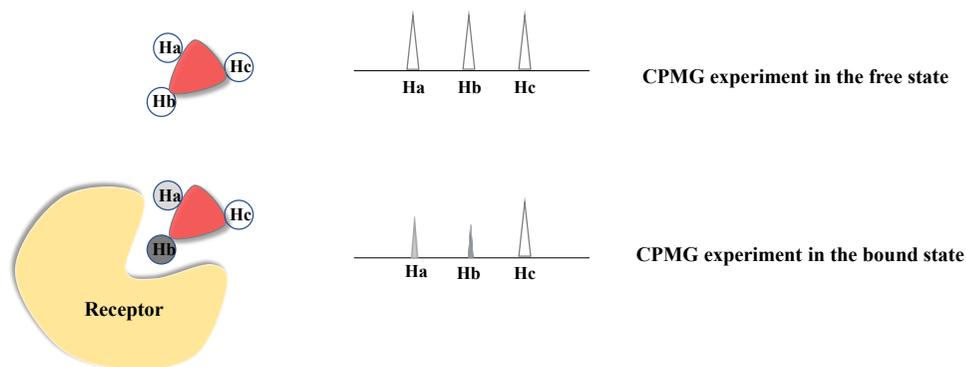


Figure 2.4. Schematic representation of Carr Purcell Meiboom Gill (CPMG) experiment. T_2 relaxation measurements are carried out for the ligand in absence and in presence of the protein. The ligand protons that bind to the receptor will assume the relaxation properties of the macromolecule, resulting in a decrease of signal intensities.

Information on the potential protein-ligand complex formation can be also achieved by measuring the diffusion properties of molecules in solution, by means of their Brownian motions. The NMR technique used in this thesis is called DOSY (Diffusion Ordered SpectroscopY). The experiment is performed using a sequence based on short-time Pulsed-Field Gradient (PFG) that generates local variation of the magnetic field, allowing molecules to be spatially labeled in the NMR tube.¹³³ The diffusion coefficients of the molecules can be calculated following the Stokes-Einstein equation:

$$D = k_b T / (6\pi\eta r_h) \quad D = \frac{k_b T}{6\pi\eta r_h} \quad [2.8]$$

where D represents the diffusion coefficient (m^2/s), k_b is the Boltzmann constant, T is the temperature, η is the viscosity of the solvent and r_h is the hydrodynamic radius.

Differently from the previous ligand-based NMR techniques, the protein-ligand ratio is low, usually going from 1:1 to 1:10. Before the acquisition of 2D DOSY NMR experiments, diffusion time and gradient parameters must be optimized depending on the system. Indeed, these parameters affect the diffusion decay and the attenuation of NMR signal intensity (figure 7 A), defined as follows:

$$I = I_0 e^{-D\gamma^2 g^2 \delta^2 (\Delta - \delta/3)} \quad [2.9]$$

where I represents the NMR signal intensity, I_0 is the reference intensity, D is the diffusion coefficient, γ is the gyromagnetic ratio and g is the gradient strength.

The analysis of the binding is then monitored by the calculation of all diffusion constants of each species in solution. Thus, a variation in the diffusion coefficient values observed passing from the free ligand to the bound state will indicate changes in the diffusion properties and may suggest the formation of protein-ligand interaction (figure 2.5).¹³⁴

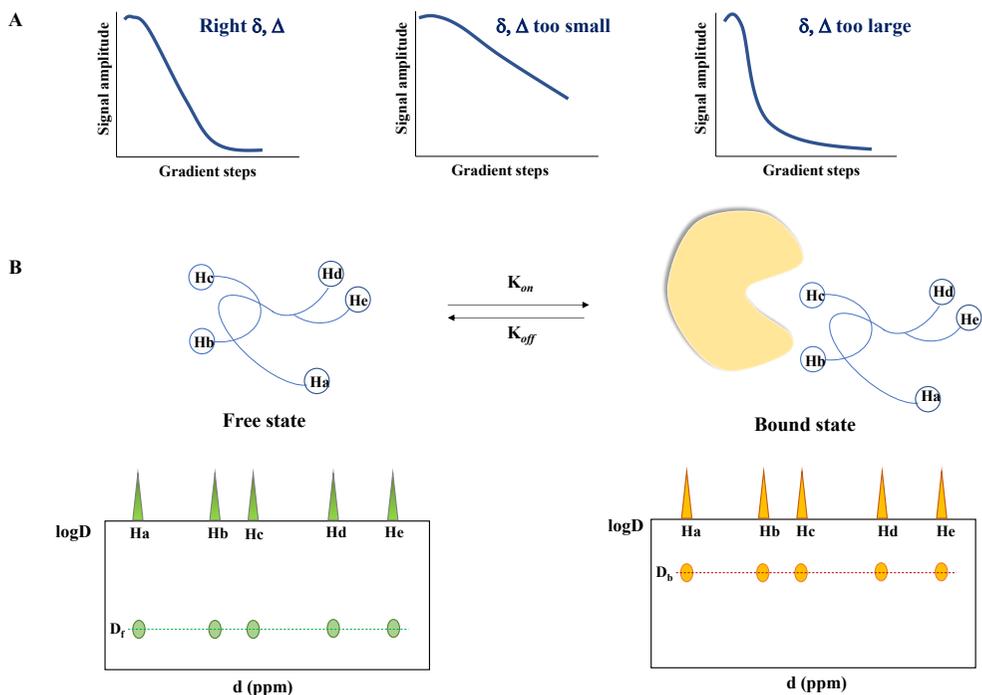


Figure 2.5. Schematic representation of Diffusion Ordered Spectroscopy (DOSY) experiment. A) Measurement of signal decay for the optimization of gradients and diffusion time parameters prior the acquisition of the 2D DOSY experiments. B) Example of 2D DOSY NMR spectra in the free and bound states, showing differences in the diffusion coefficient and properties.

2.2 Protein-based NMR approaches

Protein-based NMR methods permit the direct observation of receptor signals to characterize the interactions with cognate ligands, detecting the receptor binding site, when unknown.¹³⁵ For the acquisition of these NMR experiments, a previous assignment of the protein NMR resonances is required. A soluble and non-aggregated isotope-labelled protein (*e.g.* ^{13}C , ^{15}N , ^2H) must be expressed and purified and then

characterized by specific 3D NMR experiments for the assignment of protein backbone and lateral chains.¹³⁶

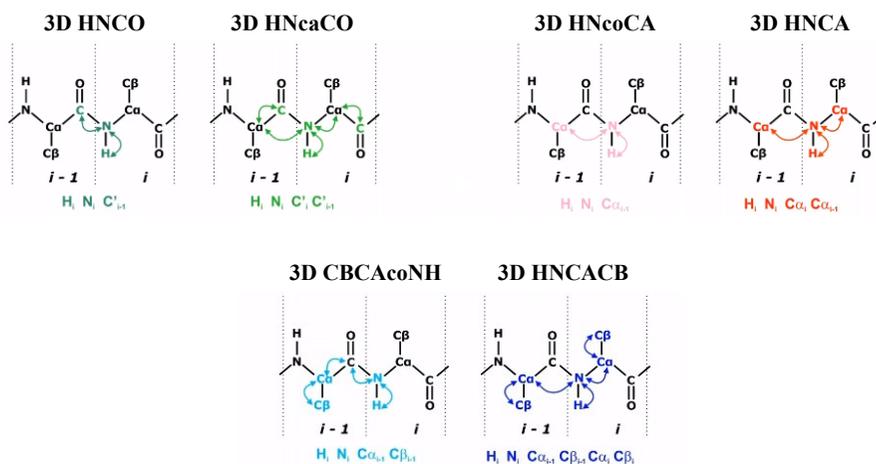


Figure 2.6. Representation of 3D NMR experiment used for the backbone resonances assignment. Each 3D experiment correlates HN resonances of each amino acid to the chemical shifts of CO, C α and C β belonging to the same and preceding amino acid.

For this reason, the limit of the protein-based technique is the macromolecule size (typically < 50 kDa). Since the molecular weight impacts the relaxation times, thus resulting in low quality ^{15}N -HSQC spectra for large macromolecules, several methodologies have been developed in the last decades. The minimization of signals overlap can be accomplished by the reduction of spin-spin relaxation, by using the triple labelling protein expression (^{15}N , ^{13}C , ^2H), where deuteration considerably extends the T_2 relaxation times, allowing narrower lines in NMR spectra. Another solution is the TROSY (Transverse Relaxation Optimized Spectroscopy) implementation of triple resonance experiments, that selects a single component of different relaxation T_2 mechanisms (due to the dipole-dipole mechanism and chemical shift anisotropy) leading to a single and sharp peak in the spectrum.¹³⁷In this thesis, 3D NMR experiments were used for the protein backbone assignment, as depicted in

figure 2.6. Spectra were analyzed by using CARA (Computer Aided Resonance Assignment) software.¹³⁸ The program allows to identify each resonance and associate it to a nuclear spin (*peak peaking*) and assign each spin system to a specific residue of the protein sequence (*matching*).

2.2.1 Chemical Shift Perturbation

Chemical shift perturbation (CSP) or chemical shift mapping (CSM) is the most common protein-based NMR experiment, which is used to map the chemical shift of a ^{15}N labelled protein when it is titrated with a ligand.^{139,140} Indeed, if the interaction occurs, the chemical shifts of the amino acids involved in the complex formation with ligand are subjected to perturbation. CSP analysis of a protein/ligand complex requires the acquisition of the ^{15}N -HSQC spectrum for the apo protein as reference. Then, sequential ^{15}N -HSQC spectra are performed upon addition of increasing ligand concentrations, ideally until the protein binding site is completely saturated.

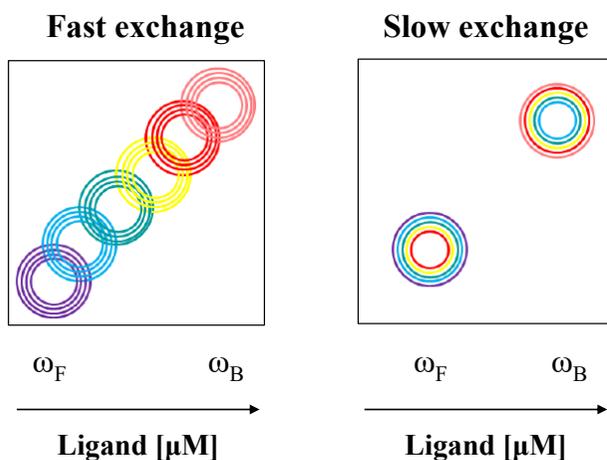


Figure 2.7. Chemical shift perturbation in protein-based NMR experiments. By adding amounts of ligand to the protein, the chemical shift can vary in the ^{15}N HSQC spectrum, depending on the protein-ligand exchange regime.

A fundamental prerequisite for chemical shift perturbation analysis is that the protein and the ligand must be dissolved in the same buffer and the measurements during the titration must be acquired under the same conditions. This is important because chemical shifts are very sensitive to differences in temperature, pH value and buffer composition, especially those of amide protons.³

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

Although most of the peaks affected by a variation of the chemical shift determines the binding site of the protein that accommodates the ligand, conformational changes of amino acids also lead to differences in resonance frequencies. Thus, the shifting of a signal is not always indicative of the vicinity to the binding interface, but it could give information about allosteric changes in the protein structure when a ligand is bound. This kind of conformational change usually happens to protein residues that are buried in the 3D structure or located far from the binding pocket.

2.3 Computational methods

Molecular modelling encompasses all computational methods used to model or mimic the behavior of molecules. 3D complexes have been investigated using computational approaches combined to experimental techniques of structural biology, such as NMR spectroscopy, X-rays crystallography and cryo-EM. Additionally, *in silico* methods

can elucidate the structural characteristics of molecules when experimental data are missing. Here, molecular docking and dynamic simulations were used.

2.3.1 Docking

In the field of molecular modelling, molecular docking is a method widely used for the structure-based drug design, thanks to the ability to predict the binding-conformation of small molecule into appropriate protein binding site, providing a model of the interaction. In general, computational docking protocols allow to find the possible binding poses of a small molecule within a particular receptor, adjusting their conformation to achieve the “best-fit”. The second step is the energy scoring of the resulting binding poses, so the energy evaluation of the ligand-target complex. Although these two steps are common in all the programs available for docking calculations,¹⁴¹ the main differences rely on the algorithm used for the computational search and the nature of the scoring function applied to rank the docked poses.^{142,143} Moreover, each program diverges for the maximum number of rotatable bonds of ligand, and therefore entails a different accuracy and computational cost. Three types of algorithms for the conformational ligand search can be chosen: shape matching, systematic search and stochastic algorithms.

Shape matching algorithms consider the geometrical overlap between two molecules, identifying the possible binding sites of a protein by a macromolecular surface search.¹⁴⁴ Systematic search algorithms are usually used for flexible-ligand docking, and all possible binding conformations are generated by exploring all degrees of freedom of the ligand. The most time efficient are the stochastic algorithms, such as Monte Carlo (MC) method and evolutionary programming (EP) where random changes in the ligand are executed.¹⁴⁵

Regarding the energy evaluation, three scoring functions can be classified: force-field based, knowledge-based and empirical-based scoring functions. Force-field based

scoring functions employ non-bonded terms of classical force fields to compute the direct interaction energies, considering van der Waals and electro-static energies as well as stretching, bending, and torsional energies. Knowledge-based scoring functions are based on the statistical analysis of interacting contacts from protein–ligand complexes identified from structure databases. Empirical scoring functions make use of several intermolecular interaction terms, such as hydrogen bond and hydrophobic interactions, for pose and affinity prediction so that fitted theoretical values are as close as possible to experimental data.

To perform a docking calculation, the 3D structure of the protein of interest is needed, and it can be provided by experimental techniques, such as X-ray crystallography or NMR spectroscopy, or it can derive from homology modelling methods. Another important, but not mandatory, requisite is the information about the protein binding site. In Autodock4¹⁴⁶ used in this thesis, information on the binding site helps to build the AutoGrid grid box, that is the grid volume where the ligand rotates freely, even in its most fully extended conformation. If the binding pocket is unknown, a grid volume big enough to cover the entire protein surface is necessary, resulting in more computational costs. In this cases, preliminary docking experiments could be performed to investigate if some regions of the protein are preferred by the ligand, in a process known as “blind docking”. In this way, a second round of docking calculations allows to choose a smaller grid where ligand can move around the protein. In Autodock program, evolutionary programming (EP) algorithms are considered, where computational models are treated as evolutionary biological processes. In particular, Autodock uses the Lamarckian Genetic Algorithm,¹⁴⁷ an EP algorithm based on the combination of the genetic algorithm (GA) and local search (LS) method that tries to find the closest conformation of the global energy minimum.¹⁴⁸ In this hybrid method, the ligand variables (translation, orientation, and conformation) are considered as a “gene” that characterize a “genotype” (ligand’s state), while the

atomic coordinates correspond to the “phenotype”. Genotypic space is characterized by mutation and crossover, whereas phenotypic space is established by the energy function to be optimized. Individual conformations search their local minima, inheriting the information to later generations (Lamarckian aspect).

AutoDock, that uses a Lamarckian Genetic Algorithm and semi-empirical free energy force field to score docked binding poses of small molecules to macromolecular targets, creates a set of docked conformation with the sum of intermolecular and internal energy components. The conformation with the lowest docked energy could be considered as ‘best’ docking result.

2.3.2 Molecular Mechanics and Dynamics simulation

Molecular mechanics and dynamics simulation are methods to calculate the structure and energy of molecules based on nuclear motions. According to the Born Oppenheimer approximation, nuclei are considered much heavier than electrons, moving much more slowly. In this assumption, atoms, characterized by a mass and size (van der Waals radius), are represented as hard and impenetrable spheres, covalently connected by elastic bonds (stretch, bend, torsion), according to Hooke’s law and characterized by non-bonded interactions (van der Waals and electrostatics). The sum of these energies defines the potential energy surface of a molecule. The atomic interactions can be modelled with simple parameterized functions determining molecule’s potential energy and geometry, called *force fields*, that can be obtained either from *ab initio* or semi-empirical quantum mechanics (QM) calculations or by fitting of experimental data. Thus, the choice of the force field is the core of the validity of any Molecular Mechanics and Molecular Dynamics simulation.

For carbohydrate structures, the major force fields used are Amber/GLYCAM¹⁴⁹ and CHARMM.¹⁵⁰ Both force fields cover the majority of common monosaccharides and contain different glycosidic linkages. GLYCAM06 is suitable for D and L

enantiomers, mono- and oligosaccharides, and for all glycosidic linkages, including *N*-glycosylation. It is the only force field in which the same atom type (Cg) at the anomeric carbon (C1) is assigned in both anomers, α and β , facilitating the simulation of ring-flipping and having equilibrium between conformers with axial and equatorial substituents at the anomeric center.

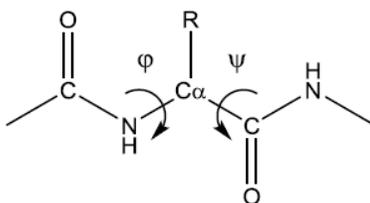


Figure 2.8. Representation of a protein backbone, with the dihedral angles which are defined in Amber ff14SB force field.

Regarding proteins, the force field mostly used is Amber ff14SB¹⁵¹, where parameters are usually calculated by *ab initio* methods¹⁵² and then fitted and validated with experimental results. Protein backbone, for example, is represented as two dihedral angles, φ (C–N–C α –C) and ψ (N–C α –C–N), both defined in the force field (figure 2.8).

2.3.2.1 Molecular mechanics of sugars

The 3D structure of carbohydrates is characterized by the sugar composition and glycosidic linkage, parameters that allow to describe shape and conformation. Ring shapes can be defined in terms of reference conformations (chair, C, twist, T, boat, B, envelope, F, skew, S). The sugar conformation is mainly described by glycosidic torsion angles Φ (H₁–C₁–O₁–C_x) and Ψ (C₁–O₁–C_x–H_x). However, when the glycosidic bond does not involve an endocyclic carbon (atom not located in the ring), such as for 1,6 linkages, the ω (O5–C5–C6–O6) angle is considered (figure 2.9). Sampling of the

ω torsion angle was described by means of the populations of the gauche-gauche (*gg*), gauche-trans (*gt*), and trans-gauche (*tg*) rotamers.¹⁵³

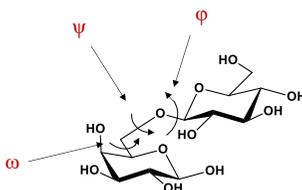


Figure 2.9. Representation of a sugar torsion angles.

The additional flexibility of the α -(1 \rightarrow 6)-linkages makes it more challenging to determine the preferential conformation in solution of oligosaccharides containing these linkages.³¹ The three cases of staggered rotamers in Newman projections are shown in Figure 2.10.

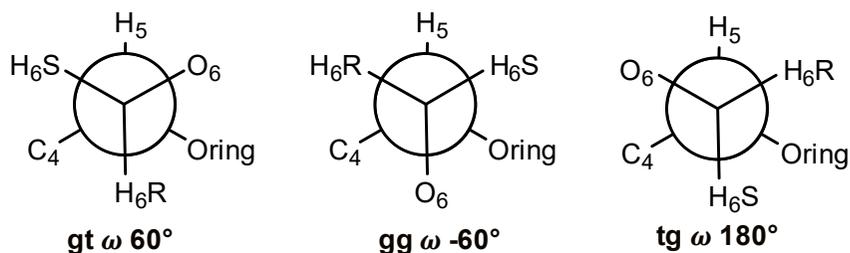


Figure 2.10 Newman projections of staggered rotameric states of the omega torsion angle along the C6-C5 bond: *gg* (gauche-gauche), *gt* (gauche-trans) and *tg* (trans-gauche).

Various methods exist for the calculation of Φ and Ψ to define the energy map, generally applied to each disaccharide unit in a glycan chain.¹⁵⁴ MM3 is the main force field used to calculate local minima and flexibility of glycosidic torsions,^{155,156} but free carbohydrate databases (such as <http://glycosciences.de>) are also available. Thus, the potential energy surface showing conformational energy for the Φ and Ψ dihedral angles can be plotted on “adiabatic” maps, representing energy graphs similar to the

Ramachandran plots used for proteins, that show the energetically favorable conformations of a carbohydrate dimer.¹⁵⁷

2.3.2.2 Molecular dynamics

Molecular dynamic simulation (MD) is one of the main tools in the computational study of biological molecules, that provides detailed information on the dynamics and conformational changes of molecules and their complexes.

In a MD simulation, the initial velocities and positions of all atoms in the biomolecular system are set up. The forces acting on each atom are calculated and, following the Newton's second law of motion, the spatial position of each atom is predicted as a function of time.¹⁵⁸ Given the atom "i" with mass "m_i" and cartesian position "x_i", F_{xi} represents the force acting on the atom during the time t:

$$\frac{d^2x_i}{dt^2} = \frac{F_{xi}}{m_i} \quad [2.10]$$

The *iter* is repeated for defined time intervals and the position and velocity of each atom in each step is evaluated, producing a trajectory that describes the atomic-level configuration of the system at every point over the time.¹⁵⁹ The first step of a MD simulation is the energy minimization. A common method is the steepest gradient in which geometry optimization of the system is executed until reaching the local minimum.¹⁶⁰ During this phase, the best approach is the inclusion of the environment in the simulation, for example considering the MD with explicit water molecules or other surrounding molecules. In this context, due to the presence of hydroxyl groups that establish several hydrogen bonds, carbohydrates have a very high affinity towards water molecules. By using explicit water, the simulation of complex occurs within a box of solvent molecules, under periodic boundary conditions (PBC) to avoid surface artefacts. Among the models that take into account of water molecules, the most used is the TIP3P.¹⁶¹ The infinite electrostatic interactions are calculated by using particle

mesh Ewald summation (PME) in which the summation into short- and long-range parts is split. The second step is based on the heating phase to remove the unfavorable contacts between solvent and solute, such as steric clashes. Here, the velocity of the atoms is increased and calculated with standard temperature-dependent Maxwell-Boltzmann distribution. The following step is the equilibration. It consists of the system relaxation under controlled energy, temperature, pressure and volume conditions. The last step is the production of the final trajectories.

Molecular dynamics simulations represent a promising tool for the investigation of carbohydrates, in free and bound states, and are usually used in combination with NMR data (*e.g.* NOE and residual dipolar coupling-based experiments) to describe in detail the conformational behavior of ligands as well as to build validated 3D models of complexes.

2.3.3 CORCEMA-ST program

CORCEMA-ST (Complete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) is a software that allows quantitative analysis of saturation transfer difference NMR (STD-NMR) data.¹⁶² This tool is based on a modification of the CORCEMA theory and enables the prediction of STD intensities from the Cartesian atomic coordinates of the ligand–receptor complex.¹⁶³ CORCEMA-ST can predict theoretical ligand STD intensities from a given molecular model of the protein–ligand complex, if specific system properties, such as dissociation constant, k_{off} , rotational correlation times of the receptor and ligand, are known. In this way, CORCEMA-ST results a valuable program for the quantitative structural interpretation of experimental STD NMR data.

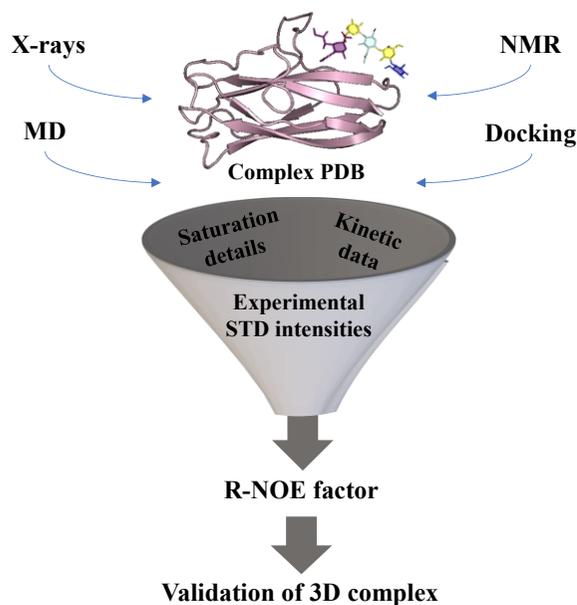


Figure 2.11. Schematic representation of CORCEMA-ST protocol for the determination of theoretical STD and the validation of 3D protein-ligand complexes.

The comparison between experimental STD build-up curves with theoretical ones allows predictions for a model of the complex that could be obtained by use of different techniques (e.g., X-ray crystallography, NMR, docking simulations). The method uses the matrix calculations, in which the Cartesian coordinates of all the protons of ligand and protein are considered within a given cut-off distance. The quality of molecular model that reproduces the experimental NMR data can be quantified by the so-called R-NOE factor. Given the proton k , R-NOE is calculated as follows:

$$R - NOE = \sqrt{\frac{\sigma W_k (STD_{exp,k} - STD_{cal,k})^2}{\sigma W_k (STD_{exp,k})^2}} \quad [2.11]$$

where $STD_{exp,k}$ and $STD_{cal,k}$ are the experimental and calculated STD intensities, respectively. A good fit between experimental and theoretical data is achieved at low R-NOE values, meaning a good validation of the 3D complex. Thus, different structural models, e.g., from different docking runs or clusters of MD, can be ranked according to how well they explain the experimental STD NMR data in order to obtain the best model. The R-NOE factor can be used as a scoring function to drive a conformational search for the ligand bound in the protein binding site. Moreover, after the identification of a good starting pose, the R-NOE value can be minimized by optimizing some key torsion angles via simulated annealing in order to find the global energy minimum of the ligand bound to the receptor. Another procedure is the refinement of the ligand geometry into the binding site by experimental STD data.

Therefore, the ensemble of experimental techniques, as those mentioned above, with computational approaches, such as molecular docking, dynamics simulations and CORCEMA-ST has been considered useful for the determination of the 3D structures of ligand-receptor complexes.

Chapter III:

Unveiling the murine and human CD22 recognition towards sialoglycans

III. Unveiling the murine and human CD22 recognition towards sialoglycans

3.1 Introduction

Siglec-2, or CD22, an inhibitory glycoprotein mainly expressed on B cells (see Chapter I, §1.4.3), upon specific recognition of α -2,6 linked sialoglycans inhibits the B cell antigen receptor (BCR) signal, developing tolerance to self-antigens and avoiding autoimmune processes and diseases. Siglecs, and in particular CD22, are considered effective glyco-immuno checkpoints within cancer immunotherapy.¹⁶⁴

The main sialoglycans expressed on mammalian tissue contain *N*-acetylated (Neu5Ac) and its derivative *N*-glycolylated (Neu5Gc) sialic acid (see Chapter I, § 1.1.2); importantly, several studies showed the presence of Neu5Gc on fetal tissues and tumor cells.^{165,166} Although humans lack the CMAH enzyme for the synthesis of Neu5Gc, they can incorporate it from dietary sources. Indeed, low levels of Neu5Gc were found on the surfaces of human secretory epithelia and small- and large-blood vessels endothelia.¹⁶⁷ Both murine and human CD22 can bind acetylated and glycolylated sialoglycans, with m-CD22 preferring Neu5Gc over Neu5Ac.¹⁶⁸ Since changes in the Neu5Gc/Neu5Ac ratio can potentially modulate Siglecs' binding and signaling properties, understanding the basis of these interactions may have therapeutic implications. Thus, we here elucidated the molecular binding of CD22 and *ad hoc* synthesized complex-type *N*-glycans and the role of *N*-glycolyl neuraminic acid (Neu5Gc) in the interaction with both orthologues (murine and human CD22). Moreover, our results could enable the development of glycomimetics for modulating the activity of Siglec-2 in autoimmune diseases and B-cell derived malignancies.

3.2 Insights into the affinity of the interaction between h-CD22 and acetylated *N*-glycans

Both human and murine CD22 (h-CD22, m-CD22) were expressed by Prof. Paul Crocker (University of Dundee, UK) as recombinant soluble IgG Fc chimeras. The interaction of h-CD22 with the trisaccharide Neu5Ac- α -(2,6)-Gal- β -(1,4)-GlcNAc- β -(CH₂)₂NH₂ (6'SL_n) was firstly assessed, representing the terminal end of complex-type *N*-glycans, typically exposed on the surface of mammalian cells, through biophysical techniques, including surface plasmon resonance (SPR) and alpha (Amplified Luminescent Proximity Homogeneous Assay) screen assays, ligand-based NMR methods and MD simulations. Then, the binding of h-CD22 with complex-type *N*-glycans was characterized, providing the ligand conformational features and 3D views of the complexes by the ensemble of NMR and computational data.

3.2.1 Biophysical techniques for detecting the binding between h-CD22 and *N*-glycans

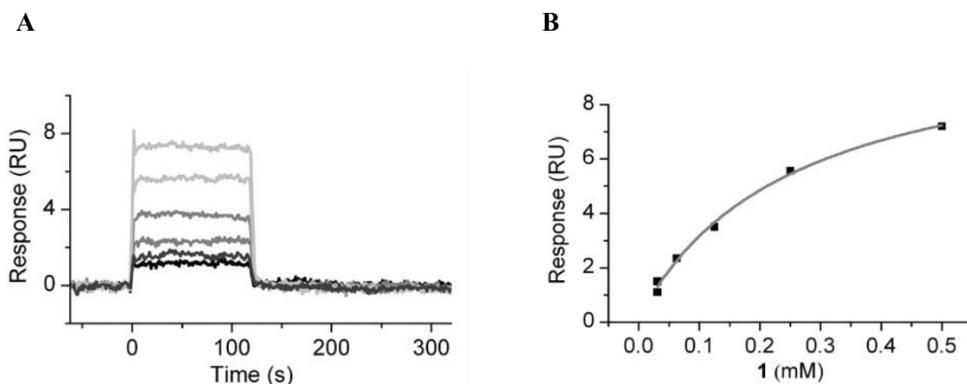


Figure 3.1. SPR analysis. A) Overlay plot of SPR sensorgrams from steady-state affinity analysis of h-CD22 binding to trisaccharide (**1**) at different concentrations. The sensorgrams were reference subtracted and blank subtracted. B) The equilibrium SPR response was plotted against the analyte concentration. Solid line represented the nonlinear curve fitting to the data (squares).

In collaboration with Prof. Antonio Randazzo group (Farmacy Department, University Federico II), SPR binding experiments¹⁶⁹ were acquired to detect the affinity of the interaction between h-CD22 and 6'SL_n. The protein was immobilized on the sensor surface and different concentrations of ligand, ranging from 0.031 mM to 0.5 mM, were injected. A dissociation constant K_D of 250 μ M was evaluated (figure 3.1).

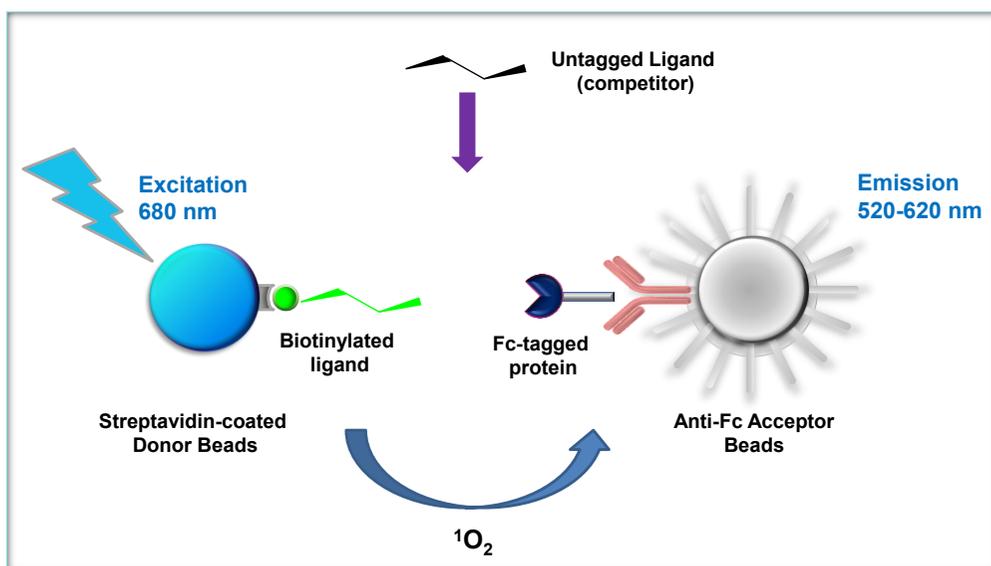


Figure 3.2. Scheme of an alpha screen experiment. Donor bead contained a photosensitizer (phthalocyanine) which converted ambient oxygen to an excited singlet oxygen, upon excitation at 680 nm. Within its 4 μ s half-life, the reactive form of O_2 could diffuse approximately 200 nm in solution. If during this time an acceptor bead was within that distance, the energy transfer from the singlet oxygen to thioxene derivatives in the acceptor bead produced a luminescent signal at 520-620 nm (alpha signal). The presence of the untagged ligand interacting with the protein competed to the biotinylated ligand, leading to a decrease of the light production.

Alpha screen experiments¹⁷⁰ were also performed for the determination of the apparent IC_{50} of different ligands binding to h-CD22 (figures 3.2 and 3.3). This technique was based on the competition of a ligand of interest to a biotinylated ligand, coated on

streptavidin donor bead, interacting with the protein of interest, captured on an acceptor bead. Upon illumination at 680 nm, the energy transfer from one bead to another produced a luminescent signal at 520-620 nm. The decrease of the alpha signal, measured as IC₅₀, was detected when an untagged ligand (the ligand of interest) bound to the protein. The lower the IC₅₀ value the stronger the binding between h-CD22 and the competitor. Our results showed the specificity of the protein towards α -2,6 sialylated ligands and a preference for complex-type *N*-glycan containing two sialic acids (ligand 2, figure 3.3), suggesting the involvement of both residues in the recognition.

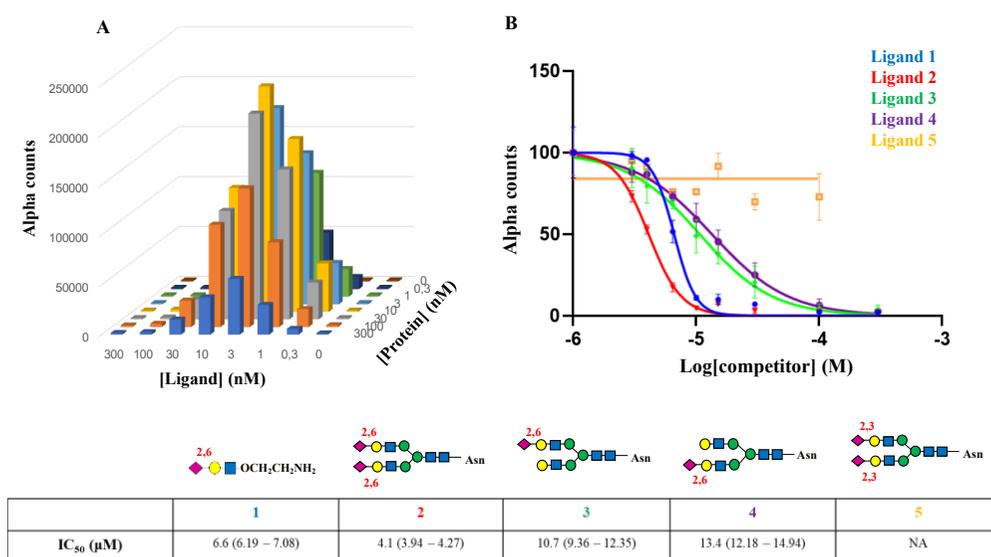


Figure 3.3. Alpha assay on h-CD22 and different sialoglycans (1-5). A) 3D bar-graph of the alpha counts in absence of the competitor to determine the optimal concentrations of protein and biotinylated ligand. B) Dose-response plot of alpha counts at different concentration of the ligand competitor (structure at the bottom) to determine the IC₅₀ (values shown in the table), calculated using GraphPad Prism. 95% confidential interval (95% CI) were listed inside parenthesis. Measurements were done in triplicate.

3.2.2 Molecular binding between h-CD22 and acetylated 6'SL_n

STD NMR experiments were carried out on CD22 in the presence of 6'SL_n trisaccharide (figure 3.4). Interestingly, STD enhancements were detected only for sialic acid and galactose residues (figure 3.4 A). The strongest STD effect belonged to the acetyl group of the sialic acid (AcK), while no STD response was observed for the *N*-acetylglucosamine residue, indicating it pointed far from the protein binding pocket.

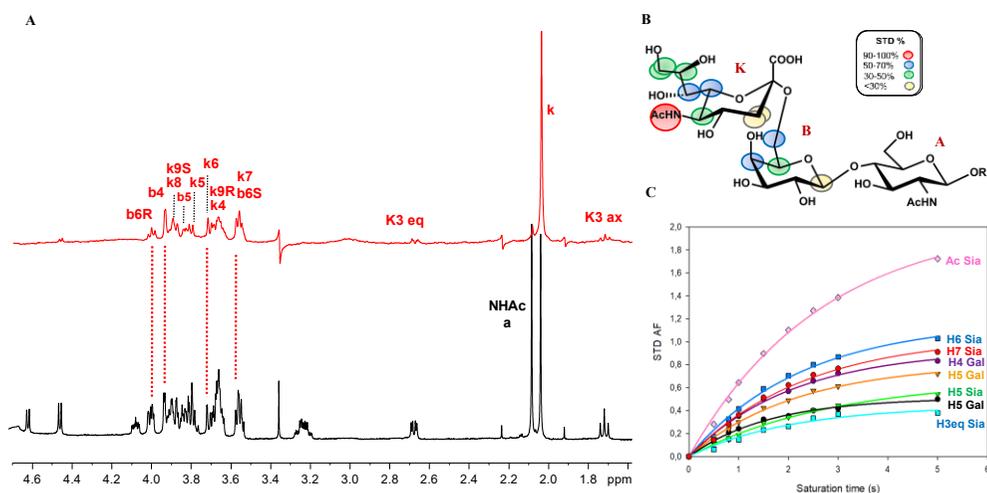


Figure 3.4. STD NMR analysis of h-CD22 and 6'SL_n. A) STD NMR spectrum (red) and off-resonance (black) acquired at saturation time of 2 s. B) Epitope map of 6'SL_n calculated by $(I_0 - I_{\text{sat}})/I_0$, where $(I_0 - I_{\text{sat}})$ was the intensity of the signal in the STD-NMR spectrum and I_0 was the peak intensity of the unsaturated reference spectrum (off-resonance). The highest signal belonging to AcK was set as 100% of STD response and the other proton signals were calculated accordingly. C) STD build-up curves at different saturation times (from 0.5 to 5 s). STD AF intensities on the y axis were calculated according to the equation [2.7] (Chapter II, § 2.1.3).

An accurate epitope mapping of the ligand was obtained by acquisition of the STD signals at different saturation times (figure 3.4 B and C). Indeed, because the intensities of the observed STD signals were correlated not only on their proximity to the receptor but also on longitudinal relaxation time (T_1), the use of STD build-up

curves was used to prevent possible misinterpretation and to overcome possible artifacts due to differences in capability to accumulate saturation in the free state.¹⁷¹ The slope of the STD build-up curve close to saturation time of 0 (STD_{fit}) corresponded to the STD intensity influenced solely by the proximity of the ligand proton to the protein (see Chapter II, § 2.1.3).

The epitope map of the trisaccharide was obtained by normalizing all the values of different protons ligand to the largest STD_{fit} , giving $STD_{epitopes\ fit}$ (table 3.1). Since the acetyl group of the sialic acid were the protons with maximum magnetization transfer, the STD intensity of this peak was set to 100% as a reference and the relative STD intensities for the other protons were normalized based on this peak intensity.¹⁷²

Table 3.1. Measured STD intensities of the trisaccharide bound to h-CD22 at different saturation times. STD_{max} values were calculated by fitting the data to a mono-exponential equation: $STD(t_{sat}) = STD_{max} * (1 - \exp(-k_{sat} * t_{sat}))$.

¹H	STD_{max}	K_{sat}	STD (fit)	STD epitopes (fit)
K Ac	2.0593	0.3725	0.7671	100%
K3ax	0.4342	0.4799	0.2083	27%
K3eq	0.4431	0.4853	0.2150	28%
K5	0.6731	0.3493	0.2351	30.6%
K6	1.1775	0.4381	0.5159	67.2%
K7	1.0562	0.4245	0.4483	58.4%
K8	0.6790	0.4431	0.3001	39.1%
k9R	0.5748	0.5763	0.3313	43.2%
H4 Gal	0.9346	0.4757	0.4445	57.9%
H6R Gal	0.8157	0.4578	0.3734	48.6%
H5 Gal	0.5122	0.6201	0.3176	41.4%

In order to obtain information about the bioactive conformation of the trisaccharide, the conformational behavior of the ligand was initially studied in the free state. The relative orientations of saccharide units are expressed in terms of the glycosidic

linkage torsion angles Φ (O5'-C1'-O6-C6) and Ψ (C1'-O6-C6-C5). For (1→6) linkages, an additional torsion angle, ω (O6-C6-C5-O5), associated to the hydroxymethyl group, provided additional flexibility to the glycosidic linkages (see Chapter II, § 2.3.2.1).

To unveil the conformational distribution of the ω torsion angle, and discriminate between *gt* (staggered conformation at 60°), *gg* (-60°), and *tg* (180°) orientations, the equation proposed by Thibaudeau *et al.*,¹⁷³ was considered:

$${}^2J(H_5, C_6) = -1,29 + 1,53\cos(\omega) - 3,68\sin(\omega) \quad [3.1]$$

The heteronuclear coupling constant ${}^2J(H_5, C_6)$ was calculated by HSQC-HECADE (heteronuclear couplings from e.COSY-type cross peaks) experiment (Figure 3.5).

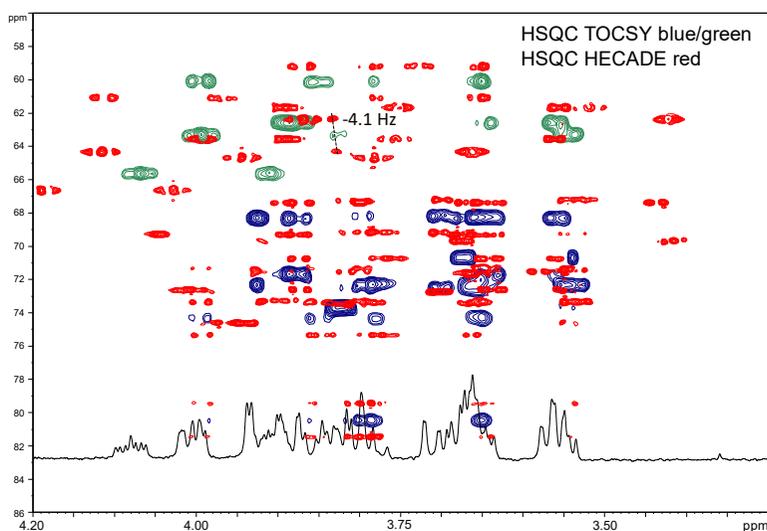


Figure 3.5. HSQC and HSQC-HECADE spectra. Projections over each cross peak yield the active heteronuclear coupling in antiphase in both dimensions. In contrast, the homonuclear proton couplings were exclusively displayed in the proton dimension (F2). Since the one-bond coupling could be assumed to be positive, the signs of the other couplings resulted from the tilt of the respective cross-peak pattern. For negative peaks, only one contour level was plotted.¹⁷⁴

The experimental value of $^2J(H_5, C_6)$, found by HSQC-HECADE NMR experiment, was -4.1. Resolving the eq. 3.1, the ω torsion angle calculated was close to 60° , that corresponded to the *gt* conformation. Moreover, the large value of the $^3J_{H5, H6proR}$ of 8,7 Hz confirmed the preference for the *gt* conformer since it implied the antiperiplanar orientation of the proton H-6proR with respect to H-5. The preference for the *gt* conformation was also detected by molecular dynamics (MD) simulation and molecular mechanics (MM) analysis (figure 3.6), performed by Prof. Sonsoles Martín Santamaría (CSIC, Madrid).

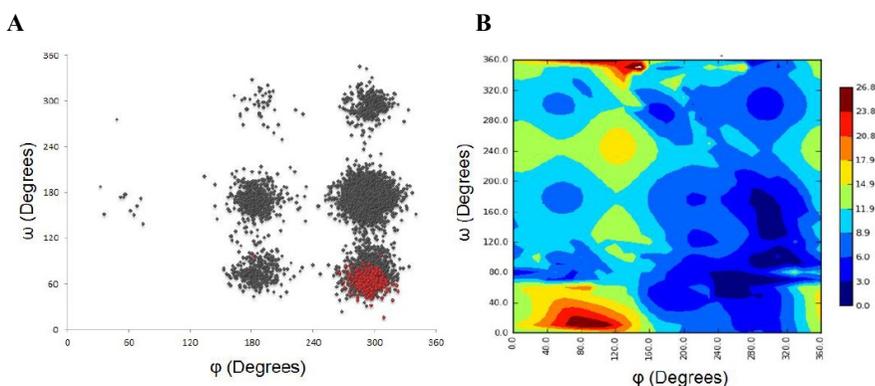


Figure 3.6. MD (A) and MM (B) analysis confirmed the preference for the *gt* conformation in the free state.

The higher *gt* population of trisaccharide suggested the interaction between the terminal NeuAc and the internal sugar residues to stabilize the *gt* conformation, and gave a “bent” conformation.^{175,176}

For the determination of the bioactive conformation of ligand upon binding to h-CD22, tr-NOESY/tr-ROESY experiments were performed and compared to the corresponding spectra in the free state (figure 3.7). The NOESY experiment acquired on the ligand alone resulted close to zero; passing to the bound state, negative NOEs were observed, confirming the binding between h-CD22 and 6'SL_n (figure 3.6 A).

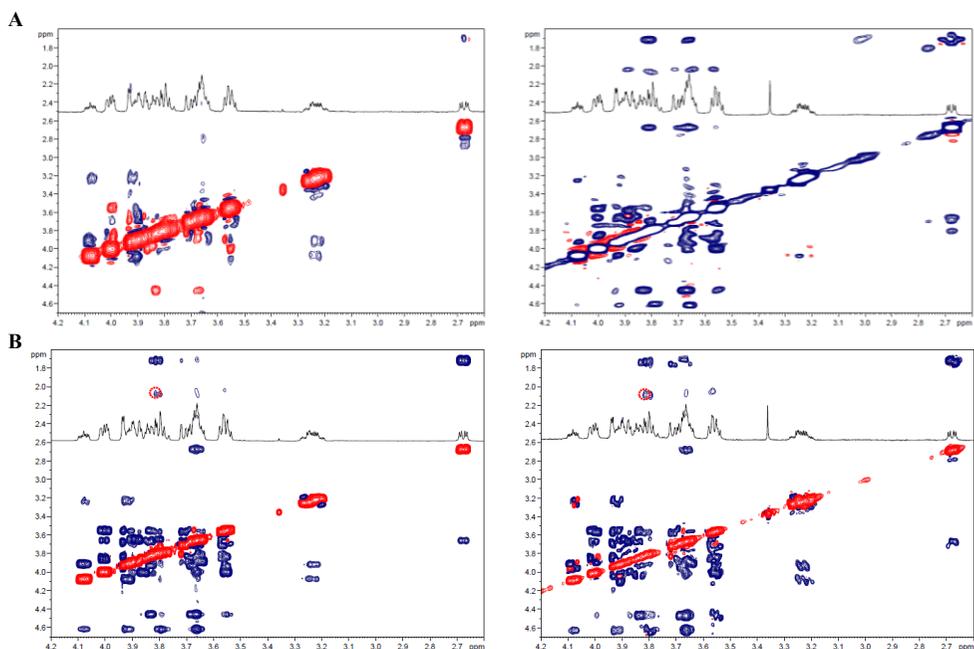


Figure 3.7. NMR experiments for the conformational analysis of 6'SL_n. A) NOESY of the free state at mixing time of 600 ms (left) and tr-NOESY of 6'SL_n bound to h-CD22 at 250 ms (right). B) ROESY NMR spectrum in the free (left) and bound states (right) at 600 and 250 ms, respectively. The key ROE between the proton at position 5 of the sialic acid and the acetyl group of the N-acetylglucosamine residue (red circle in the spectra) was indicative of the ligand bent conformation.

Comparing the free and bound states (figure 3.7 B), no significant differences were detected. Notably, the key ROE contact between H-5 of the sialic acid (K5) and the acetyl group of the GlcNAc (AcA) was indicative of the ligand bent conformation, as suggested by the main population for ω of 60° . The evaluation of *inter-glycosidic* NOE/ROE contacts allowed the description of the binding mode. The cross peaks intensities, measured at different mixing times, from 100 to 600 ms, were used to extract the $^1\text{H}-^1\text{H}$ cross relaxation rate (σ) and to calculate the proton-proton inter-residual distances (table 3.2).

Table 3.2. Experimental and theoretical ^1H - ^1H distances. The experimental inter-proton distances were calculated by ROESY and tr-ROESY in the free and bound state (estimated error 5–10%). The theoretical distances were calculated from MD simulation on the trisaccharide in the free state. Distances were calculated in Angstrom (\AA).

Distance	Exp. Free state	Exp. Bound state	Calc. $\phi = -60^\circ$ (bent conformation)	Calc. $\phi = 180^\circ$
B1-B5	2.56	2.55	2.60	2.60
B1-A4	2.52	2.40	2.40	2.40
B6S-B5	2.97	2.89	2.90	2.90
K3ax-B6R	4.10	4.10	4.20	2.30
K3ax-B6S	4.20	4.30	4.40	2.50
K3eq-B6R	/	/	4.60	3.37
K3eq-B6S	/	/	4.90	3.80
K3ax-K5	2.70	2.60	2.60	2.60
A1-A5	2.57	2.56	2.60	2.60

Comparing the NOE build up curves in the free and bound state (figure 3.8), it was possible to observe the maximum intensity at short mixing times when 6'SL_n bound to h-CD22.

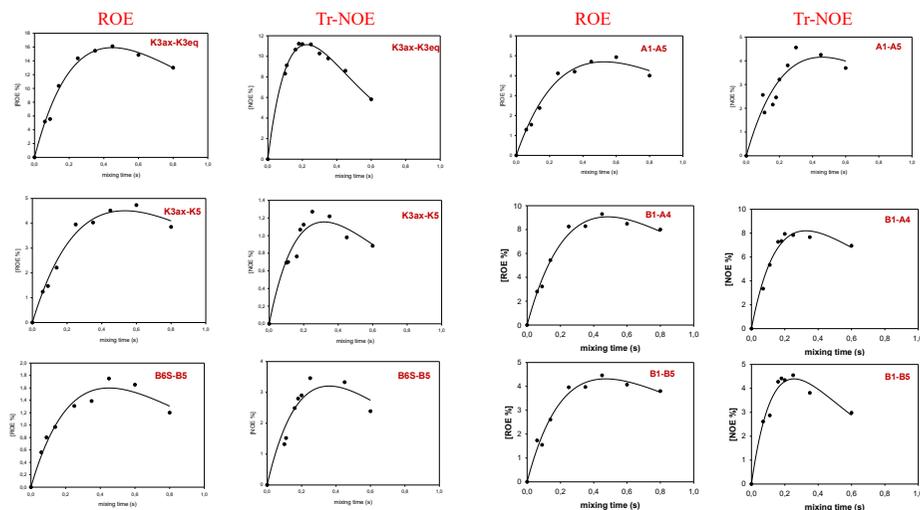


Figure 3.8. NOE and ROE build-up curves. As expected, the maximum enhancement for trNOEs was observed at significantly shorter mixing times t_{mix} for the complex with respect to

the ligand alone in solution. The mixing time values were: 0.10, 0.12, 0.15, 0.20, 0.25, 0.35, 0.45 and 0.60 s.

Indeed, the ligand interacting with the protein underwent a rather sluggish tumbling characterized by a longer molecular correlation time with respect to the small ligand in the free state. As detected above, inter-proton distances of the ligand in free and bound states did not differ significantly, meaning that no large conformational changes were observed upon binding, and thus, the bent conformation of 6'SL_n was maintained (figure 3.9).

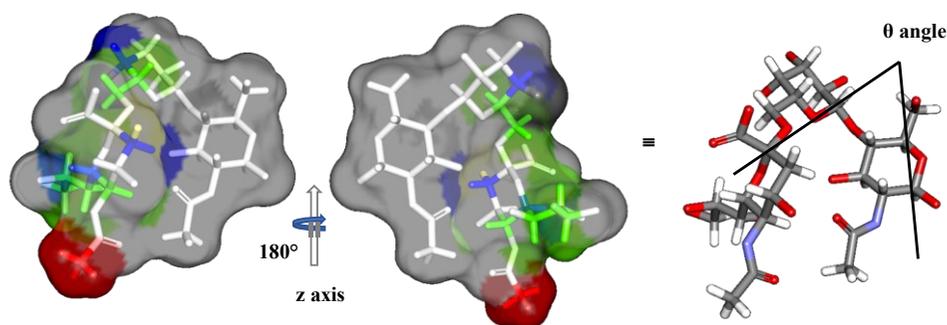


Figure 3.9. STD-derived epitope mapping on the molecular envelope of 6'SL_n in its bioactive conformation, with color code according to the observed STD effects. When ω was 60° (*gt* conformation) the ligand was characterized by an umbrella-like topology with the angle θ , defined by the C2, C1, and C1 atoms of the residues Neu5Ac, Gal, and GlcNAc, respectively (going from the nonreducing end to the reducing end) smaller than 110° (see on the right).

Regarding Φ ($H_1-C_1-O-C_X'$) and Ψ ($C_1-O-C_X'-H_X'$) glycosidic torsion angles, the conformational space was defined by molecular mechanics (MM) calculations using Maestro (Schrödinger). First, the two disaccharides Neu5Ac- α -(2,6)-Gal and Gal- β -(1,4)-GalNAc were constructed for analyzing the energetically accessible conformational regions³⁵ (figure 3.10). Given the above NMR data, the CH₂OH group of the galactose moiety was set in the *gt* conformation, the most stable conformation for a *galacto*-configured sugar unit. The corresponding adiabatic energy maps for Φ and Ψ were shown in figure 3.10. The NeuAc- α -(2,6)-Gal disaccharide populated three

energetic minima in the free state, characterized by a dihedral angle $\Psi \sim (180^\circ)$, while Φ oscillated between three minima.

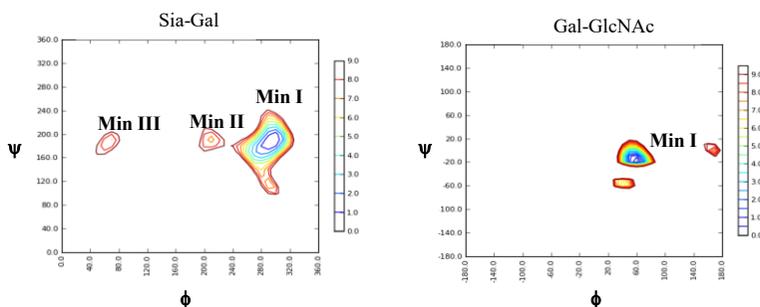


Figure 3.10. Adiabatic energy maps of the glycosidic linkages of 6'SL_n. The most populated minima, in the free state, were indicate as "I".

In collaboration with the group of Prof. Sonsoles Martín Santamaría (CSIC, Madrid), MD simulations were performed on the trisaccharide alone and bound to the protein (PDB: 5VKM). By comparing the NOE derived distances with the calculated ones (table 3.2) it could be concluded that, although there was an equilibrium between different conformational states (figure 3.10), Minimum I, characterized by the following dihedral angles, $\Phi(\sim 60^\circ)$, $\Psi(\sim 180^\circ)$, $\omega(\sim 60^\circ)$ for Neu5Ac- α -(2,6)-Gal and $\Phi(\sim 60^\circ)$, $\Psi(\sim 0^\circ)$ for Gal β -(1,4)-GalNAc was the most populated both in the free and the bound states.

In order to predict a 3D model of the protein-ligand complex (figure 3.11 A), the most representative model obtained from MD was used to predict the theoretical STD effects by means of CORCEMA-ST protocol (see Chapter II, § 2.3.3). The program was implemented on MATLAB and the theoretical STD were compared with the corresponding experimental STD NMR results (figure 3.11 C). A good accordance between experimental and theoretical values were observed (figure 3.11 B). Indeed, consistent with STD NMR data, the CORCEMA-ST prediction showed several

protons of the sialic acid exhibiting high STD values, confirming their strong contribution to the interaction with the protein.

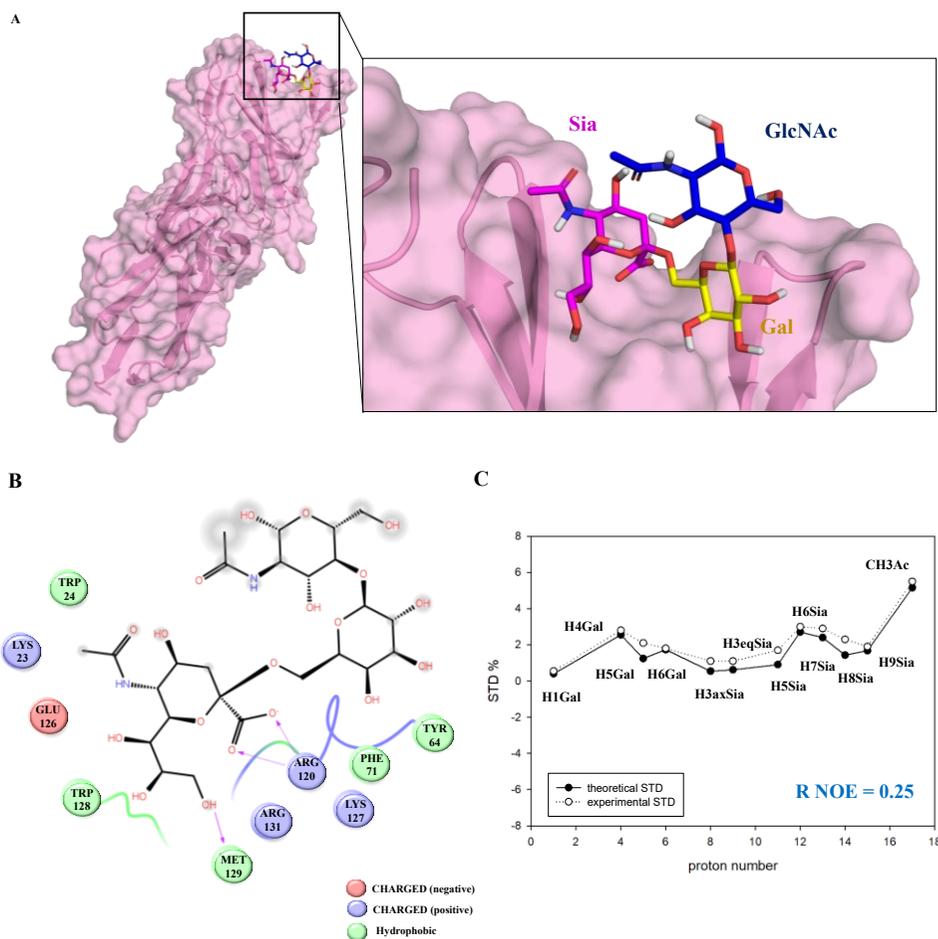


Figure 3.11. Molecular interaction between h-CD22 and 6'SLn. A) 3D best view of h-CD22 and 6'SLn obtained by MD simulation compared to the CORCEMA-ST. B) 2D interactions occurring in the complex. C) Plot of the theoretical and experimental STD calculated by CORCEMA-ST. The R-NOE of 0.25 validated the 3D complex of h-CD22 and 6'SLn.

Galactose unit showed lower STD effects and the predicted STD values for the *N*-acetylglucosamine moiety were negligibly small, confirming that it was not in intimate contact with the protein; thus, the 3D complex was validated with CORCEMA (figure 3.11). In particular, the highest STD response found for the acetyl group of sialic acid (AcK) was explained by the close contacts with Trp24, Trp128 and Glu126 side chains, and in particular by the H-bond between the NH of the sugar and the carbonyl group of Lys127. The glycerol chain of the sialic acid was also pointed toward the h-CD22 binding pocket, with the high STD effect found for the H7 making CH- π interaction with Trp128. Regarding the H9 protons, only one of the methylene protons was oriented towards Trp128 indole group, whereas the other H9 was solvent exposed.

Interestingly, 3D models containing lower populated dihedral angles of the trisaccharide were considered in the CORCEMA-ST program (data non shown); however, the R-NOE values was very high, and thus the corresponding complexes excluded.

3.2.3 Molecular binding between h-CD22 and complex-type *N*-glycan

Once the binding between h-CD22 and the undecasaccharide representing a typical complex-type *N*-glycans was assessed by alpha screen assay (see § 3.2.1), the molecular features of the interaction were characterized by NMR. Interestingly, STD NMR spectrum and ligand epitope mapping (figure 3.12) indicated that only sialic acid and galactose units were involved in the binding, following a recognition pattern comparable to shorter ligands (§ 3.2.2).

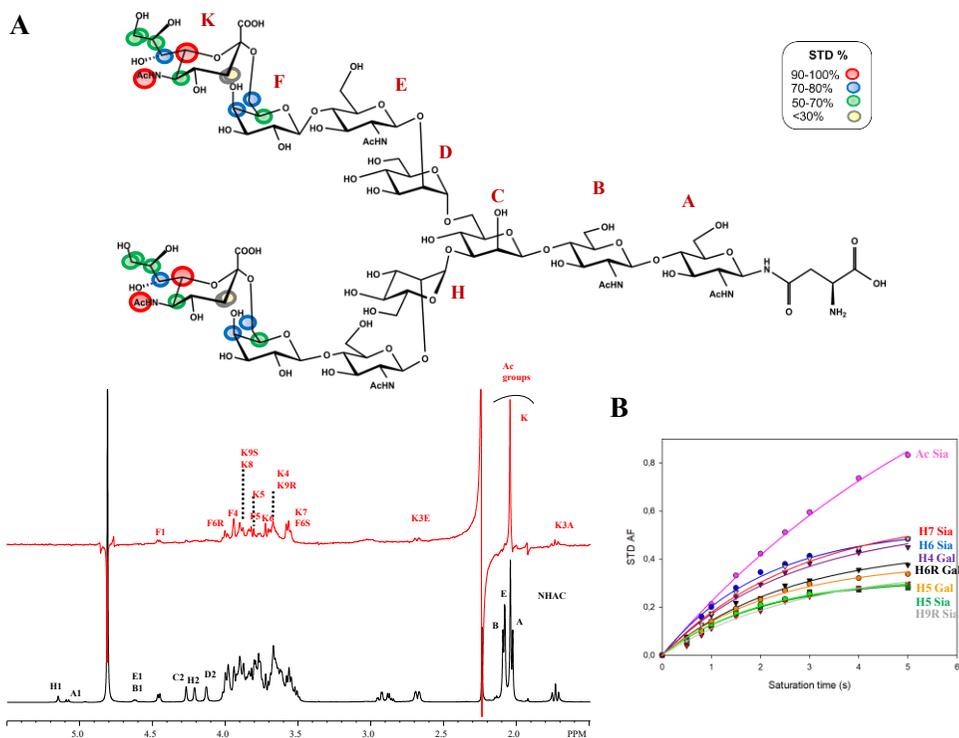


Figure 3.12. STD NMR analysis of h-CD22 and complex-type *N*-glycan. A) The STD signals (red spectrum) and the epitope mapping of the undecasaccharide revealed the recognition of only sialic acid and galactose residue from h-CD22. The strongest STD signal was observed for the acetyl group protons (Ac) of the sialic acid B) STD build-up curves.

STD build-up curves were also acquired (figure 3.12 B) by collecting spectra at different saturation times (from 0.5 s to 5 s). The acetyl group protons of the sialic acid gave the maximum magnetization transfer and was set to 100% (table 3.3).

Since the STD NMR of the undecasaccharide (figure 3.12) did not allow to discriminate between the arms of the ligand, STD NMR experiments on the corresponding monosialylated *N*-glycans, whose alpha screen assays were previously shown (§ 3.2.1), were also performed (figure 3.13).

Table 3.3. Experimental STD intensities of *N*-glycan bound to h-CD22 at different saturation times. STD_{max} values were calculated by fitting the data to a mono-exponential equation: $STD(t_{sat}) = STD_{max} * (1 - \exp(-k_{sat} * t_{sat}))$.

¹ H	STD _{max}	K _{sat}	STD (fit)	STD epitopes (fit)
K Ac	1.6192	0.1484	0.2402	94.6%
K5	0.3232	0.4965	0.1605	63.2%
K6	0.5310	0.4778	0.2537	100%
K7	0.5933	0.3566	0.2116	83.4%
K8	0.3158	0.5034	0.1590	62.7%
k9R	0.3666	0.3560	0.1305	51.4%
H4 Gal	0.5473	0.3739	0.2045	80.6%
H5Gal	0.3895	0.4410	0.1719	67.7%
H6R Gal	0.4495	0.3851	0.1731	68.2%

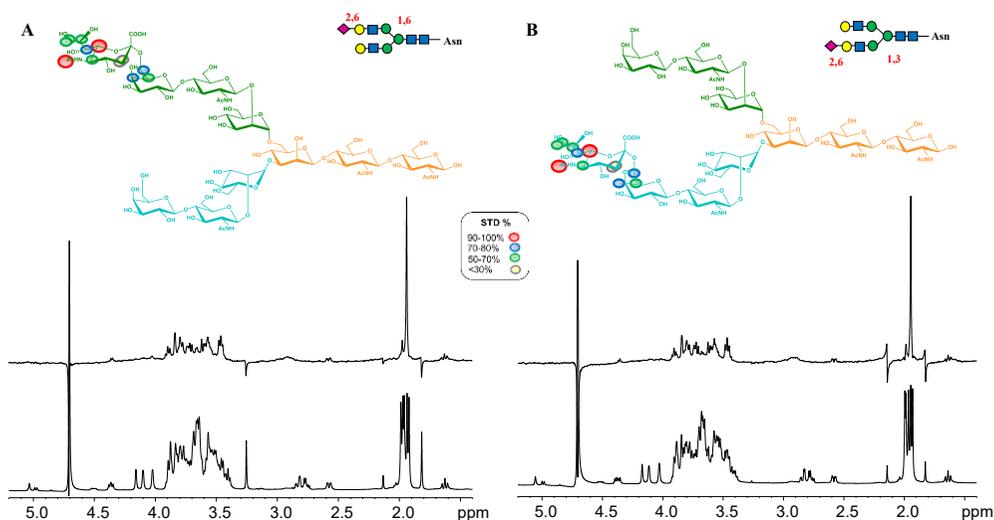


Figure 3.13. STD NMR spectra and epitope mapping of the h-CD22 with monosialylated *N*-glycans. A) Sialic acid containing branch on the Man- α (1,6)-Man glycosidic linkage. B) Sialic acid containing branch on the Man- α (1,3)-Man glycosidic linkage.

The STD NMR profiles were comparable, indicating a similar recognition of the two ligands from h-CD22. Thus, it was possible to deduce that the sialic acid-galactose

units of both arms of the undecasaccharide (figure 3.13) were both and similarly recognized by h-CD22.

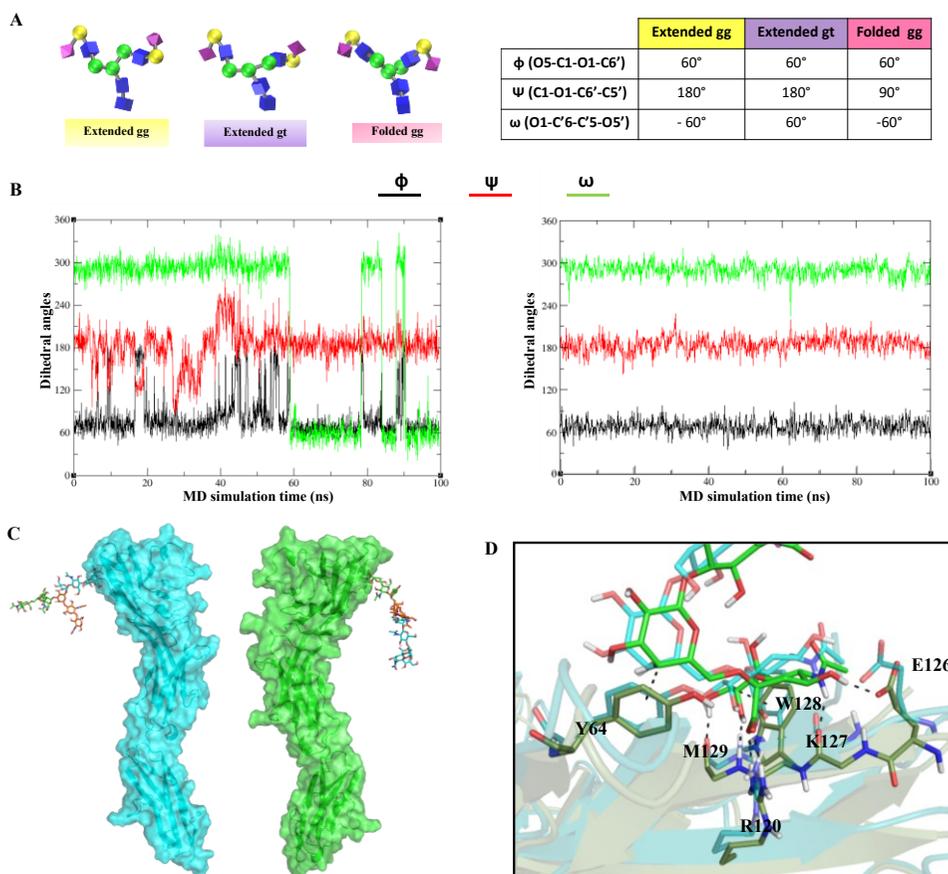


Figure 3.14. Conformational analysis of complex-type *N*-glycan into h-CD22. A) The flexibility of the ligand was given by the presence of Man- α -(1,6)-Man glycosidic linkage, leading to the equilibrium of different conformations in the free state, grouped as “extended” and “folded” shapes. B) MD simulations, performed by Prof. Sonsoles Martín Santamaría (CSIC, Madrid), in the free (left) and bound (right) states. The “extended gg” conformer was selected. C) 3D complexes of the undecasaccharide with h-CD22 via its 1-6 branch (green) and its 1-3 branch (blue), showing the involvement of only sialic acid and galactose moieties. D) A snapshot from MD simulation showing H-bond interactions (dashed lines) between the residues of each branch of undecasaccharide (blue and green) into h-CD22 binding site.

Regarding the conformational analysis of the complex-type *N*-glycan, the presence of the additional ω (O5–C5–C6–O2') torsion angle around the α -(1,6) linkage added a major flexibility to the ligand. Indeed, among the two roughly grouped “folded” and “extended” forms,^{177,178} the combination of NMR and MD results demonstrated that the biantennary sialylated *N*-glycans in complex with h-CD22 preferentially chose an extended conformation, with the lactosamine branches adopting the umbrella-like topology (figure 3.14).¹⁷⁹

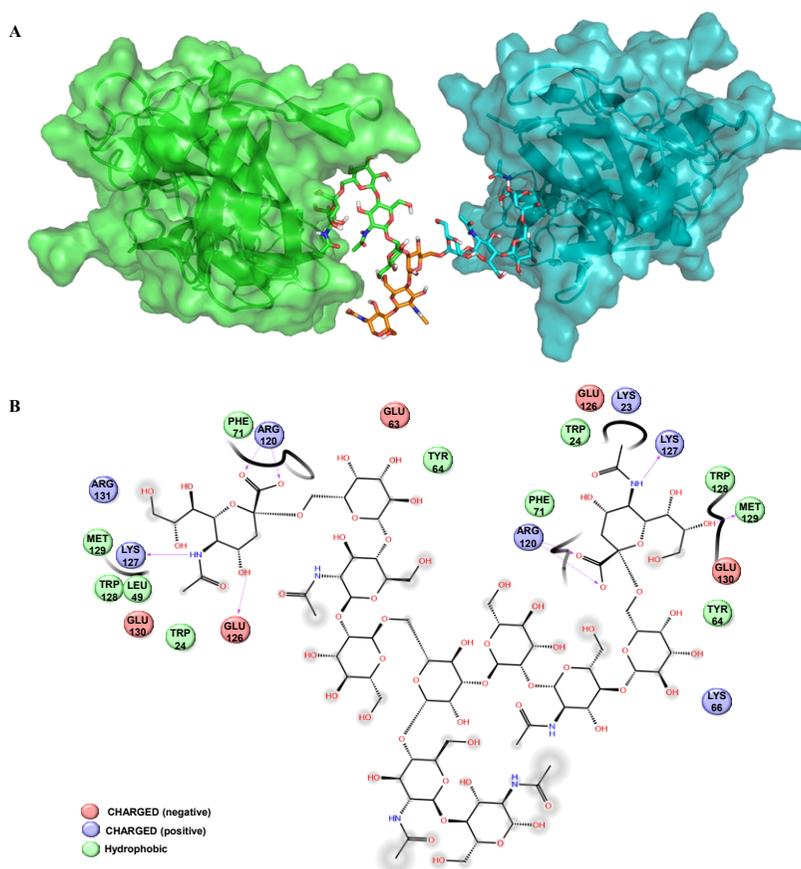


Figure 3.15. The extended geometry of the ligand allowed the simultaneous interaction with two h-CD22. A) 3D top view of the ternary complex of (h-CD22)₂ – complex-type *N*-glycan. The stability of the complex was monitored by MD simulations. B) 2D plot showing

hydrophobic and polar interactions occurring between the undecasaccharide and the two proteins. The two branches of the ligand were recognized in the same way.

Interestingly, the extended gg conformation of the complex-type *N*-glycan allowed the simultaneous interactions with two h-CD22 proteins. MD simulations confirmed the stability of this ternary complex (figure 3.14), that maintained amino acids contacts analogously to those bound to the trisaccharide (§ 3.2.1).

Moreover, it was observed that complex-type *N*-glycan belonging to glycosylation pattern of h-CD22 could interact with two h-CD22 proteins, forming a quaternary complex (figure 3.15), leading to implications in biological aspects in the formation of CD22 homo-oligomers on B cells, favoring the *cis* interactions (Chapter I, §§ 1.3.2 and 1.3.3).

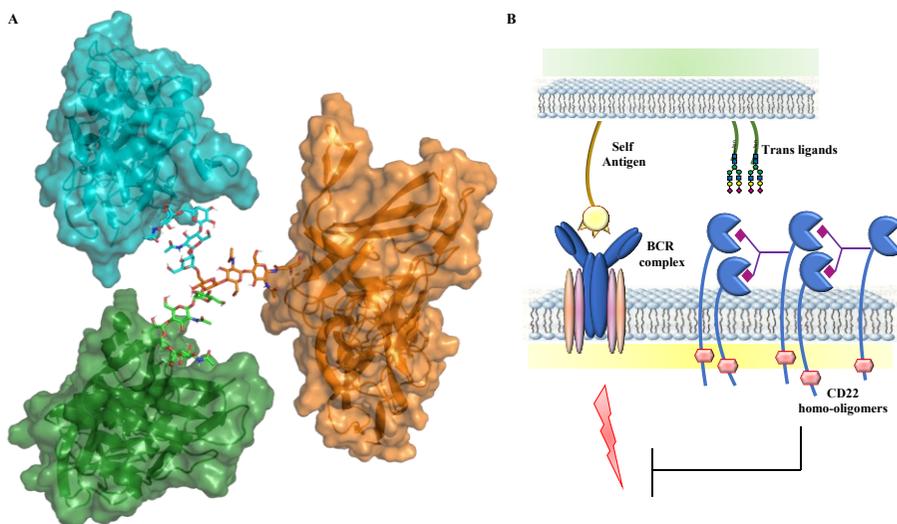


Figure 3.16. Quaternary complex of (h-CD22)₃ – complex-type *N*-glycan. A) 3D top view of the complex obtained by MD simulation. B) Biological effects of the quaternary complex allowed to the formation of *cis* interactions. Here, the glycosylation site of the protein was at position Asn112.

3.3 Comparison of human and murine CD22 interactions with *N*-glycans

The molecular interactions of human and murine CD22 to acetylated and glycolylated *N*-glycans were further investigated.

3.3.1 Determination of the binding affinities of h- and m-CD22 to acetylated and glycolylated 6'SL_n

Firstly, the binding affinities were evaluated by fluorescence analysis (figure 3.17). Fluorescence titrations of increasing amounts of sialoglycans into a fixed concentration of the proteins were performed and binding constants (K_b) were calculated. The results indicated a similar recognition of the sialoglycans from h- and m-CD22, all exhibiting a binding constant in the micromolar range.

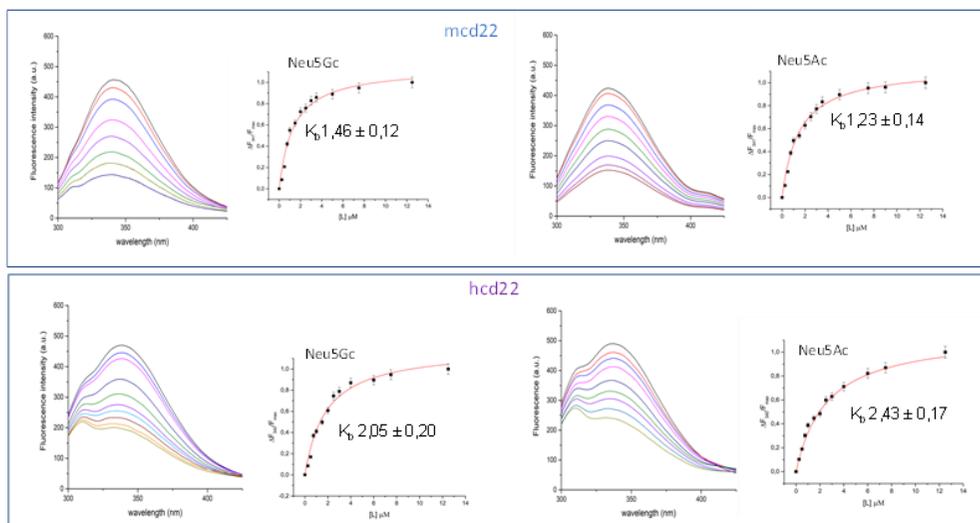


Figure 3.17. Fluorescence titrations experiments of h- and m-CD22 with acetylated (Neu5Ac) and glycolylated (Neu5Gc) sialoglycans. The binding isotherm and the values of the binding constants (K_b) were also reported.

3.3.2 Comparison of human and murine CD22 interactions with acetylated *N*-glycans by NMR

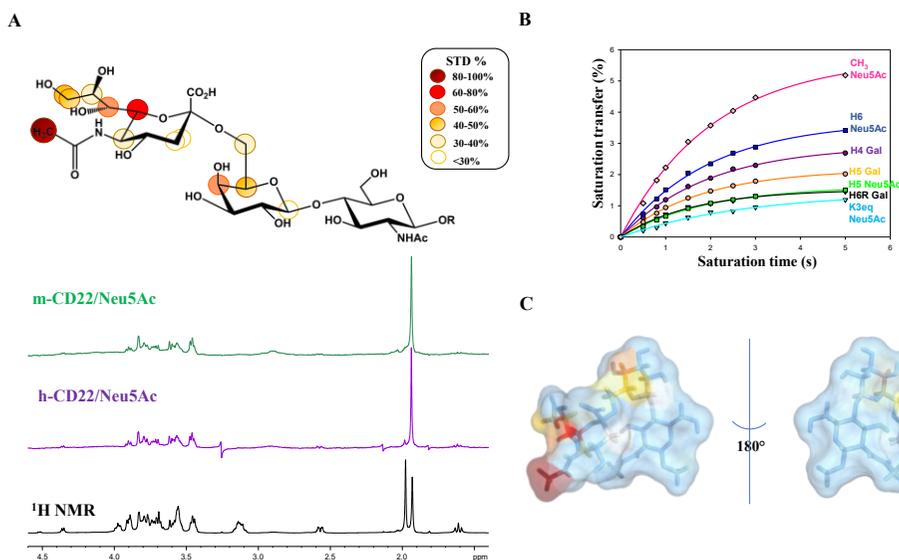


Figure 3.18. Analysis of the interactions between m-CD22 and the acetylated 6'SLn. A) The STD NMR spectra between the h- and m-CD22 interacting with the acetylated 6'SLn provided the same epitope mapping. STD effects lower than 10% were not indicated. B) STD build-up curves of the m-CD22 and acetylated 6'SLn. C) Bent conformation of the ligand with the STD color code.

The comparison of NMR spectra when h- and m-CD22 bound to Neu5Ac containing glycans revealed a similar binding epitope (figure 3.18, table 3.4); a comparable behavior was detected for complex-type *N*-glycans (not shown).

Table 3.4. Experimentally measured STD intensities of acetylated 6'SLN bound to m-CD22 at different saturation times.

I_H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
AcK	5.7320	0.4884	2.7995	100%
K6	3.7105	0.5080	1.8849	67.3%
B4	2.9244	0.5187	1.5169	54.2%
B5	2.1719	0.5580	1.2119	43.3%
B6R	1.5247	0.6170	0.9407	33.6%
K5	1.6331	0.5256	0.8584	30.6%
K3E	1.4283	0.3632	0.5188	18.5%

3.3.3 Comparison of human and murine CD22 interactions with glycosylated *N*-glycans

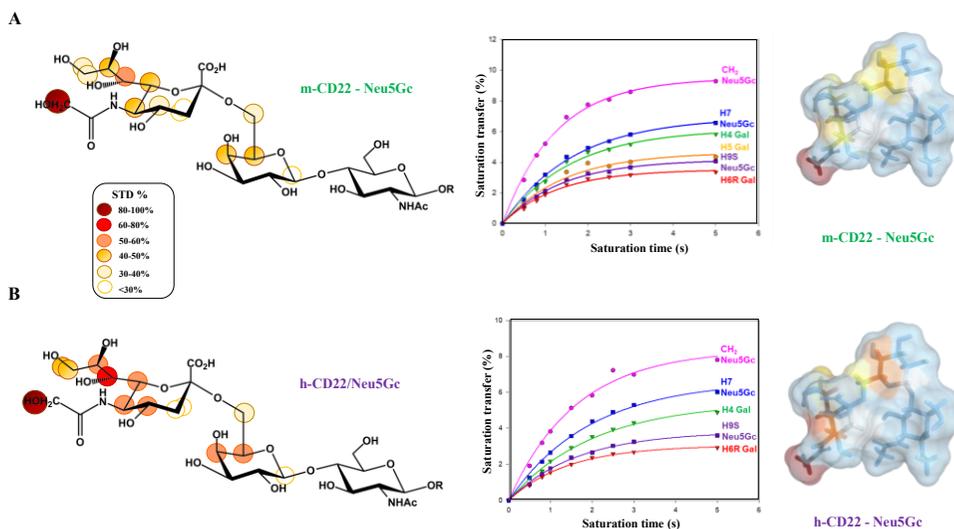


Figure 3.19. STD NMR analysis of A) m-CD22 and B) h-CD22 with the glycosylated 6'SLN. The epitope maps were calculated from the ratio $(I_0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ was the STD signal and I_0 was the peak intensity of the unsaturated reference spectrum. STD effects lower than 10% were not indicated.

STD NMR experiments revealed that the glycolylated ligand was similarly recognized by both h- and m-CD22, giving a comparable epitope mapping (figure 3.19) (tables 3.5 and 3.6).

Table 3.5. Experimental STD intensities of glycolylated 6'SL_n bound to m-CD22.

¹H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH₂ Neu5Gc	9.4963	0.8017	7.6132	100%
H7 Neu5Gc	6.9388	0.6038	4.1897	55.0%
H4 Gal	6.1757	0.6239	3.8530	50.6%
H5 Gal	5.7199	0.6108	3.4937	45.9%
H9S Neu5Gc	4.1764	0.7117	2.9723	39.0%
H6R Gal	3.5266	0.7722	2.7232	35.8%

Table 3.6. Experimental STD intensities of glycolylated 6'SL_n bound to h-CD22.

¹H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH₂ Neu5Gc	8.3838	0.6205	5.2021	100%
H7 Neu5Gc	6.6503	0.5120	3.4049	65.4%
H4 Gal	5.4644	0.4935	2.6967	51.8%
H9S Neu5Gc	3.8238	0.6113	2.3375	44.9%
H6R Gal	3.0547	0.7042	2.1511	41.3%

As found for the acetylated 6'SL_n, the sialic acid—galactose moiety was the only determinant of the binding to both h-CD22 and m-CD22 (figure 3.19). The highest STD was attributed to the glycolyl moiety of Neu5Gc, set to 100%. Then, H7 sialic acid was saturated more than 50%. Protons H5, H6, and H8 of Neu5Gc and H5 and

H4 of Gal showed %STD in the range of 40%–50%. The lowest STD effects were detected for the diastereotopic H3 protons of sialic acid. The GlcNAc residue (A) was solvent exposed since no STD signals were found.

Regarding the conformational analysis of the glycolylated ligand, MD simulations on the free state were firstly run (figure 3.20). Differently from the Neu5Ac trisaccharide, where the ϕ torsion angle around Neu5Ac and Gal linkage could populate different minima ($-60^\circ/60^\circ/180^\circ$), the Neu5Gc glycan preferentially adopted a conformation with ϕ of approximately -60° in the free state.

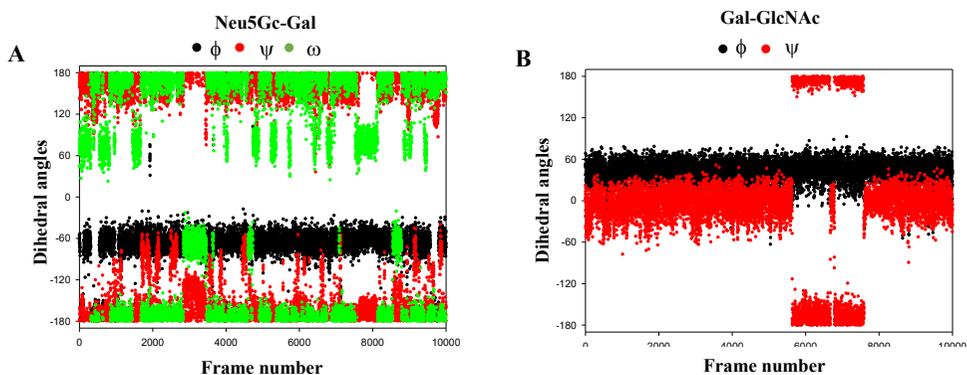


Figure 3.20. MD simulations of the glycolylated 6'SL_n in the free state. A) Torsion angles around Neu5Gc- α -(2,6)-Gal linkage. B) Torsion angles around Gal- β -(1,4)-GlcNAc linkage.

The analysis of the ^1H - ^1H inter-proton distances calculated by the NOESY experiment also confirmed a population corresponding to a ϕ torsion angle of -60° (figure 3.21 and table 3.7). Moreover, glycolylated 6'SL_n bound to h- and m-CD22 adopted a similar bioactive conformation, as evidenced by NOE experiments.

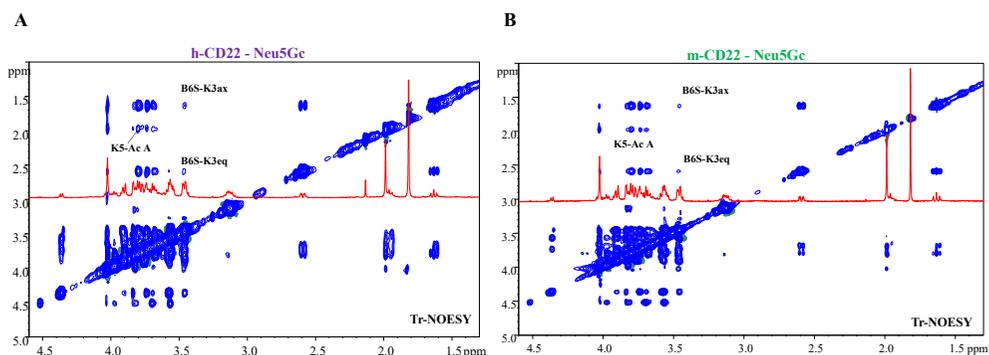


Figure 3.21. Tr-NOESY of the glycosylated 6'SL_n upon binding to A) h-CD22 and B) m-CD22. The spectra were comparable, indicating a similar bioactive conformation of the ligand into both h- and m-CD22.

Table 3.7. Theoretical and experimental ¹H-¹H inter-proton distances (calculated in Å) of the glycosylated trisaccharide in the free and bound states with h- and m-CD22 (estimated error 5–10%).

Distances	Family I	Family II	Free state Exp. distances	Exp. h-CD22 bound state	Exp. m-CD22 bound state
	$\Phi = -60^\circ$ $\Psi = 180^\circ$ $\omega = 60^\circ$	$\Phi = 180^\circ$ $\Psi = 180^\circ$ $\omega = 60^\circ$			
H3_{eq} Neu5Gc - H6S Gal	4.93	3.84	4.50	4.61	4.72
H3_{eq} Neu5Gc - H6R Gal	4.58	3.37	nd	nd	nd
H3_{ax} Neu5Gc - H6S Gal	4.43	2.53	4.11	4.80	4.90
H3_{ax} Neu5Gc - H6R Gal	4.25	2.35	nd	nd	nd
H5 Neu5Gc - CH₃ GlcNAc	4.30	9.60	nd	4.96	4.81

In particular, the absence of NOE contacts between the H6-proR of Gal and the diastereotopic (axial and equatorial) H3 protons of Neu5Gc and the key NOE occurring between the acetyl group of GlcNAc and H5 of sialic acid (Table 3.7) detected in the tr-NOESY spectra (figure 3.21) indicated that the glycosylated ligand assumed a bent conformation, as found for the acetylated 6'SLn (§ 3.2.2).

Since the 3D structure of m-CD22 was not available, homology modelling using the crystal structure of h-CD22 (PDB: 5VKJ)⁷⁸ as structural template was performed by Dr. Rosa Ester Forgiione (see details in Chapter VIII). The most relevant differences between h- and m-CD22 lied in the replacement of Lys23^h, Tyr64^h, and Lys127^h with Asp25^m, Phe68^m, and Arg131^m, which slightly affected the polarity of the binding site (figure 3.22).

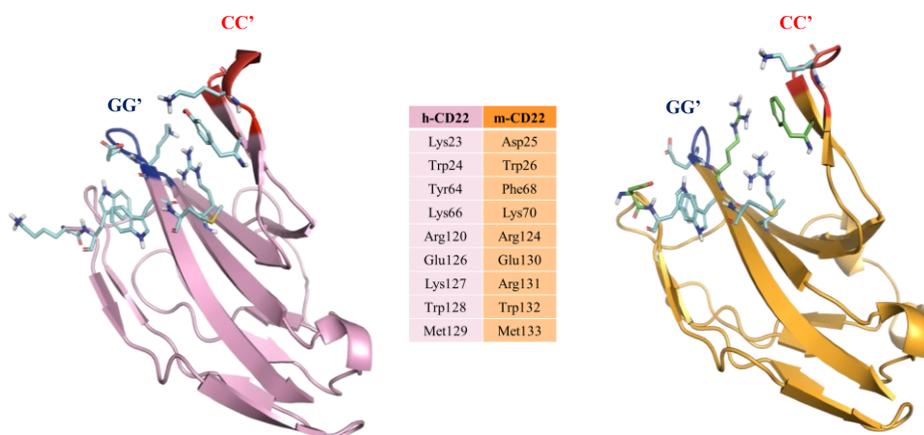


Figure 3.22. Comparison of the *N*-terminal V-set domains of h-CD22 (pink, PDB: 5VKM) and the homology model of m-CD22 (orange). Common residues in the binding sites were colored in cyan, while different residues were highlighted in green. All these residues were listed in the table.

The model of m-CD22 was subjected to 100 ns MD simulation to determine the pose at lowest energy to use in complex with the ligand (data not shown). Then, Neu5Gc

ligand, in the most populated gt conformation, was docked into h-CD22 and m-CD22 binding sites by means of AutoDock 4.2 and the corresponding complexes, showing lower relative energy and the highest populated cluster, were chosen for running 100 ns MD simulation.

The root mean square deviation (RMSD) of the m-CD22/Neu5Gc and h-CD22/Neu5Gc complexes indicated a good stability of the poses along the MD simulation, as shown by the ligand RMSD values within ~ 1.5 -2 Å and by the many contacts monitored (figure 3.23).

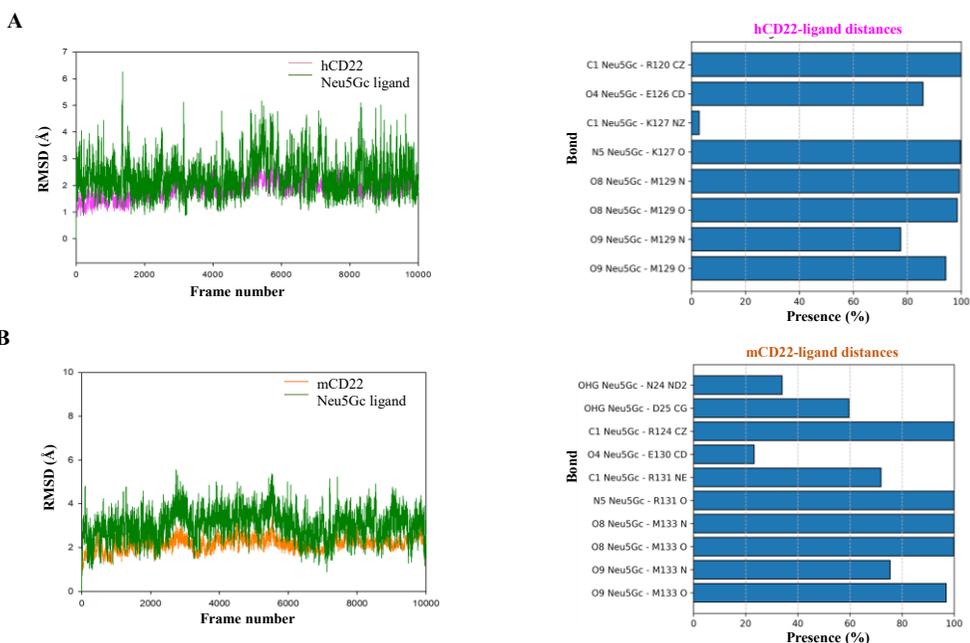


Figure 3.23. RMSD (left) and most representative inter-molecular distances (right) among the A) h-CD22/Neu5Gc and B) m-CD22/Neu5Gc ligands.

To assess the best 3D view of the glycosylated ligand into h-CD22 and m-CD22, different poses from the cluster analysis from the MD simulations were considered in

the CORCEMA-ST program. Those showing the best fit between theoretical and experimental STD data were shown in figures 3.24 and 3.25.

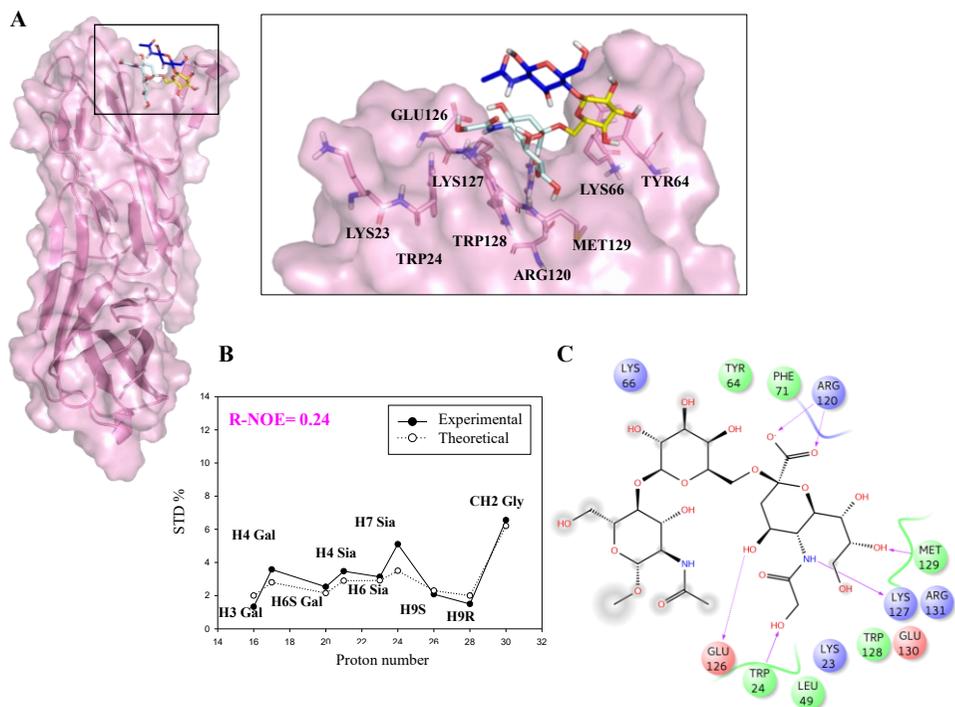


Figure 3.24. Human-CD22/Neu5Gc ligand complex. A) Best 3D view from MD simulation. B) 2D plot comparing the theoretical (solid line) and experimental (dashed line) STD values by CORCEMA-ST. C) 2D plot of the interactions representing the interactions between the glycolylated 6'SLn and h-CD22 binding site residues, deriving from the best representative frame from the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone.

Regarding h-CD22/Neu5Gc ligand complex, the contacts established along the MD simulation were similar to those found between h-CD22 and the Neu5Ac ligand (figure 3.4). The most stable interactions were detected for the carboxylate group of Neu5Gc involved in the salt bridge with the guanidine group of the highly conserved Arg120, and for the N-glycolyl group of Neu5Gc in the H-bond with Lys127 (figure

3.23 A). Furthermore, the methylene protons of *N*-glycolyl group of Neu5Gc made hydrophobic interactions with Trp24 and Trp128 residues. Recurring H-bonds were also found between the glycerol moiety of the glycosylated ligand with Met129 backbone oxygen and amide, as well as CH- π interactions with Trp128. Moreover, OH4 Neu5Gc engaged a polar interaction with Glu126. As for the galactose, it was found its involvement in the CH- π interaction with Tyr64 aromatic residue, whereas GlcNAc residue was far from the h-CD22 surface for most part of the simulation.

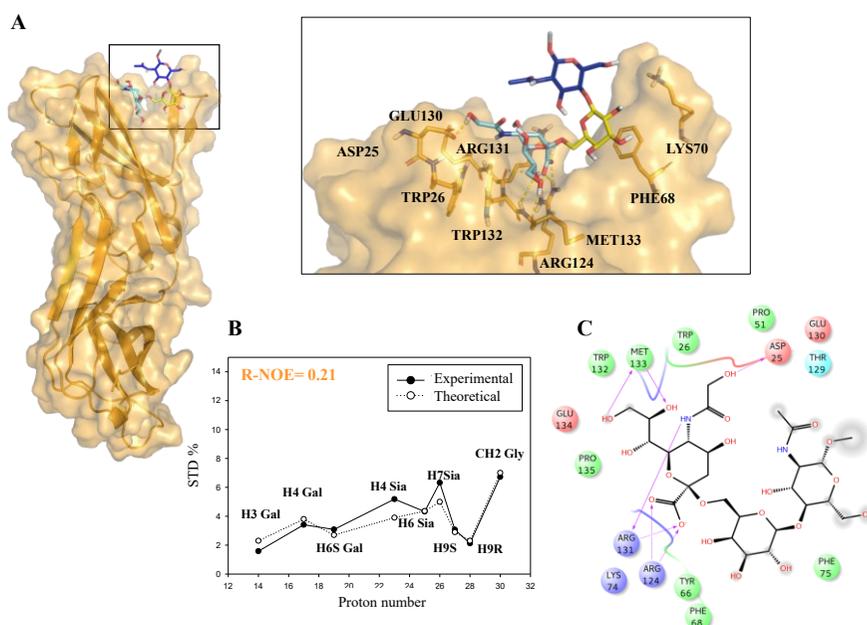


Figure 3.25. Murine-CD22/Neu5Gc ligand complex. A) Best 3D view from MD simulation. B) 2D plot comparing the theoretical (solid line) and experimental (dashed line) STD values by CORCEMA-ST. C) 2D plot of the interactions representing the interactions between the glycosylated 6'SLn and m-CD22 binding site residues, deriving from the best representative frame from the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone.

Regarding m-CD22/Neu5Gc ligand complex (figure 3.25), the pattern of interactions was similar to the human ortholog, with Neu5Gc and Gal residues involving in several

polar and hydrophobic interactions, whereas GlcNAc was excluded from the h-CD22 binding site. Additionally, Arg131 played the same role of Lys127 in h-CD22 receptor, making an H-bond between its backbone oxygen and the amide nitrogen of Neu5Gc N-glycolyl moiety. Another difference from h-CD22 complex was the involvement of the hydroxyl group of Neu5Gc in polar interaction with Asp25, as also supported by CORCEMA-ST.

Indeed, different poses of m-CD22/Neu5Gc ligand complex, lacking the H-bond between OH at position 5 of Neu5Gc and Asp25, provided higher R-NOE values due to the significantly lower STD value attributed to the N-glycolyl moiety (data not shown).

3.3.4 Comparison of acetylated and glycolylated 6'SL_n into the h-CD22 and m-CD22 binding sites

Once refined the 3D structure of m-CD22, the protein in complex with acetylated 6'SL_n was analyzed (§3.3.2) to describe the 3D structure (figure 3.26). Thus, the bent conformation of Neu5Ac containing glycan accommodated into h- and m-CD22 binding sites in similar manner (figures 3.4 and 3.26).

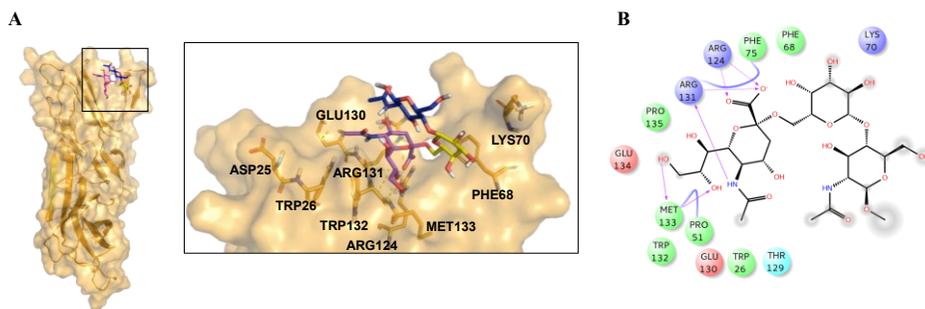


Figure 3.26. Murine-CD22/Neu5Ac ligand complex. A) 3D view obtained by the most representative cluster from the MD simulation. B) 2D interactions establishing between m-CD22 and the acetylated 6'SL_n.

Overall, the combination of theoretical and experimental results showed that h- and m-CD22 similarly recognized both acetylated and glycolylated ligands (Figure 3.27). A slight difference was found in the shape and polarity of the cavity of the two orthologs, given by the presence of Asp25 only in m-CD22 receptor, which further assumed an optimal position for binding to longer ligands (figure 3.27 B).

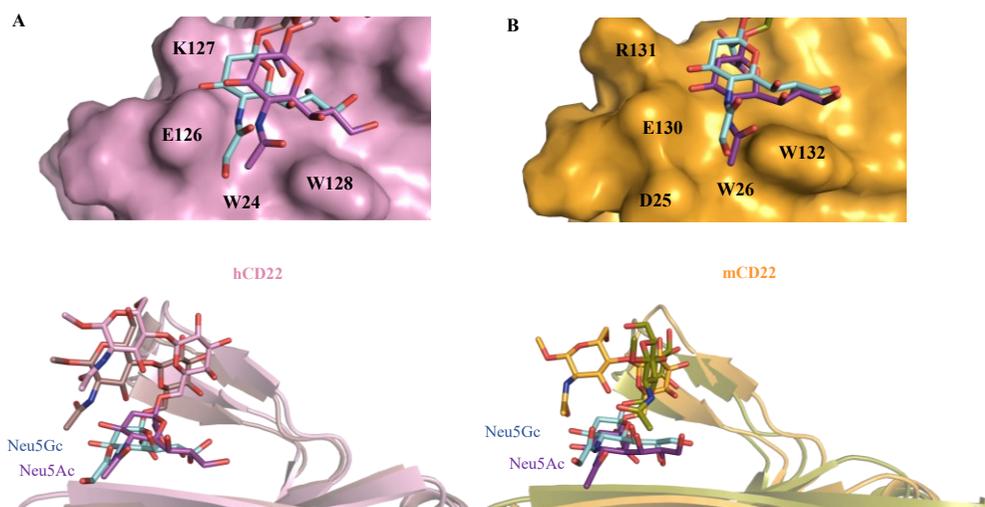


Figure 3.27. Comparison of the interaction of Neu5Ac/Neu5Gc ligands with h-CD22 (A) and m-CD22 (B).

3.4 Discussion

Given the role of Siglec-2 in immune tolerance, unveiling the molecular determinants of the interactions with *N*-glycans exposed on mammalian cells could provide rational opportunities for further developing CD22 potent inhibitors and/or regulators. To better define the structural basis of recognition, we studied the molecular interactions of various sialylated *N*-glycans by human and its ortholog murine CD22. The combination of ligand-based NMR experiments, such as STD and tr-NOESY,

together with molecular mechanics, homology modelling, docking, molecular dynamics and the use of CORCEMA-ST, provided information on the ligands' conformational behavior and 3D structure of the complexes. We found that m-CD22 binding site was comparable to h-CD22, accommodating acetylated and glycolylated ligands in a bent umbrella-like topology, although the presence of Asp25^m in place of Lys23^h allowed to establish additional H-bonds with the hydroxyl group of the glycolyl moiety. Overall, the sialic acid-galactose moiety of *N*-glycans was the unique portion recognized by CD22 receptors. Indeed, the investigation of the h-CD22 binding to biantennary complex-type *N*-glycans revealed that all the other residues were solvent exposed. Interestingly, the extended conformation adopted by longer *N*-glycans allowed to simultaneously interact with different h-CD22 proteins, causing the formation of homo-oligomers on the same B cell and, thus, favoring the *cis* interactions.

Chapter V:

Molecular details of sialoglycans recognition by *Streptococcus gordonii* Siglec-like adhesins

V. Molecular details of sialoglycans recognition by *Streptococcus gordonii* Siglec-like adhesins

5.1 Introduction

Infective endocarditis (IE) is an inflammatory disease that provokes the infection of the endocardium and the damage of the heart. The inflammation is generally caused by the entrance of bacteria, including *Staphylococci*, *Streptococci*, and *Enterococci*, in the bloodstream and their consequent colonization in the heart. However, *Streptococcal* bloodstream infections (BSIs) are considered the most frequent causes of IE. Indeed, oral microbiota is considered a crucial factor associated to the IE progression and commensal oral microbiota strains from *Streptococcus gordonii* and *Streptococcus mitis*, typically associated with caries and dental plaques, are frequently found in patients affected by IE. The incidence of this disease has remained unchanged over the past 30 years. Antibiotic administration is generally used although patients can develop antibiotic resistance and the treatment can therefore fail. In these cases, or when the infection is uncontrolled, resorting to surgery is required. Thus, new measures for IE prevention and reduced antibiotic treatments or surgery intervention have to be considered.¹⁸⁰ Adherence and colonization of Gram positive bacteria into the bloodstream is usually mediated by the presence of surface adhesins that interact with host glycoproteins found on the surface of damaged valves.^{181,182} Streptococcal species involved in the IE etiology often contain Siglec-like adhesins, serine-rich repeat glycoproteins (SRRPs) that recognize carbohydrates containing terminal sialic acid moiety on human salivary mucins and platelets (see Chapter I, §1.5.1).^{107,113} Since the mechanisms of sialoglycan recognition and binding by Siglec-like adhesins have not been determined, we here investigated at molecular level the binding regions of SLBR-B and SLBR-H from *S. gordonii* strains M99 and DL1, respectively, to different α -2,3-sialoglycans, including the different affinity and binding specificity,

together with the dynamic range of conformations adopted by the SLBR–sialoglycan complexes.

5.2 Steady-state fluorescence analysis

The binding profile of SLBR-B and SLBR-H with sialoglycans was first investigated by steady-state fluorescence analysis, which provided values of dissociation constant K_D . Both proteins were titrated with sialyl-T-antigen linked to threonine to recall the *O*-glycan. Furthermore, given the higher selectivity of SLBR-H, we also characterized the binding between SLBR-N and 3'-sialylactosamine (3'SL_n). Overall, the dissociation constant values were all in the micromolar range (figure 5.1).

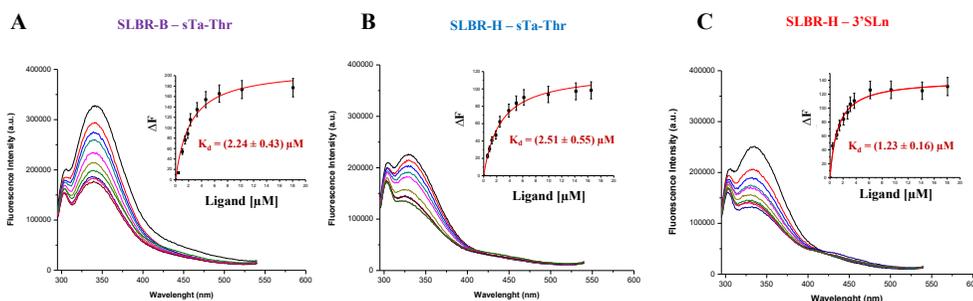


Figure 5.1. Fluorescence analysis of Siglec-like adhesins and sialoglycans for the determination of dissociation constant (K_d) using one site binding analysis. A) Titration of SLBR-B with sTa-Thr. B) Titration of SLBR-H with sTa-Thr. C) Titration of SLBR-H with 3'SL_n. For each data point, 10% Y error bars were shown. The insets show the non-linear fits for the titrations with R^2 values of 0.98, 0.97 and 0.96 respectively.

5.3 Molecular binding of sTa-Thr to SLBR-H

STD NMR experiments were performed on SLBR-H and sTa-Thr. Several STD enhancements were detected in the STD spectrum (figure 5.2) meaning that the three saccharide units of sTa-Thr were involved in the binding with SLBR-H. The highest signals were attributed to H7 Neu5Ac and H4 and H6 Gal, exhibiting STD% above 80%. The other protons of these residues also received a good magnetization from the

protein, except for the diastereotopic (axial and equatorial) H3 of Neu5Ac and H2 of Gal, that provided lower STD values. The GalNAc residue contributed to the binding process to a lesser extent.

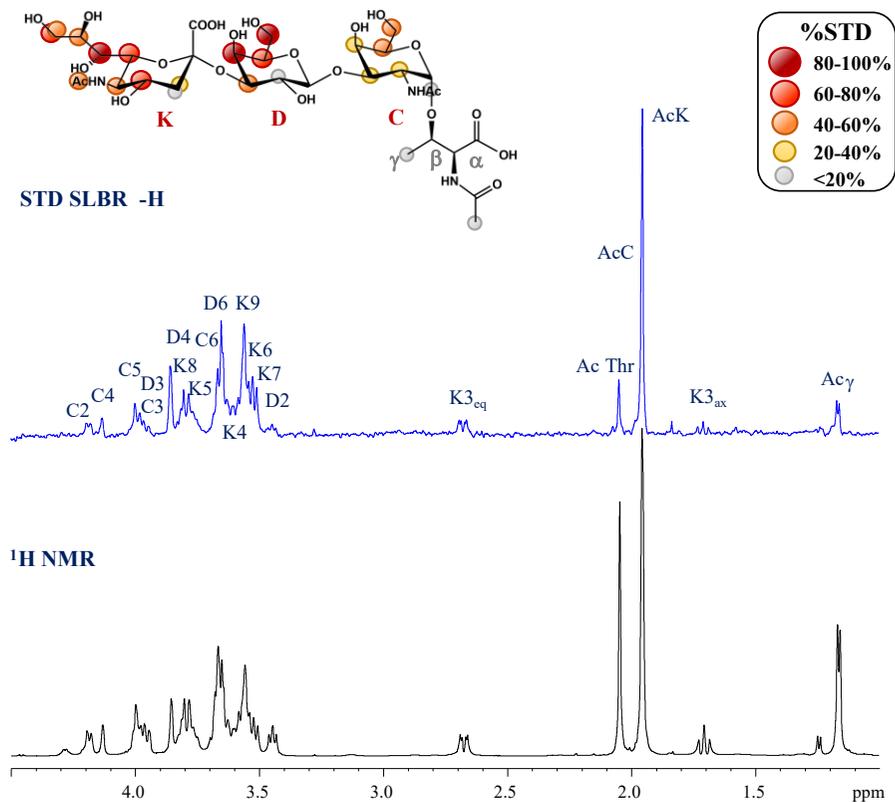


Figure 5.2. STD NMR analysis of SLBR-H and sTa-Thr. The epitope mapping of the ligand revealed that all the residues contributed to the binding process, with the threonine portion more solvent exposed.

To understand how the ligand was placed into SLBR-H, tr-NOESY and molecular modelling were complemented with STD NMR data. Based on the PDB structure (6EFD), sTa-Thr was manually docked into the protein binding site, the complex was minimized and subjected to MD simulation (figure 5.3). Comparing the free and bound states of the torsion angles around the glycosidic linkages, no significant

differences were detected, except for the ϕ (C1-C2-O-C3') around Neu5Ac and Gal residues, that exclusively selected a ϕ value of -60° upon binding (figure 5.3 A and C).

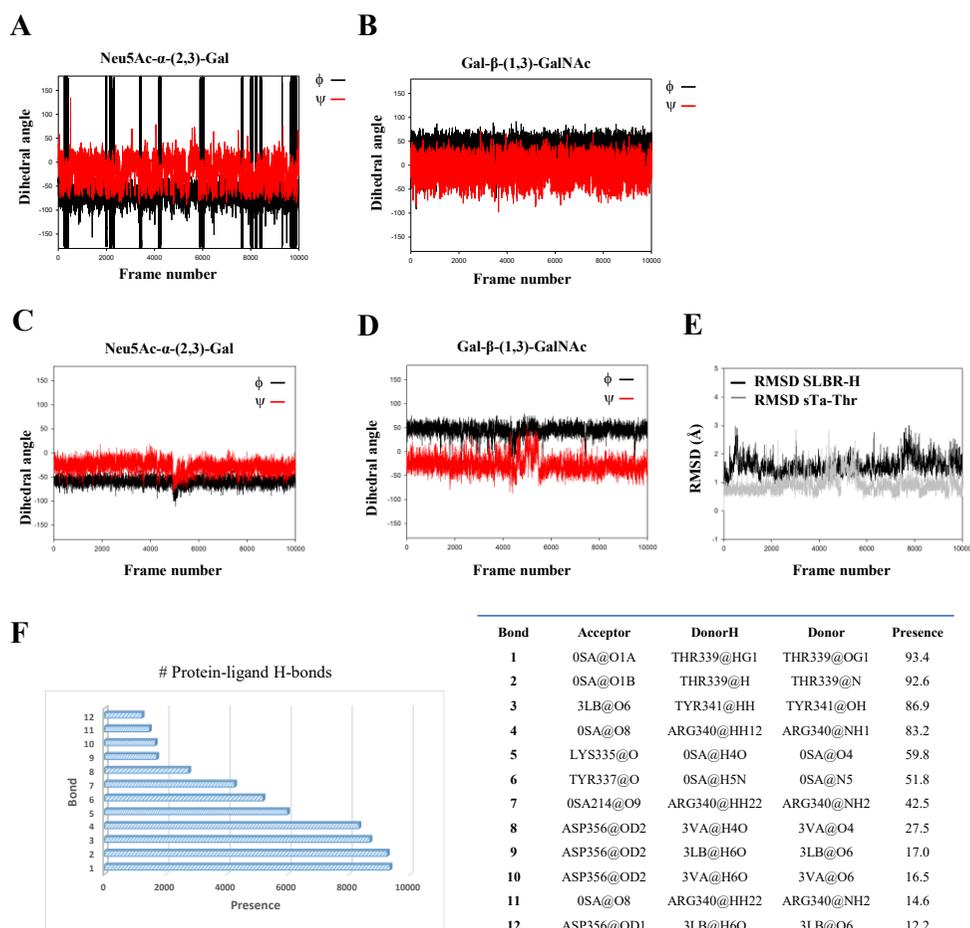


Figure 5.3. MD analysis of SLBR-H and sTa-Thr. Torsion angles monitored in the free (A-B) and bound (C-D) states. E) RMSD trend of SLBR-H and sTa-Thr. F) Description of the main H-bonds interactions monitored along the MD simulation.

Since the NOESY spectrum of the ligand alone was close to zero, a TROESY experiment was acquired. The comparison of TROESY and tr-NOESY (figure 5.4)

spectra also confirmed the similar behavior of the ligand in absence and in presence of SLBR-H since no significant changes were detected.

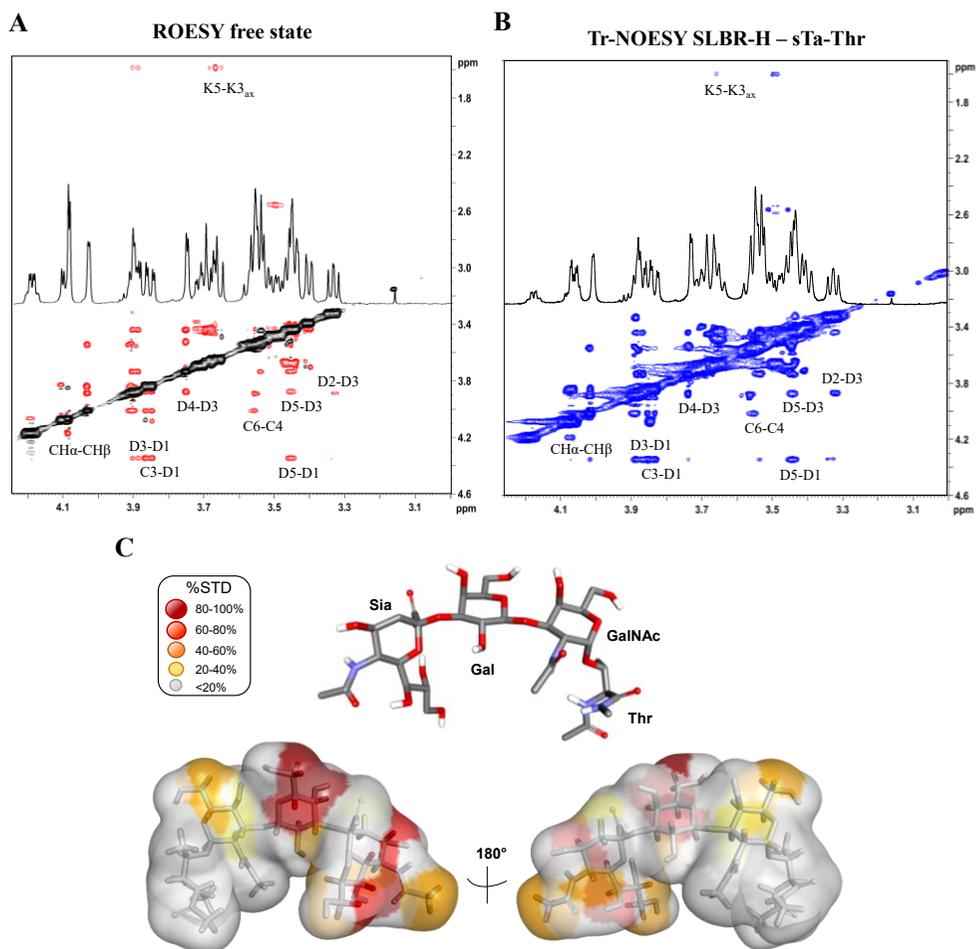


Figure 5.4. Conformational analysis of sTa-Thr in the free state and bound to SLBR-H. A) TROESY spectrum of sTa-Thr alone. B) Tr-NOESY spectrum of sTa-Thr bound to SLBR-H. C) Bioactive conformation of sTa-Thr with the protons colored according to STD NMR results.

RMSD profile monitored along the simulation (figure 5.3 D) showed the stability of the complex. Regarding the network of interactions at the protein-glycan interface

(figures 5.3 F and 5.5), numerous H-bonds were monitored, mainly involving the amino acids of the Φ TRY motif of the F-strand.

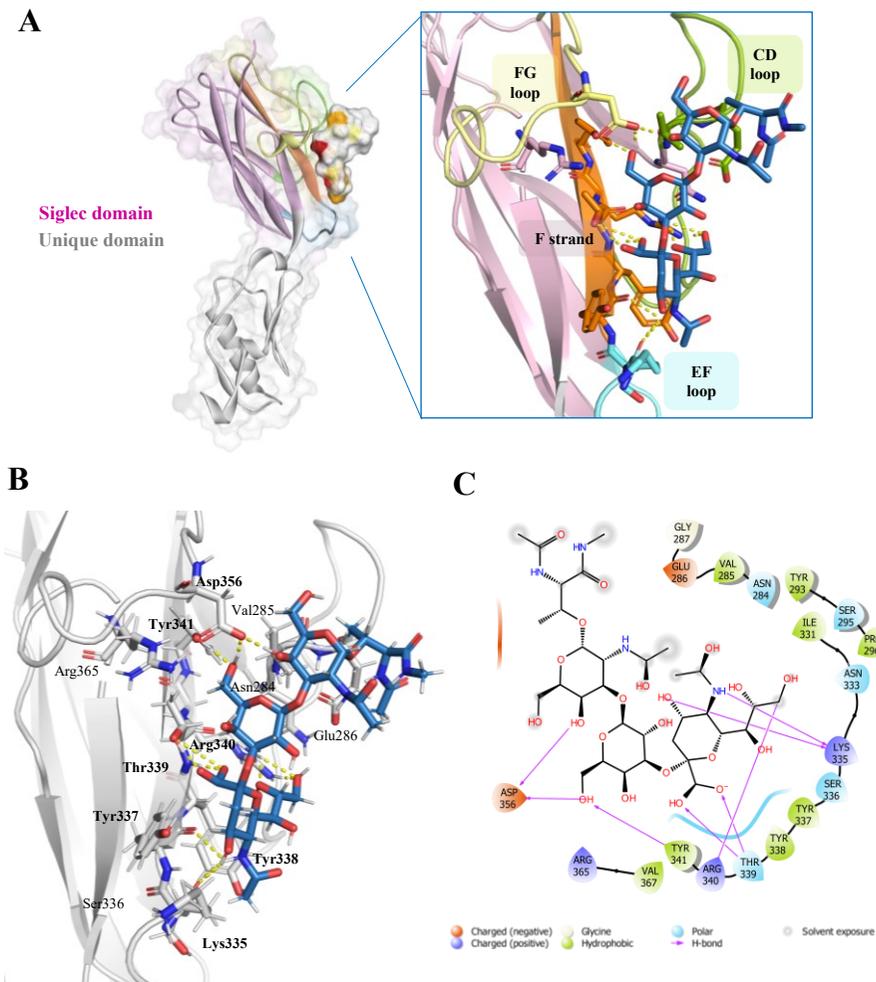


Figure 5.5. Description of SLBR-H and sTa-Thr complex. A) 3D view of the best structure selected by cluster analysis, highlighting the CD, EF and FG loops involved in the binding. B) 3D complex describing the interactions at protein-ligand interface. C) 2D diagram of interactions between SLBR-H and sTa-Thr.

In particular, Thr339 established the most recurrent interactions with the carboxylate group of Neu5Ac, as also the guanidinium group of Arg340 with O8 and O9 Neu5Ac. An important H-bond was detected between Tyr341 hydroxyl group and O6 Gal, the latter also interacted with Asp356 carboxylate group, explaining the high response in the STD spectrum (figure 5.2). Regarding the other atoms of sialic acid, OH at position 4 of Neu5Ac bound to the backbone carboxyl group of Lys335, and the amide group at C5 interacted with Tyr337 carboxyl group. According to the low STD responses (figure 5.2), less recurrent H-bonds between OH at positions 4 and 6 of GalNAc were monitored with Asp356 (figure 5.3 F). Additionally, the aromatic Tyr338 and Tyr341 amino acids, belonging to the Φ TRY consensus motif, were involved in pi-alkyl interactions with positions 8 of Neu5Ac and 6 of Gal, respectively. Furthermore, hydrophobic interactions between H7 Neu5Ac and the proximal CH of Tyr338 likely explained the highest STD response in the spectrum (figure 5.2).

5.4 Molecular binding of sTa-Thr to SLBR-B

The molecular interaction of SLBR-B and sTa-Thr was investigated by a combination of NMR and MD simulation. As previously shown by fluorescence analysis (figure 5.1), the recognition of the sTa-Thr by SLBR-B was comparable to SLBR-H. Indeed, STD NMR indicated that all the residues of sTa-Thr were involved in the interaction and thus gave STD signals in the spectrum (figure 5.6). The sialic acid was accommodated in the proteins binding site, and gave the highest STD enhancements, as found for H7 and H6.

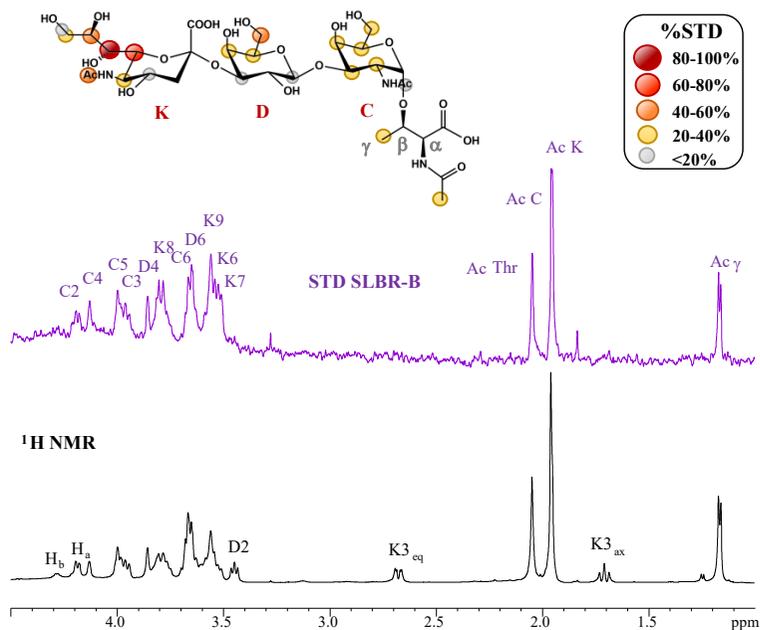


Figure 5.6. STD NMR analysis of SLBR-B and sTa-Thr. The epitope mapping revealed an involvement of all residues in the recognition with SLBR-B.

The acetyl group and the glycerol chain of Neu5Ac also received a good magnetization from SLBR-B, showing STD% around 40%. The other protons of the ligand contributed to the binding with STD% below 40% (figure 5.6). Based on the crystal structure (PDB: 5IUC), sTa-Thr was docked into SLBR-B to run MD simulations (figure 5.7).

Overall, the complex resulted stable along the simulation (figure 5.7 A-C), with the ligand conformation unchanged with respect to the free state (figure 5.3 A and B).

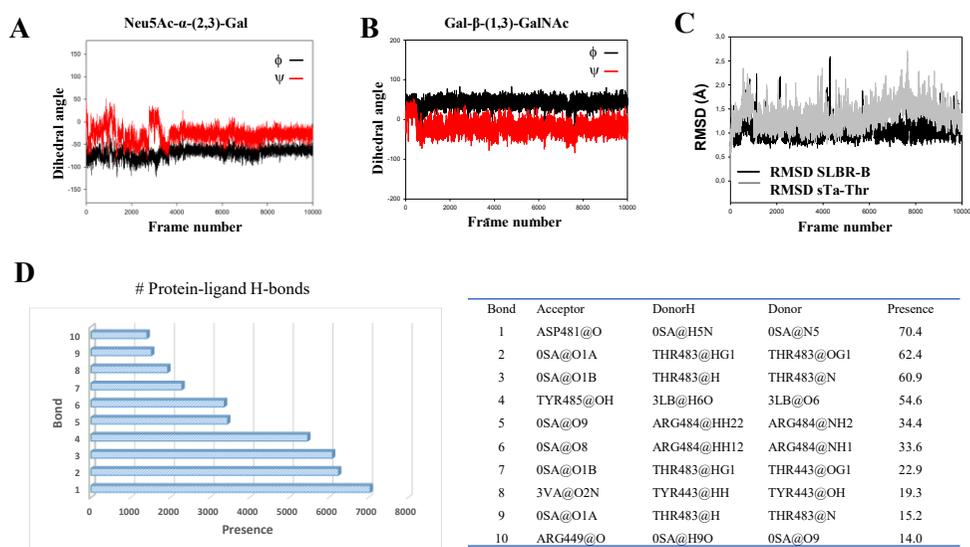


Figure 5.7. MD analysis of SLBR-B and sTa-Thr. Torsion angles monitored in the bound (A-B) state. C) RMSD trend of SLBR-B and sTa-Thr. D) Description of the main H-bonds interactions monitored along the simulation.

Indeed, although ϕ torsion angle around Neu5Ac and Gal linkage could populate two minima (-60° and 180° , figure 5.3 A), the preference for ϕ of -60° was already observed in the free state. For this reason, no changes in NOE contacts were detected (figure 5.8).

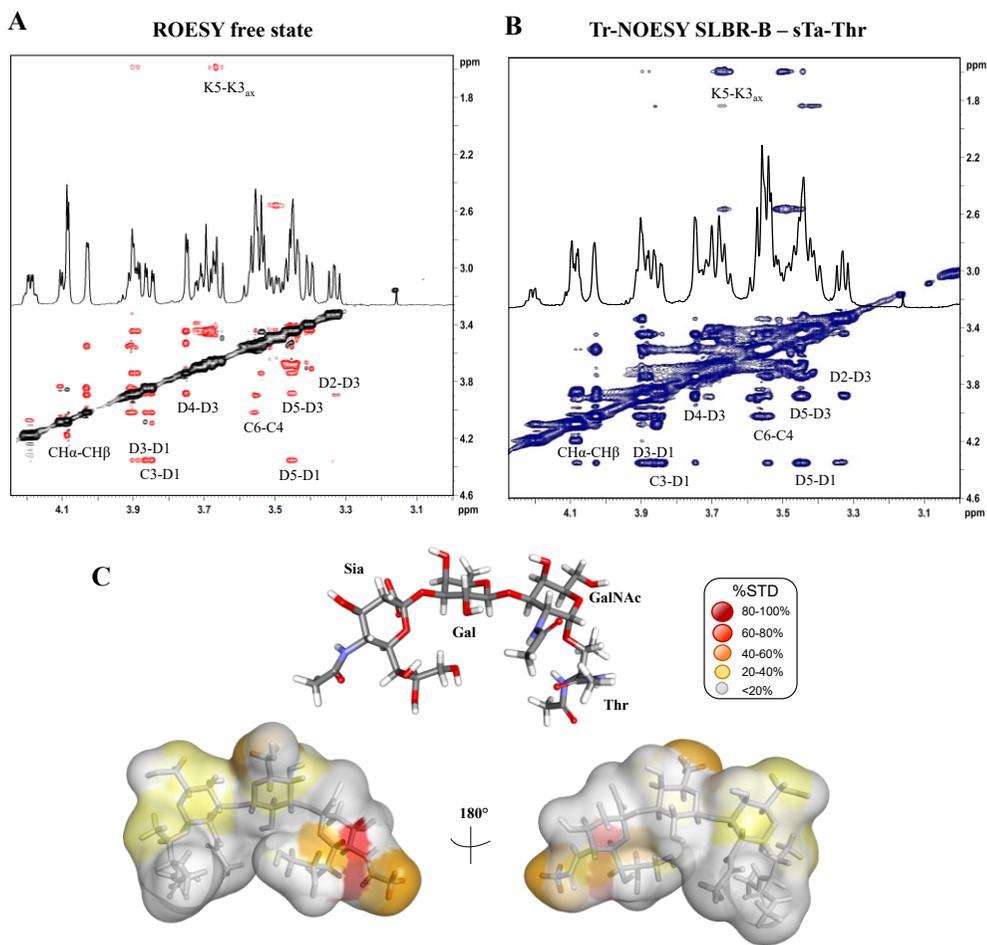
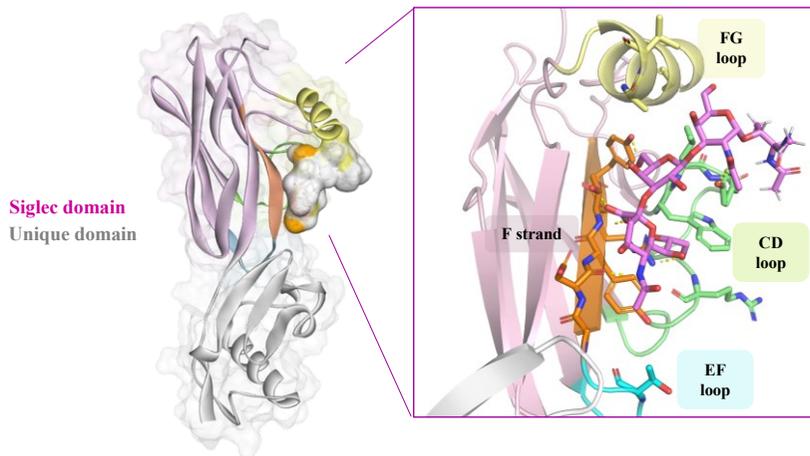


Figure 5.8. Conformational analysis of sTa-Thr in the free state and bound to SLBR-B. A) TROESY spectrum of sTa-Thr alone. B) Tr-NOESY spectrum of sTa-Thr bound to SLBR-B. C) Bioactive conformation of sTa-Thr with the protons colored according to STD NMR results.

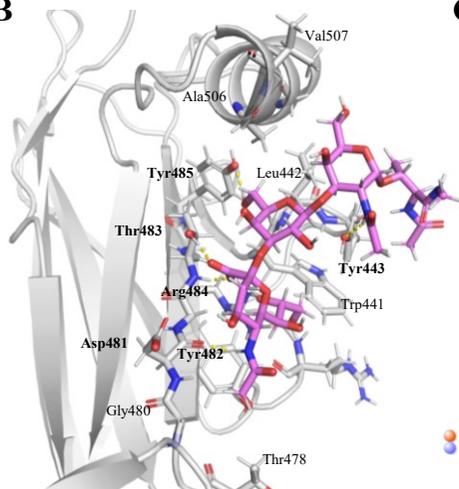
Although RMSD profile of the MD simulation showed a good 3D complex formation, the polar interactions occurring at protein-ligand interface were not retained for the entire simulation (figure 5.7 D). The main contacts were found with sialic acid, while

Gal and GalNAc were less recognized by SLBR-B, supporting the STD data (figure 5.6).

A



B



C

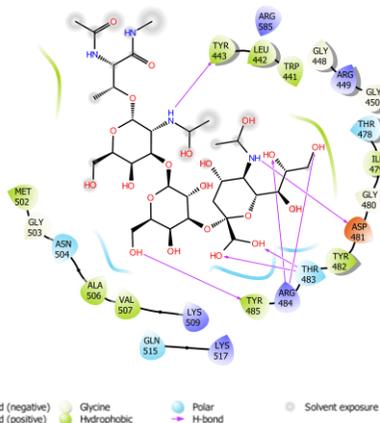


Figure 5.9. Description of SLBR-B and sTa-Thr complex. A) 3D view of the best structure selected by cluster analysis, highlighting the CD, EF and FG loops involved in the binding. B) 3D complex describing the interactions at protein-ligand interface. C) 2D diagram of interactions between SLBR-B and sTa-Thr.

Among the main contacts, the amide and carboxyl groups of Neu5Ac established H-bonds with Asp481 and Thr483, respectively (figures 5.7 D and 5.9), and OH at position 6 of Gal interacted with Tyr485. No significant contacts were found for GalNAc, that indeed showed lower STD responses (figure 5.6). The lower contribution of Gal and GalNAc residues was likely due to the interactions between the threonine moiety of the ligand and the α -helix of SLBR-B FG loop (figure 5.9 A and B), which might contribute to Gal and GalNAc units movement from the protein surface.

5.5 Molecular binding of 3'SL_n to SLBR-H

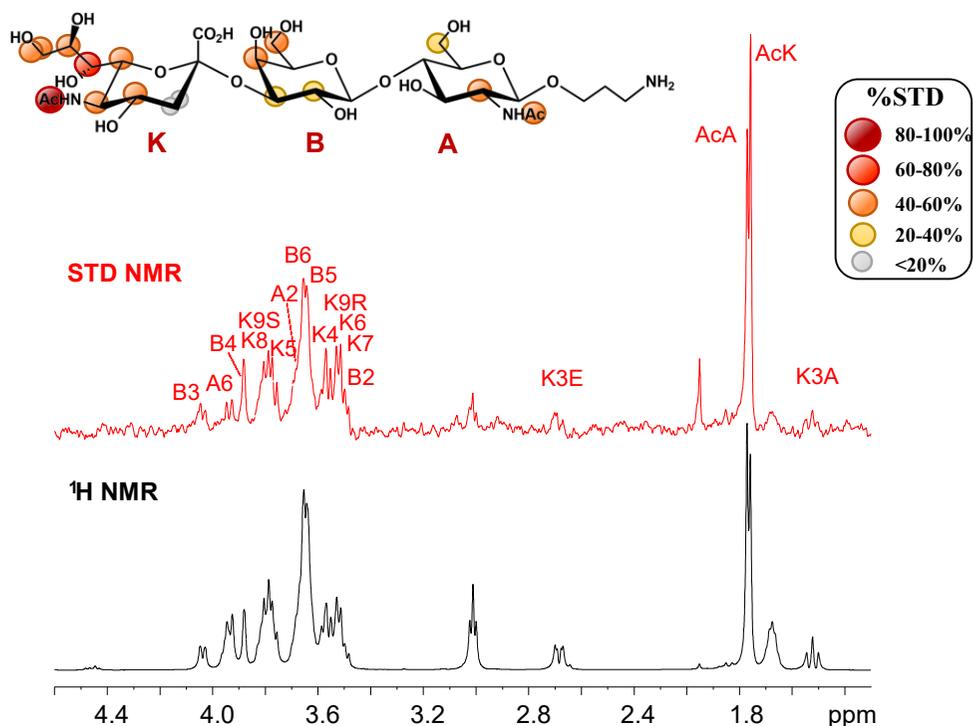


Figure 5.10. STD NMR analysis of the 3'SL_n recognized by SLBR-H.

Due to the ample selectivity of SLBR-H towards different sialoglycans, we also investigated the molecular binding of this Siglec-like adhesin with 3'SLN. As expected, STD NMR revealed a complete recognition of the ligand by SLBR-H, with high STD signals compared to the off-resonance (figure 5.10).

The epitope mapping indicated the sialic acid as the residue most involved in the interaction. Indeed, the highest STD signal was attributed to the acetyl group; important effects were also found for the glycerol chain and for H4, H5 and H6 of Neu5Ac. A good magnetization transfer also affected H4 and H6 Gal as well as H2 and the acetyl group of GlcNAc.

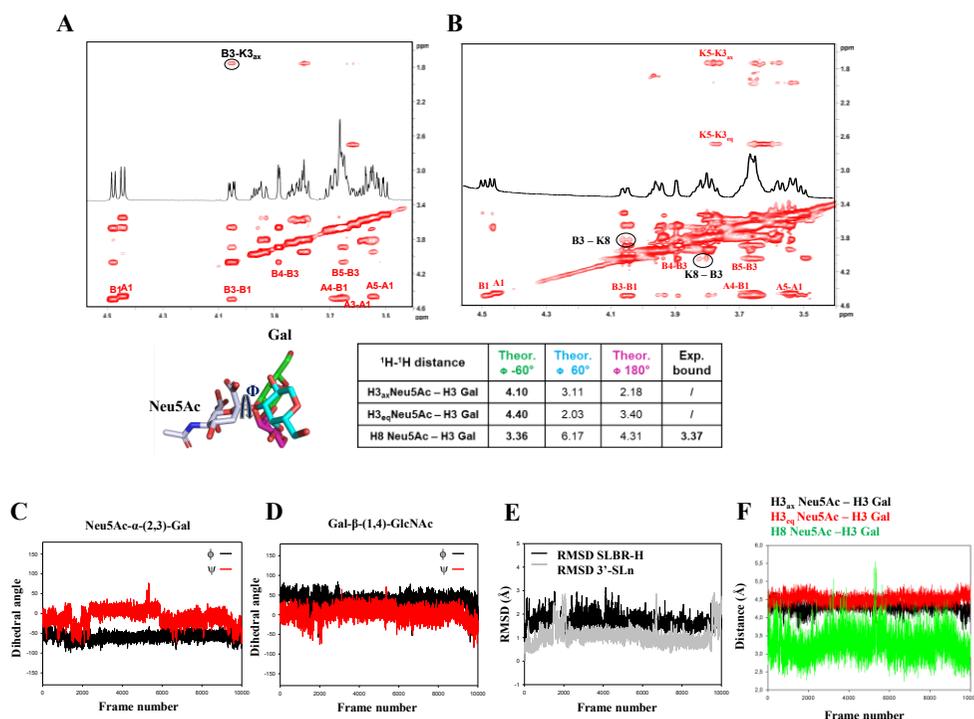


Figure 5.11. Conformational analysis of 3'SLN in the free state and bound to SLBR-H. A) NOESY and B) tr-NOESY spectra. The presence of key NOE B3-K8 and the absence of B3 with diastereotopic K3 (axial and equatorial) in the bound state suggested the conformer selection for ϕ around Neu5Ac and Gal of -60° . C-D) Torsion angles monitored along 100 ns

MD simulation in the bound state. E) RMSD profile of the 3D complex. F) Inter-nuclear distances corresponding to the key NOE contacts monitored during MD simulation.

Regarding the conformational behavior of the ligand (figure 5.11) in the free state, an equilibrium of three different conformations, corresponding to φ $-60^\circ/60^\circ/180^\circ$ around the Neu5Ac- α -(2,3)-Gal glycosidic linkage, was established (table in figure 5.11).⁹⁵ Upon binding, tr-NOESY (figure 5.11 B) data, confirmed by MD simulation (figure 5.11 C-F), displayed a bioactive conformation where 3'SL_n preferentially adopted a φ torsion angle around -60° . Indeed, the key NOE correlation between H3 Gal and H8 Neu5Ac (B3-K8), with the corresponding *inter*-proton distance, and the absence of the NOE contacts between H3 Gal with the diastereotopic H3 protons of Neu5Ac were indicative of a conformer selection upon binding, as also monitored by MD simulation (figure 5.11 C-F).

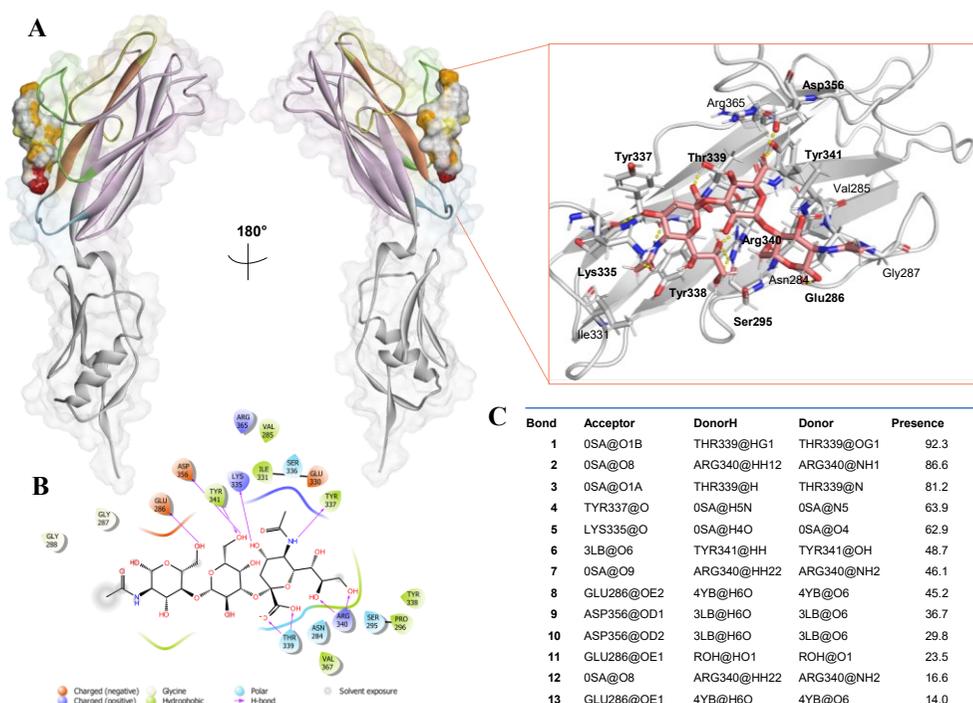


Figure 5.12. Description of SLBR-H and 3'SL_n complex. A) 3D view of the best structure selected by cluster analysis from MD simulation, describing the interactions at protein-ligand interface. B) 2D diagram of interactions between SLBR-H and 3'SL_n. C) Description of the main H-bonds interactions monitored along the simulation.

The 3D complex of SLBR-H and 3'SL_n showed a good stability along MD trajectory (figure 5.11 E); a network of various interactions was monitored (figure 5.12), allowing to define an extended and polar binding site of SLBR-H. The most stable contacts were found for the amino acids of the consensus sequence (figure 5.12 C), especially Thr339 and Arg340 with the carboxyl group and glycerol chain of Neu5Ac, reinforcing binding and tuning the ligand orientation into SLBR-H (figure 5.12 A) binding site. Interestingly, Tyr338 played a fundamental role in 3'SL_n accommodation, establishing hydrophobic interaction with H7 Neu5Ac, that gave high STD response (figure 5.10). As for the galactose, O6 interacted with Tyr341 of the consensus sequence, while the proton at the same position made an H-bond with the carboxylate of Asp356. Moreover, OH at position 6 of GlcNAc established contacts with Glu286, according to STD data.

Overall, STD NMR and MD results indicated a complete accommodation of 3'SL_n into SLBR-H binding pocket.

Given the polar nature of binding site of SLBR-H and the importance of water molecules in mediating protein-ligand interactions, WaterLOGSY experiments were further employed. Although these experiments on the previous systems did not give additional information on the interactions (data not shown), a general decrease of almost all proton signals was observed for 3'SL_n bound to SLBR-H with respect to the WaterLOGSY (WL) acquired in the free state. Moreover, a positive signal referring to the H9 Neu5Ac was indicative of the existence of water molecules specifically surrounding this position of 3'SL_n (figure 5.13).

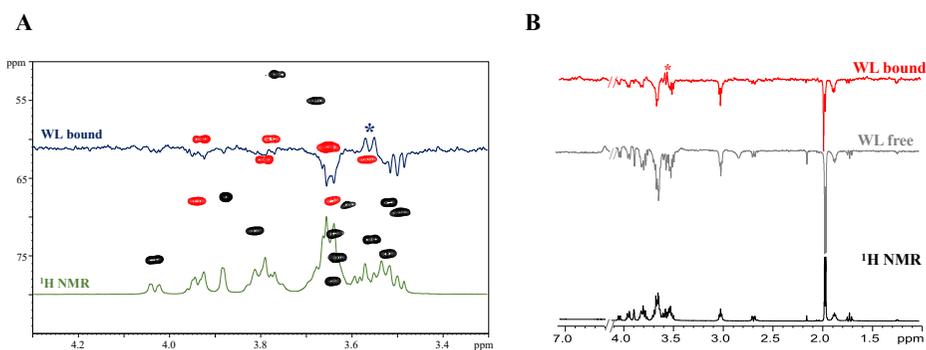


Figure 5.13. WaterLOGSY experiments on SLBR-H and 3'SL_n. A) HSQC and WL experiments of 3'SL_n interacting with SLBR-H. B) Comparison of the WL experiments in the free (grey) and bound (red) states.

A detailed analysis of MD simulation confirmed the presence of a water density stably surrounding OH9 proton of sialic acid (figure 5.14 A). A specific water molecule that acted as bridge between 3'SL_n and SLBR-H was found, thus mediating an H-bond between OH at C9 of Neu5Ac and OH of Ser295 (figure 5.14 B). The stability of these H-bonds was also monitored along the MD simulation, by measuring the distances between the atoms involved (figure 5.14 C).

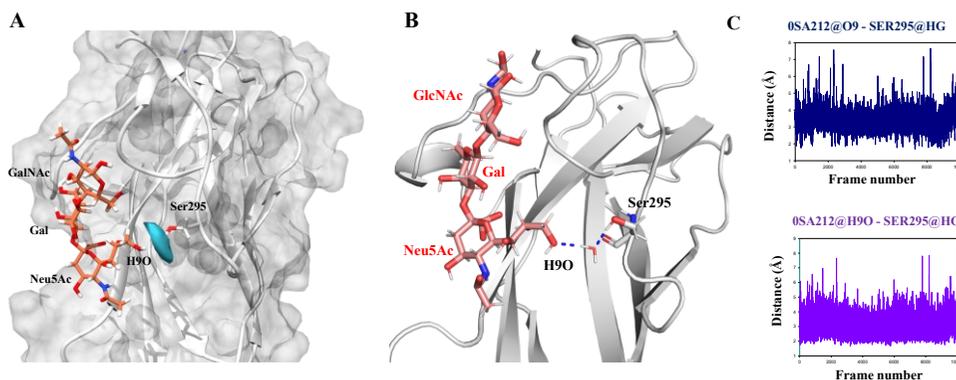


Figure 5.14. Computational analysis implemented with waterLOGSY results in order to define the water molecule serving as bridge between Ser295 of SLBR-H and the atom at position 9 of sialic acid of 3'SL_n. A-B) 3D view of SLBR-H and 3'SL_n complex, highlighting the water

molecule at protein-ligand interface. C) Stability of the distances monitored among the protein's and ligand's atoms involved in the interaction with the water molecule.

5.6 The ganglioside GM1b as novel ligand for SLBR-H and SLBR-B

Since GM1b is a ganglioside containing the Neu5Ac- α -(2,3)-Gal- β -(1,3)-GalNAc epitope, its interaction with both Siglec-like adhesins was explored.

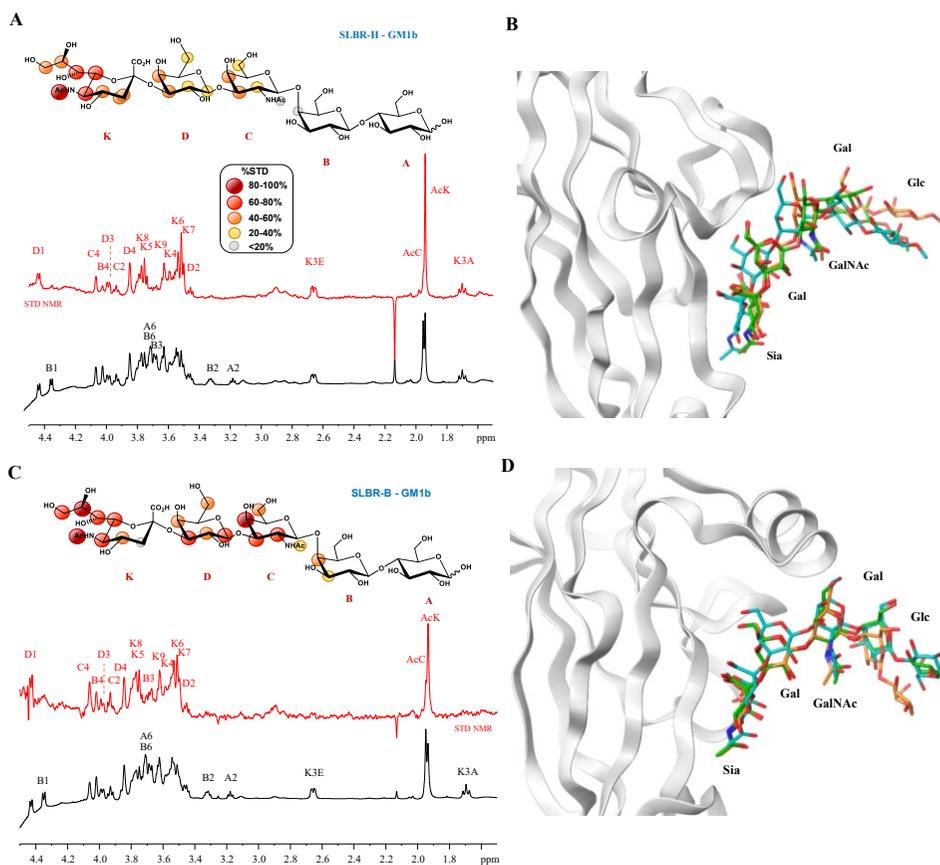


Figure 5.15. Analysis of GM1b interacting with Siglec-like adhesins from *S. gordonii*. A) STD NMR and epitope mapping of GM1b with SLBR-H. B) 3D complex of SLBR-H and GM1b extracted by MD simulation. C) STD NMR and epitope mapping of GM1b with SLBR-B. D) 3D complex of SLBR-B and GM1b extracted by MD simulation.

Interestingly, SLBR-H and also of the more specific SLBR-B were able to recognize the ganglioside, with the portion mostly involved in the binding corresponding to the sTa epitope (figure 5.15).

STD NMR analysis showed the acetyl group of Neu5Ac (K unit) as the main signal in the spectra; STD responses were also detected for Gal (D) and GalNAc (C) residues; Gal (B) and Glc (A) residues were instead more solvent exposed (figure 5.15 A and C). As illustrated by the 3D complexes (figure 5.15 B and D), the higher STD responses detected for GM1b GalNAc (C residue) when bound to SLBR-B rather than SLBR-H (figure 5.15 A and C) were likely ascribable to its proximity to SLBR-B α -helix (figure 5.15 D).

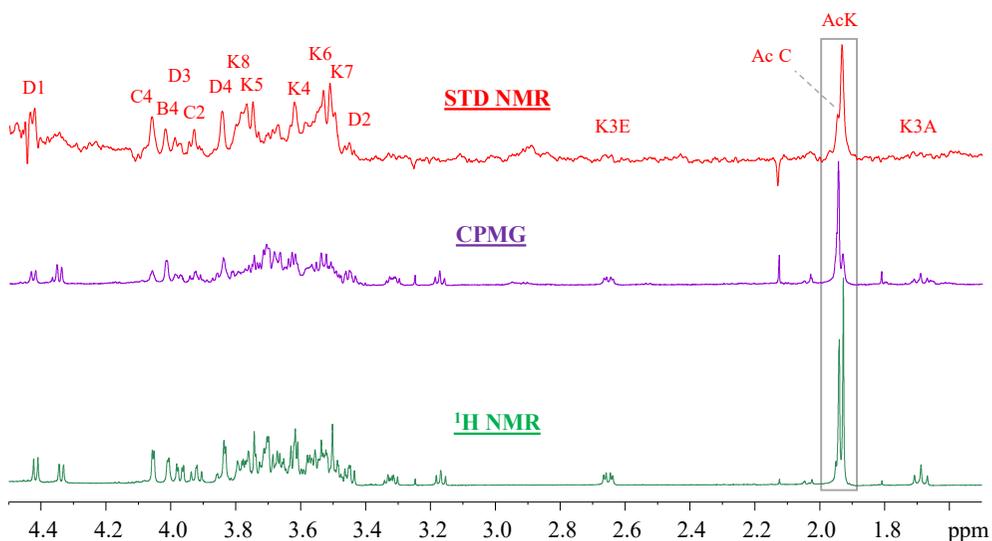


Figure 5.16. Carr Purcell Meimboom Gill NMR experiment of SLBR-B and GM1b. The protons that decreased their relative intensity in the CPMG (purple) for the T2 relaxation effect, as indicated for the acetyl groups around 2 ppm, increased their intensity in the STD (red) at the same time.

Carr Purcell Meimboom Gill (CPMG) relaxation-time-edited NMR experiments further confirmed the interactions between GM1b and SLBR-B (figure 5.16). The

decrease of several signals in CPMG spectra passing from the free (green spectrum) to the bound state (purple spectrum) indicated a change in transverse relaxation time (T_2) of the ganglioside upon binding with SLBR-B. Consequently, the lower CPMG signals in the bound state represented the protons closer to the binding pocket of the protein. Thus, the complementary CPMG and STD NMR techniques supported the recognition of GM1b by SLBR-B. Moreover, the signals showing decreased intensity in the CPMG experiment corresponded to the protons that gave the higher STD effects (figure 5.16).

5.7 Discussion

The role of Siglec-like adhesins in bacterial colonization of heart and consequent pathogenesis of infective endocarditis has been demonstrated *in vitro* and *in vivo*.¹¹² Previous site-directed mutagenesis of amino acids of SLBR-B and SLBR-H crucial for the binding, such as R484E^{SLBR-B}, Y485F^{SLBR-B} and T339V^{SLBR-H} or R340E^{SLBR-H} belonging to the consensus sequence, resulted in significant reduction of the infection.¹¹³ However, the description of sialoglycans binding to Siglec-like adhesins was not determined yet. Thus, we proposed a global view of the interaction of SLBR-H and SLBR-B, Siglec-like adhesins respectively expressed on DL1 and M99 *S. gordonii* surface, with different sialoglycans usually exposed on MG2/MUC7 salivary mucins and/or human platelet glycoprotein GPIb α .¹⁸³ The analysis of glycans' epitope and conformational behavior, and the description of 3D complexes allowed to describe the structural and topological features driving recognition and binding processes. To simulate the *O*-glycan typically exposed on platelets or salivary glycoproteins, we evaluated the binding between both SLBR-H and SLBR-B with sTa linked to threonine (sTa-Thr), revealing a comparable recognition, with the main involvement of the sugar portion and Thr more solvent exposed. Our results are consistent with streptococcal adherence to host O-glycoproteins, whose sugar portion

is linked to a protein backbone (figure 5.17), that can influence the entire mucin conformations. For example, the presence of proline residues to Ser or Thr, often found in O-glycosylation sites, characterizes the glycopeptide shape that assumes an extended rod-like structure form.¹⁸⁴ Thus, the higher rigidity of threonine linked-*O*-glycans with respect to serine linked-*O*-glycans toward the peptide backbone may favor the interaction with adhesins like SLBR-H and SLBR-B.

Regarding the comparison of the contacts of sTa-Thr into SLBR-H and SLBR-B, we could affirm that the residues of the consensus sequence exhibited the same interactions with the ligand. However, the amide group of Neu5Ac established H-bonds with Asp481 in SLBR-B and with Tyr337 in SLBR-H, both amino acids belonging to the F-strand. Another difference was found in the recognition of O6 GalNAc, that interacted with Tyr443 of the CD loop in SLBR-B, and with Asp356 of the FG loop in SLBR-H (figures 5.5 and 5.9).

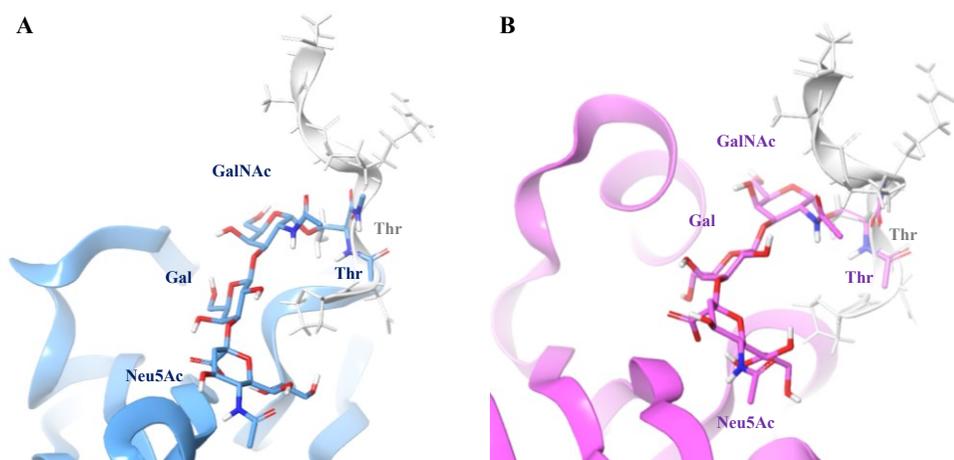


Figure 5.17. Superimposition of SLBR-H (A) and SLBR-B (B) with sTa-Thr (colored structures) to a generic *O*-glycan expressed on mucins (in grey, structure PDB: 5T78). The orientation assumed by Thr toward the solvent in both complexes was coherent with the presence of a potential amino acid backbone linked to the sTa usually found in natural *O*-glycans.

We also inspected the molecular binding of both adhesins to a longer ligand containing sTa epitope, the ganglioside GM1b, as new potential ligand (5.15). Indeed, despite ligands repertoire recognized by Siglec-like adhesins have been described,^{115,117} the possibility to bind other glycan substrates, like gangliosides, ubiquitous in many tissues (e.g., brain, gut and vessels), cannot be excluded (yet).^{185,186,187}

Furthermore, given the broader selectivity of SLBR-H, we then characterized its binding with 3'-SLn, identifying a specific conformation adopted by the ligand upon binding and discovering the presence of resident water molecules at protein-ligand interface (figures 5.11 and 5.14).

Overall, although similar in the V-set Ig domains architecture in terms of loops and β -sheets, mammalian Siglecs and bacterial Siglec-like adhesins showed different binding mode towards glycans. The importance of the CD, EF and FG loops conformation in SLBR-H and SLBR- B structures was widely defined in tuning the selectivity of ligands, mechanistically different from binding via the CC' loop of the analogous sialic acid-binding domain in mammalian Siglecs. Comparing the epitope maps of α 2,3 sialoglycans recognized by Siglec-like adhesins, SLBR-H and SLBR-B, and Siglec-10 (previously studied by our group),⁹⁵ we observed a different accommodation of the ligand into the proteins' binding site (figure 5.18).

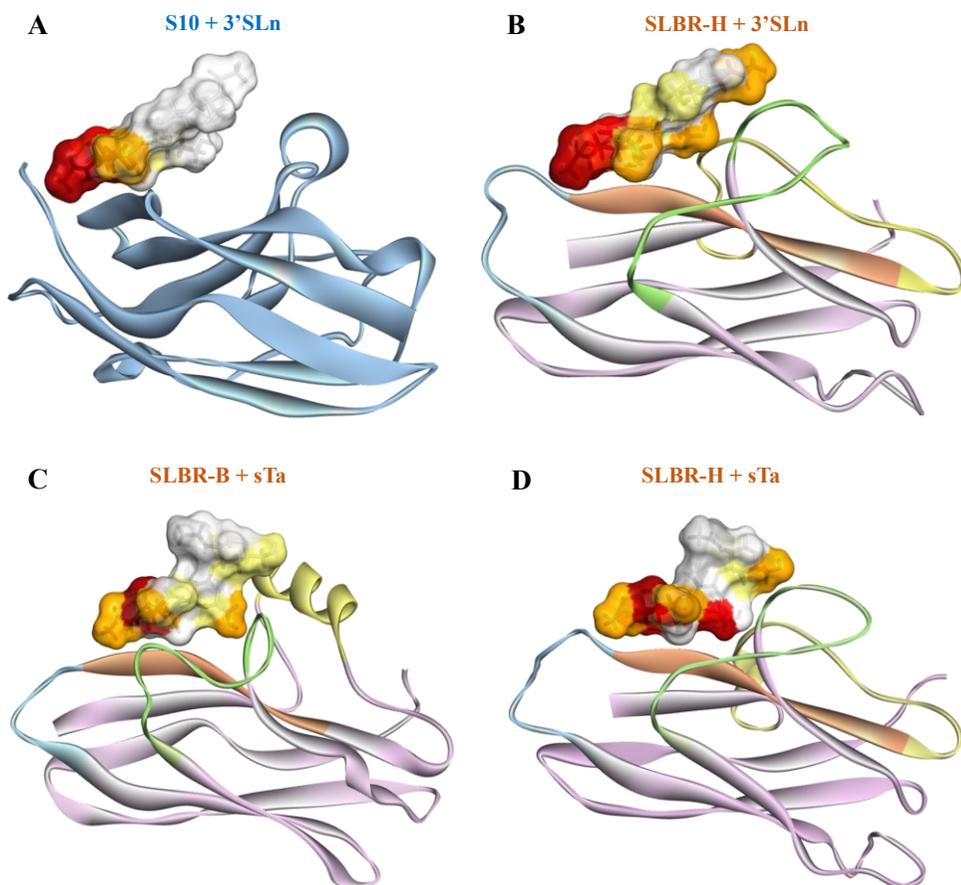


Figure 5.18. Comparison of α 2,3 sialoglycans into Siglec and Siglec-like adhesin binding sites. A) Epitope mapping of 3'SL_n bound to Siglec-10, as example of the binding mode of Siglecs. B) Epitope mapping of 3'SL_n bound to SLBR-H. C) Epitope mapping of sTa bound to SLBR-B. D) Epitope mapping of sTa bound to SLBR-H.

Indeed, while all the residues of the sugar were recognized by all Siglec-like adhesins (figure 5.18 B-D), 3'SL_n was anchored to Siglec-10 through contacts with sialic acid and a partially with galactose, while the third sugar moiety was completely excluded from the binding pocket. This kind of recognition was observed for all Siglecs, interacting with also different sialoglycans (figure 5.19).

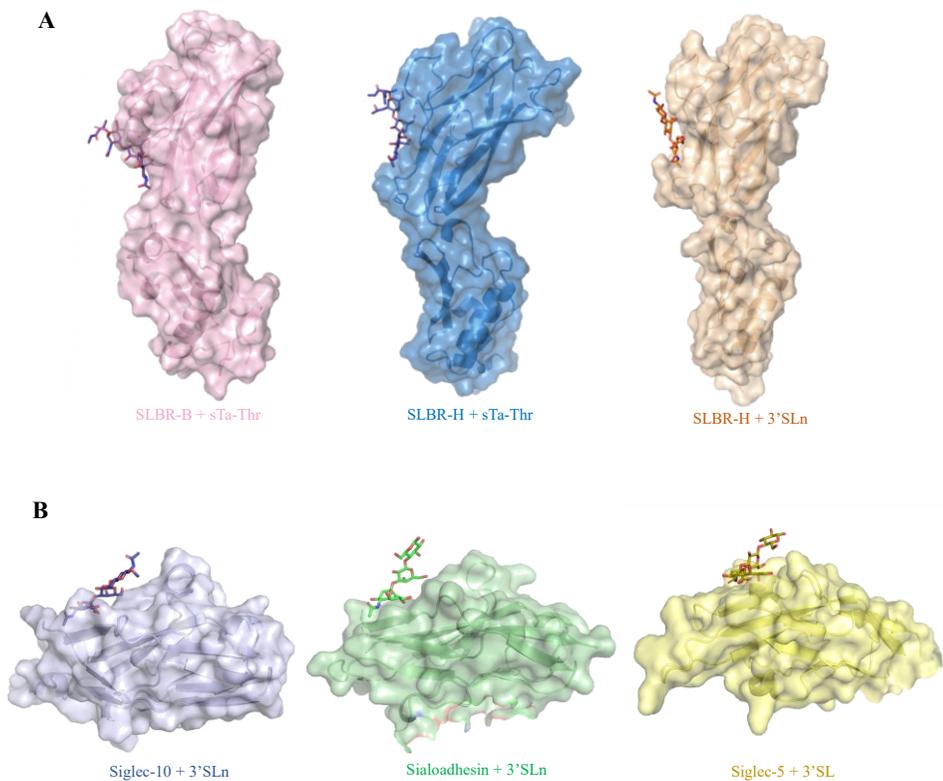


Figure 5.19. Comparison between Siglec-like adhesin and Siglec binding sites. A) 3D view of the structures: SLBR-B and sTa-Thr (pink), SLBR-H and sTa-Thr (blue), SLBR-H and 3'SLn (wheat). In all the complexes, the ligands were completely accommodated into the corresponding protein binding sites. B) 3D view of the structures: Siglec-10 and 3'SLn (light blue), Sialoadhesin and 3'SL (green, PDB: 1QFO), Siglec-5 and 3'SL (yellow, PDB: 2ZG3). In these complexes, sialic acid and galactose were only involved in the binding, while the third sugar was solvent exposed.

The binding mode of Siglec-like adhesins and Siglecs was indeed different. In all Siglecs, for example, a conserved arginine in the F strand was crucial in the formation of salt bridge with the carboxylate of sialic acids. In the Siglec-like adhesins, instead, the carboxyl group of Neu5Ac always established hydrogen bonds with the threonine of the consensus motif (Thr339^{SLBR-H} and Thr483^{SLBR-B}), while the arginine residues

(Arg340^{SLBR-H} and Arg484^{SLBR-B}) interacted with the glycerol moiety of sialic acid. On the other hand, in Siglecs binding site, the glycerol chain of sialic acid interacted with aromatic residues (in particular, a conserved tryptophan) due to the formation of intra-sheet disulfide bond between specific β -strands that allowed to expose the hydrophobic amino acids.⁴⁷ Although aromatic residues were also found in Siglec-like adhesins, causing for example the important interaction with H7 Neu5Ac, the binding pocket of these proteins was mainly polar, due to the presence of several H-bonds at the interface.

Since no vaccine or anti-adhesive drug was approved against IE and given their role in the progression of the infection, Siglec-like adhesins are considered attractive targets for the development of new therapeutics against IE. Within this goal, we provided the molecular mechanisms at the basis of glycans' recognition by the streptococcal adhesins, crucial for the design of novel therapeutics to prevent or treat IE disease, such as the development of specific inhibitors that do not interfere with normal, important Siglec interactions.

Chapter VII:

Molecular binding of different *Staphylococcus aureus* wall teichoic acid glycoforms recognized by specific antibodies

VII. Molecular binding of different *Staphylococcus aureus* wall teichoic acid glycoforms recognized by specific antibodies

7.1 Introduction

Staphylococcus aureus is a Gram-positive bacterium that causes severe infections, including bacteremia, staphylococcal toxic shock syndrome, endocarditis, and osteomyelitis.^{188,189} This pathogen asymptotically colonizes 30% of the human population. In particular, wall teichoic acids (WTAs),¹⁹⁰ cell surface glycopolymers linked to the thick layer of peptidoglycan, contributes to the nasal colonization, by developing invasive diseases (see the structure in Chapter I, § 1.2.2).^{191,192} Given the growing health concern in regard to the diffusion of antibiotic-resistant MRSA (methicillin-resistant *S. aureus*), the development of new non-antibiotic therapeutic approaches is urgently required, as highlighted by the World Health Organization.¹⁹³

Depending on the environmental conditions, WTA-expression on *S. aureus* strains can vary, leading to different structures of these glycopolymers.¹⁹⁴ WTA modifications, recognized by both innate and adaptive immune components, are caused by the action of specific glycosyltransferases, which attach a N-acetylglucosamine (GlcNAc) moiety to the Rbo backbone, at different stereochemistry (α or β) and positions (Chapter I, § 1.2.2.1). These WTA GlcNAc residues represent antigenic epitopes for certain monoclonal antibodies (mAbs),¹⁹⁵ and thus they are considered promising targets for active¹⁹⁶ and passive immunization strategies as well as antibiotic delivery that can halt the infection.¹⁹⁷ Previous studies have shown that the use of a “minimal epitope” constituted by a single RboP monomer with a β -1,4-GlcNAc could be recognized by patient-derived MRSA-targeting mAbs.¹⁹⁸ Nevertheless, information about the molecular binding of mAbs with different glycoforms (β -1,3- and α -1,4-GlcNAc) is currently lacking. In collaboration with Prof. Jeroen Codée (University of Leiden), Prof. Nina M. van Sorge (University

of Amsterdam), Dr. Mark J. van Raaij and Dr. Pablo Soriano Maldonado (Spanish National Research Council), we here explored the binding of two mAbs, *i.e.*, the anti- β -GlcNAc specific mAb 4497 and the anti- α -1,4-GlcNAc-WTA mAb 4461, to previously well-defined synthetic WTA fragments, which vary in RboP-chain length, glyco-type (β -1,3-, β -1,4, and α -1,4), position and number of GlcNAc residues on the RboP backbone.¹⁹⁹ These studies allowed to define the molecular basis behind the cross-reactivity of mAb 4497 that can bind both β -1,4- and β -1,3-GlcNAc moieties on Rbo chain and to identify structural aspects that allow recognition of α -1,4-GlcNAc WTA by mAb 4461.²⁰⁰

7.2 Molecular investigation of IgG mAb 4461 with α -GlcNAc modified wall teichoic acids

The interactions between the monoclonal antibody IgG 4461 with synthetic glycosylated WTA oligomers were explored, demonstrating its ability to specifically bind to RboP WTA decorated with α -1,4-GlcNAc moieties.

7.2.1 IgG mAb 4461 binding to α - WTA-trimer

STD NMR experiments were initially performed on the WTA-trimer (figure 7.1), containing a single GlcNAc residue at position 4 of the internal RboP subunit. As inferred from the intense STD signals belonging to the sugar protons, it was clear the stronger contribution to the interaction of the carbohydrate ring with respect to the ribitol chain, which gave rise to lower STD effects. To identify the ligand protons that were in contact with the IgG mAb 4461 antibody surface, the highest STD enhancement belonging to the acetyl group of GlcNAc was set to 100%, and all the other STD percentages were normalized accordingly (figure 7.1). The protons H3, H4 and H5 of GlcNAc were also significantly recognized by the antibody, with %STD values between 60% and 90%. Moreover, certain transfer of magnetization also

involved H2 and the anomeric proton of GlcNAc, resulting in STD effects in the range of 40-60%. Lower STD responses were instead obtained for H6 protons of GlcNAc, that seemed to point farer from the antibody surface as compared to the other sugar protons. The STD signals coming from the RboP chain were principally attributed to the portion of the “internal” ribitol phosphate unit (indicated as B in figure 7.1), showing STD enhancements around 20-40%. Thus, the STD NMR analysis revealed the importance of the GlcNAc moiety in the binding with IgG mAb 4461, providing the main STD signals in the spectrum; some protons of the RboP chain of WTA-trimer were also recognized by the antibody, but to a lesser extent.

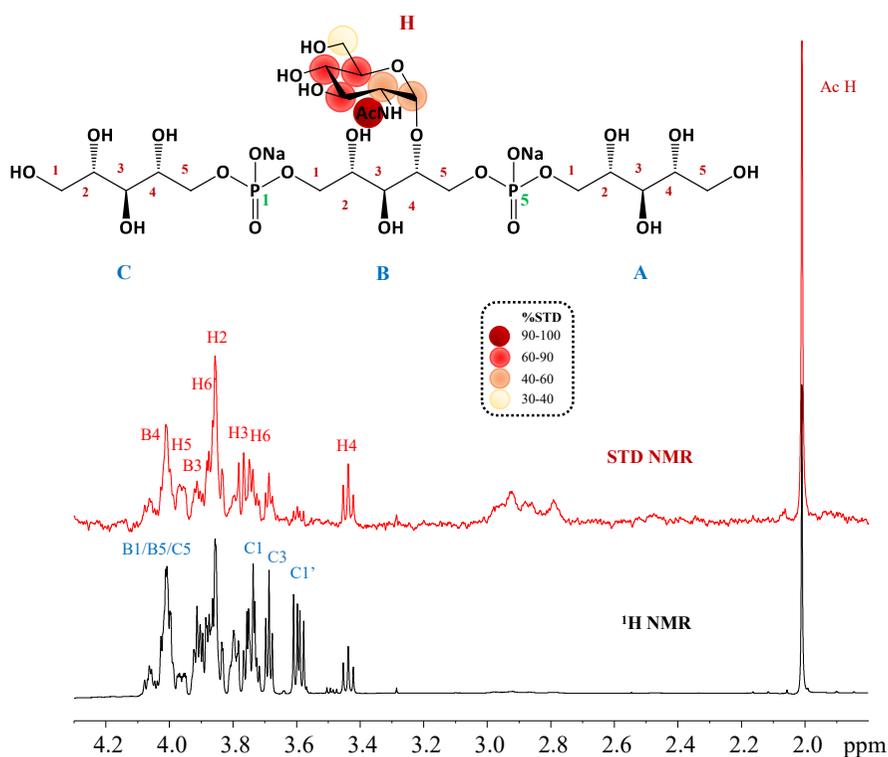


Figure 7.1. STD NMR analysis of IgG mAb 4461 and (α -1,4)-WTA-trimer. STD-NMR spectrum (red) of IgG mAb 4461 –WTA-trimer and the unsaturated reference spectrum (black)

and relative epitope mapping (only protons exhibiting %STD above 30% were indicated in the epitope mapping). Low STD signals were indicated in blue.

Our collaborator Dr M.J. van Raaij and Dr P. Soriano Maldonado (Spanish National Research Council) solved the crystal structure of IgG mAb 4461 in complex with the (α -1,4)-WTA-trimer at 1.4 Å resolution. Then, the complex was subjected to molecular dynamic simulation (figure 7.2). IgG mAb 4461 recognized the α -GlcNAc modified WTA in a cavity formed between complementarity-determining regions (CDRs) of the heavy and light chains of the antibody, with most of the interactions mediated almost entirely through the sugar ring.

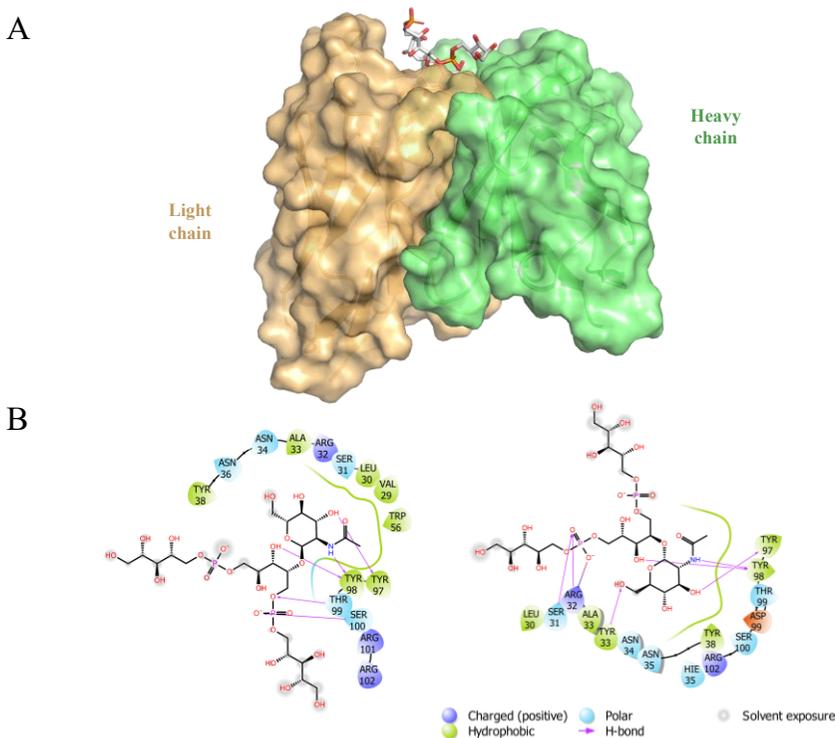


Figure 7.2. MD analysis of IgG mAb 4461 and (α -1,4)-WTA-trimer. A) 3D view of the complex obtained from MD simulation. B) Diagrams of 2D interactions corresponding to the main poses deriving from MD: solid arrows represent hydrogen bonds with functional groups

of the amino acids of the antibody; the other residues in the binding pocket participate in polar and hydrophobic interactions.

Polar, hydrophobic and hydrogen bonds were indeed observed at IgG mAb 4461 – WTA-trimer interface, especially surrounding the pyranose ring. Tyrosine residues were found in the antibody binding site to establish important H-bonds with GlcNAc sugar; in particular, the oxygen of hydroxyl group of Tyr97 and Tyr98 of the heavy chain respectively interacted with OH at position 3 and the amide of GlcNAc, both giving high STD signals (figure 7.1). Regarding the phosphate groups, one of the oxygens of P5 (see nomenclature in figure 7.1) not involved in the phosphodiester linkage made an H-bond with the hydrogen of Ser100. However, another contact between the same phosphate group and the Thr99 was monitored along the simulation, but to a lesser extent. The interaction between P1 with Ser31 and Arg32 were also found in the MD, although less recurring along the simulation. Tyr98 was also detected in establishing a H-bond with the proton at position 3 of the Rbo-B chain, according to the NMR data. Despite the flexibility of the Rbo A and C arms, interactions of the C unit and the antibody could be slightly detected, while the A unit was solvent exposed (figure 7.2 B).

7.2.2 IgG mAb 4461 binding to α - WTA-hexamer

The interaction of 4461 with a longer WTA fragment, *i.e.* a RboP hexamer containing two α -GlcNAc units (figure 7.3) was also investigated. From the analysis of the STD NMR spectra, it was revealed that the portions of the RboP backbone decorated by the GlcNAc moiety (namely C and F units) were differently recognized by the antibody since they displayed different STD enhancements. As depicted in the epitope mapping of WTA-hexamer (figure 7.3 A), the “internal” GlcNAc contributed more to the binding than the “terminal” sugar. The evidence of this result was clear, for

example, when comparing the two STD signals belonging to the acetyl groups of the two GlcNAc residues: the acetyl group of the “terminal” sugar showed only the half of %STD with respect to the corresponding signal of the “internal” GlcNAc, instead set to 100%. Accordingly, the proton H1 of the “internal” GlcNAc revealed a major STD contribution with respect to the anomeric proton of the “terminal” sugar ring. A 3D view of ligand accommodation into the IgG mAb 4461 binding pocket was achieved by manually docking the WTA-hexamer into the monoclonal antibody (figure 7.3 B). The hypothetical 3D structure showed the propensity to an end-on insertion binding mode with the “internal” sugar moiety deeply located in the antigen-binding site, establishing the main interactions with IgG mAb 4461, while the RboP units tended to protrude from the binding site. Overall, the results confirmed the importance of the GlcNAc sugar modification on the WTA RboP backbone for the recognition process suggesting a preference of the monoclonal antibody for the “internal” sugar, embedded in the binding site together with the protons of the neighboring Rbo chains, while the “terminal” sugar as well as Rbo units E and F were further away from the binding pocket.

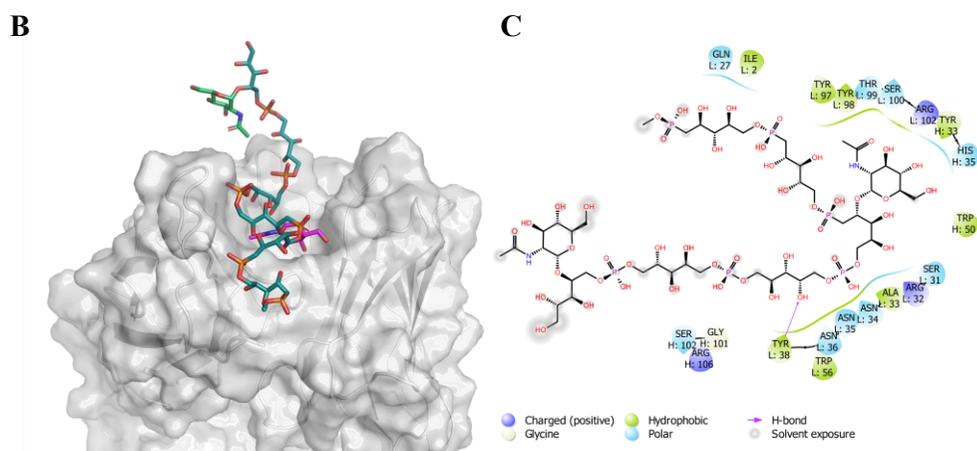
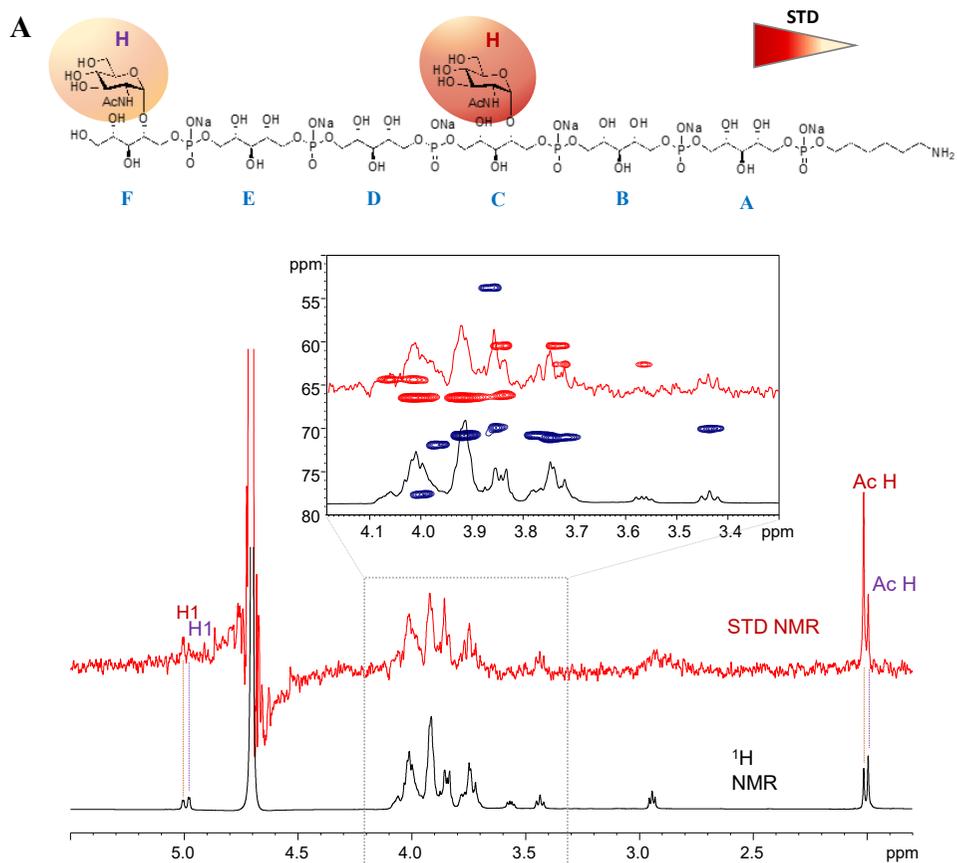


Figure 7.3. STD NMR and 3D view of IgG 4461 and WTA-hexamer. A) STD NMR spectrum and epitope map of WTA-hexamer interacting with IgG mAb 4461. B) 3D view of IgG mAb 4461 – WTA-hexamer: the “internal” sugar into the antibody binding site was colored in pink, the “external” GlcNAc was colored in green. C) 2D plot of interactions resulting from the manual docking: solid arrows represent hydrogen bonds with functional groups of the amino acids of the antibody; the other residues in the binding pocket participate in polar and hydrophobic interactions.

7.3 Molecular investigation of IgG mAb 4497 with β -GlcNAc modified wall teichoic acids

IgG mAb 4497 in complex with WTA modified with β -GlcNAc sugar at position 4 and at position 3 of the inner RboP subunit was explored by STD NMR in combination with computational studies.

7.3.1 IgG mAb 4497 binding to β -(1,4)-WTA-trimer

The binding profile of the WTA-trimer containing a β -linked GlcNAc residue at C-4 of the internal RboP subunit was examined with IgG mAb 4497 and the epitope mapping, showing the STD effects, were achieved accordingly (figure 7.4). From the STD NMR responses, the sugar ring resulted the main residue contributing to the interaction with the antibody. The acetyl group was set to 100%, (figure 7.4) and %STD effects around 50% were observed for protons H2, H3, H4 and H5 of GlcNAc. Lower STD enhancements belonged to the H6 protons of the sugar as well as to the protons of the RboP B unit. Finally, slight STD signals were observed for the protons of the C unit.

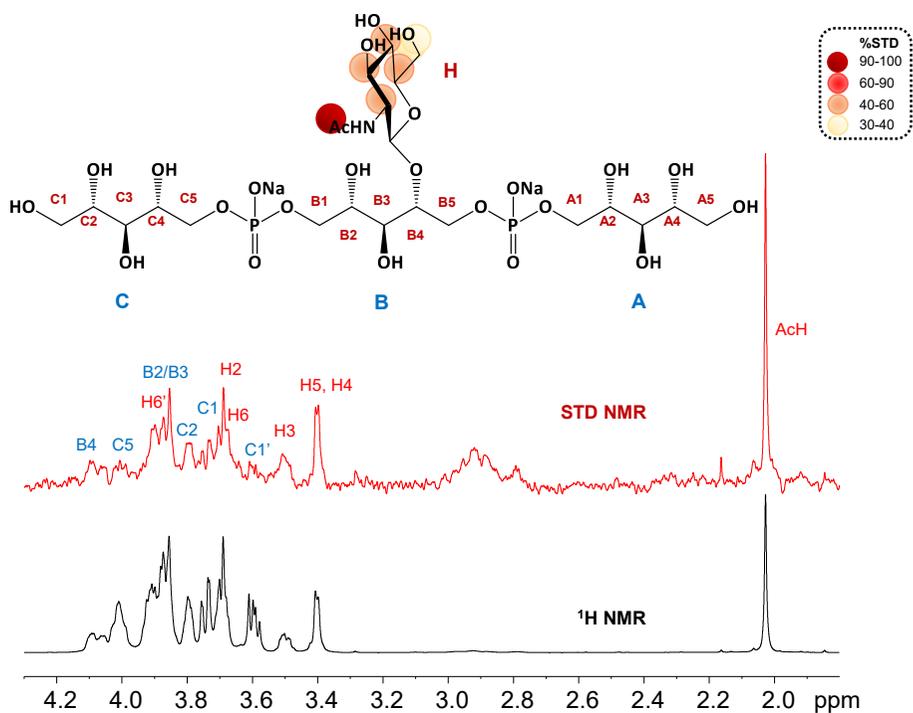


Figure 7.4. STD NMR of IgG mAb 4497 and β -(1,4)-WTA-trimer and relative ligand epitope map (only protons exhibiting %STD above 30% were indicated in the epitope mapping).

The crystallographic structure of IgG mAb 4497 and β -(1,4)-WTA-trimer was subjected to MD simulation to monitor the behavior of the complex in solution and determine the molecular interactions along the dynamic. A representative pose of IgG mAb 4497 and β -(1,4)-WTA-trimer belonging to the most populated cluster obtained from 100 ns MD simulations was depicted in figure 7.5.

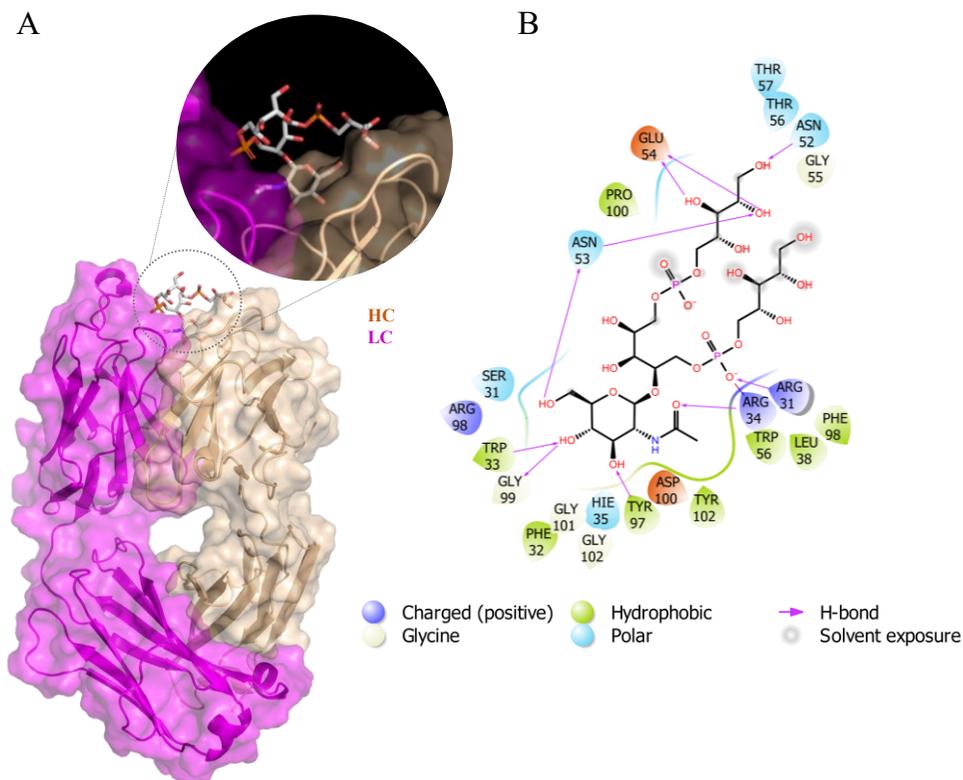


Figure 7.5. Best pose of the complex between IgG 4461 and β -(1,4)-WTA-trimer obtained from MD simulation. A) 3D view and B) 2D plot of the interactions between IgG mAb 4497 and β -(1,4)-WTA-trimer. Solid arrows represent hydrogen bonds with functional groups of the amino acids of the antibody; the other residues in the binding pocket participate in polar and hydrophobic interactions.

The GlcNAc pyranose was recognized by the heavy chain residues of the antibody, stacking with the Trp33 of CDR H1, which side chain indole NH interacted with OH at position 4 of the sugar. Moreover, a very stable hydrogen bond was monitored between Gly99 and OH at position 4 of GlcNAc, maintained for the 97% of the MD simulation time, contact already found in the structure of 5-phosphate β -WTA analog bound to 4497 CDRs (6DWA).²¹⁴ It is worth to know that in the already published structure another glycine (Gly101) belonging to the heavy chain of the antibody

established a contact with O3 of GlcNAc; however, our MD results showed that OH at position 3 of the sugar made an H-bond with the hydrogen of the hydroxyl group of Tyr97 of the CDR light chain L3, that remained stable for the 90% of the simulation. Regarding the pyranose ring, fewer stable interactions were found between the oxygen of Asp100 with O3 of GlcNAc, as well as the hydroxyl group at position 6 and the oxygen of Asn53. Interestingly, simultaneous and stable contacts between PO5 with Arg31 and Arg34 belonging to the light chain of the antibody were detected, in particular both guanidinium of the arginine residues made interactions with the oxygen not involved in the phosphodiester bond (figure 7.5 B). Furthermore, contacts between Glu54 and protons of the Rbo-C2 and Rbo-C3, and, to a lesser extent, N52 and N53 with Rbo-C1 and Rbo-C2 were also found, indicating that the arm C of the ligand participated into the binding, while the arm A was solvent exposed.

7.3.2 IgG mAb 4497 binding to β -(1,3)-WTA-trimer

The STD NMR analysis on β -(1,3)-WTA-trimer, the ligand modified with a GlcNAc residue at position 3 of the inner residue of the RboP, in complex with IgG mAb 4497 was performed (figure 7.6). The main STD contribution came from the GlcNAc unit. In particular, the highest STD signal belonged to the acetyl group of the sugar, thus set to 100%; protons H2, H3, H4 and H5 of GlcNAc received a strong magnetization from the antibody, with %STD ranging between 60-90%. Lower STD effects were found for the H6 protons. These data suggested a strong involvement of the sugar residue into the antibody. Some protons belonging to the RboP chains, especially those of the unit modified by the sugar (B unit), were also recognized by IgG mAb 4497. Indeed, the proton at position 3 resulted to be involved in the binding, as also H2, H3 and H5 of the C residue, although they exhibited lower STD effects.

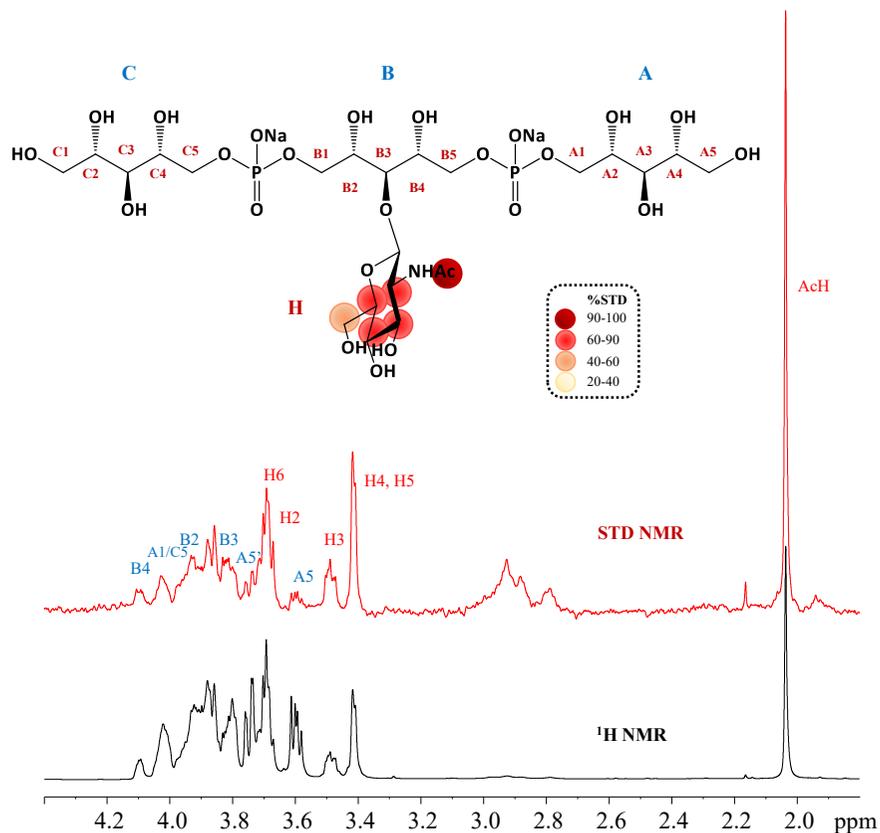


Figure 7.6. STD NMR of IgG mAb 4497 and β -(1,3)-WTA-trimer and relative ligand epitope map (only protons exhibiting %STD above 30% were indicated in the epitope mapping).

The crystallographic structure of IgG mAb 4497 and β -(1,3)-WTA-trimer was subjected to 100 ns MD simulation to monitor the behavior of the complex in solution and determine the molecular interactions along the dynamic. A representative pose of IgG mAb 4497 and β -(1,3)-WTA-trimer belonging to the most populated cluster obtained from MD was depicted in figure 7.7. Importantly, Trp33 aromatic ring belonging to the heavy chain CDR H1 of the antibody stacked against the face of the GlcNAc pyranose, stabilizing the sugar in the cavity formed between heavy CDRs

and light CDR loops of IgG 4497. Several amino acids participated to the recognition of the sugar moiety that was involved in polar and hydrophobic interactions with the antibody. Further important hydrogen bonds, especially with hydroxyl groups at position 3 and 4 of GlcNAc, were formed with CDR heavy chain. Among these, OH at position 4 was involved in the binding with Gly99 as well as OH at position 3 of GlcNAc with Asp100 and Tyr97, maintained for almost the entire MD simulation time. Further H-bond interactions were detected between OH at position 4 of GlcNAc and HN of Trp33, and OH at position 6 with the amide of Asn53 belonging to the CDR heavy chain, although repeated along the simulation to a lesser extent.

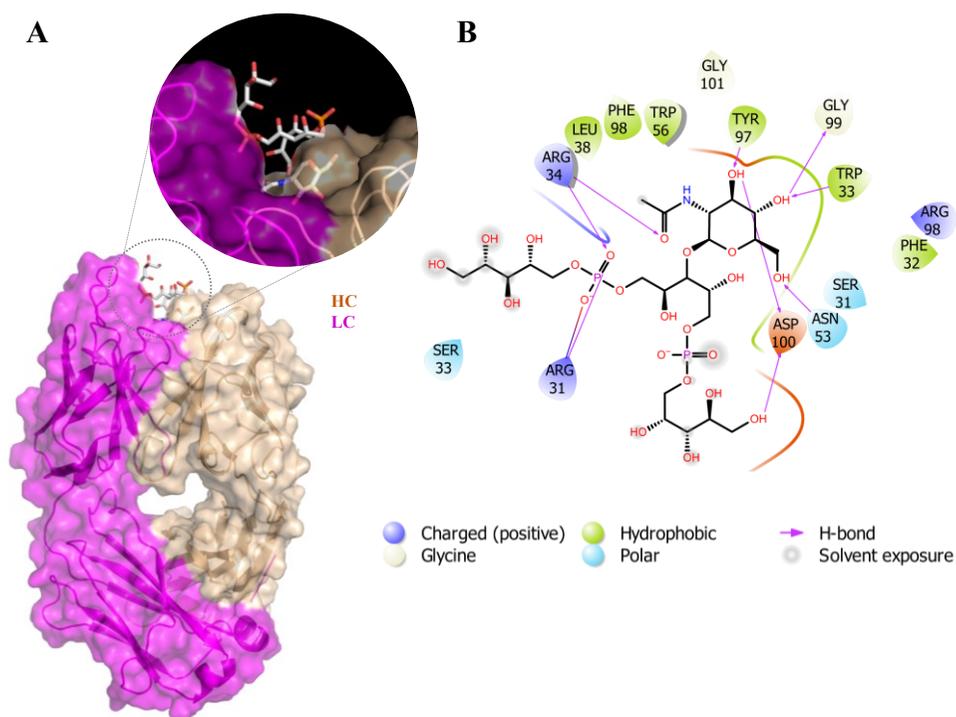


Figure 7.7. Best pose of the complex between IgG 4461 and β -(1,3)-WTA-trimer obtained from MD simulation. A) 3D view and B) 2D plot of the interactions between IgG mAb 4497 and β -(1,3)-WTA-trimer. Solid arrows represent hydrogen bonds with functional groups of the

amino acids of the antibody; the other residues in the binding pocket participate in polar and hydrophobic interactions.

The amino acids contacts with the phosphate groups of β -(1,3)-WTA-trimer were also monitored by MD. From the calculation of the number of contacts monitored along the simulation, Arg31 and Arg34 interacted with one of the oxygens not involved in the phosphodiester linkages of one phosphate group (figure 7.7). Although the A and C units resulted flexible along the trajectory, a partial involvement of one RboP chain could be observed (figure 7.7 B).

7.4 Discussion

The recognition of three minimal binding epitopes constituting a single GlcNAc of different stereochemistry (α or β) at different positions (3 and 4) of a ribitol backbone, representing possible modifications of WTAs composing *S. aureus* cell wall, have been investigated with specific monoclonal antibodies. Due to the antibiotic-resistance developed, unraveling antibody-WTA interactions at the atomic level could represent the starting point to generate novel therapeutics against *S. aureus*. By means of the combination of spectroscopic and computational approaches, we confirmed the importance of carbohydrate role in WTA structures, since the GlcNAc was the main residue recognized by the antibody, as previously reported.²¹⁴ Additionally, we found the involvement of two flanking phosphates and ribitol moieties in the binding, showing that the flexibility of the RboP backbone is crucial to accommodate both the β 1,3- and β 1,4-regioisomers. Furthermore, for the first time, we examined the molecular binding of IgG mAb 4461 with α 1,4-GlcNAc WTA-trimer and -hexamer, showing the preference of the “internal” α -GlcNAc-residue into the Ab-surface, as demonstrated by TA-microarray (not shown) and NMR studies.

Overall, the WTA-ligands into their specific monoclonal antibody structures displayed an “end-on” insertion binding mode³⁰ with the antigen-binding site amino acids forming numerous contacts mainly with the GlcNAc residue and also interactions with the RboP backbone.

General conclusions

Glycans are important mediator of biological functions, that include structural and modulatory properties, molecular mimicry and specific recognition by, most commonly, glycan-binding proteins that in turn underly both physiological, health and pathological events.

In this thesis, new insights into understanding the protein-ligand molecular binding processes within different biological events have been provided, to set the basis for the design of novel therapeutic and diagnostic treatments. In this framework, a fundamental prerequisite is the detailed description of the molecular interaction process, here achieved by an ensemble of advanced NMR techniques, biophysical methods, and computational studies. Indeed, determining the ligand epitope and its bioactive conformation in the receptor binding site as well as knowing the nature of the interactions at protein-ligand interface are crucial information to rationally design and optimize high affinity glycomimetics against human diseases.

Chapter VIII:
Material and Methods

VIII. Material and Methods

8.1 Production of recombinant proteins

8.1.1 Siglec-2 (related to Chapter III)

Human and murine CD22 proteins were kindly provided by Prof. Crocker (University of Dundee, UK). The N-terminal Ig-like domains of CD22 fused to the Fc region of mouse IgG2b proteins were expressed in Chinese Hamster Ovary cell as described in the published protocol.⁷⁹

8.1.2 Siglec-7 (related to Chapter IV)

8.1.2.1 Molecular cloning

The vectors were linearized and restricted by AgeI-HF and KpnI endonucleases. The reaction was prepared by using 3 µg of plasmid, 2 µl of NEB1 buffer, 1 µl of each restriction endonucleases (10 U of each enzyme), and ddH₂O to the final volume of 20 µl. The reaction was incubated at 37°C for 2h and the restricted plasmids were analyzed and extracted from the agarose gel electrophoresis (1%) and purified using the commercial kit PCR clean-up.

A synthetic vector pcDNA3.1(+) encoding the gene of interest was used as a template. The fragments encoding for the constructs of interest were amplified by polymerase chain reaction (PCR). The PCR reaction was prepared by adding 100 ng of template DNA, 2 µl of forward primer, 2 µl of reverse primer to have a final concentration of 5 µM, 5 µl of 5 × Q5 polymerase enhancer, 0.5 µl of 10 mM deoxyribonucleotides mix (dNTPs), 5 µl of 5 × Q5 buffer, and 0.2 µl of Q5 polymerase. ddH₂O was added to the final volume of 25 µl. The PCR tubes were placed in a thermal cycler programmed with the following steps:

A. 3 min – 95°C

B. 30 cycles: 30 s – 95°C; 30 s – 58°C; 1 min – 72°C

C. 5 min – 72°C

After finishing the program, the temperature in the thermal cycler was decreased to 4°C.

The PCR products were analyzed by agarose gel (1%) and purified by PCR clean-up. Then, the DNA fragments were cloned into their corresponding linearized expression plasmids employing a ligation protocol with T4 DNA ligase. The ligation mixture was prepared by mixing 100 ng of restricted vector, 50 ng of restricted DNA, 2 µl of 10 x T4 DNA ligase buffer and 1 µl of T4 Dna ligase enzyme. ddH₂O was added to the final volume of 20 µl. The PCR tubes were placed in a thermal cycler programmed with the following steps:

A. 2 h – 16°C

B. 25 min – 65°C

After finishing the program, the temperature in the thermal cycler was decreased to 4°C.

Under sterile conditions, competent *E. coli* TOP10 cells were transformed with the plasmids containing the construct of interest. The mixture was incubated on ice for 30 minutes, followed by heat shock at 42°C for 45 sec and 2 min on ice. Then, 1 ml of SOB medium preheated at 37 °C was added to the transformation and the samples were incubated at 37°C for 60 min. The transformed cells were then centrifuged for 3 min at 2000 × g. Part of the supernatant was removed, while the cell pellet was resuspended in the remaining supernatant and plated on a SOB agar plate containing 100 µg/ml ampicillin. The plate was incubated O.N. at 37°C.

Colonies were randomly selected from each plate and analyzed by colony PCR. The colony PCR was prepared in PCR tubes with 10 µl of Master Mix, 2 µl of reverse primer and 2 µl of forward primer. ddH₂O was added to the final volume of 20 µl. The PCR tubes were placed in a thermal cycler programmed with the following steps:

- A. 3 min – 95°C
- B. 20 cycles: 30 s – 95°C; 30 s – 55°C; 1,30 min – 72°C
- C. 5 min – 72°C

After finishing the program, the temperature in the thermal cycler was decreased to 4°C. The PCR products were analyzed by agarose gel electrophoresis (1%). The colonies corresponding to the PCR products were used to inoculate a small volume of SOB media for the minipreparation of plasmid DNA. Then, the high-copy plasmids from *E. coli* cells were isolated using the commercial NucleoSpin Plasmid kit.

Once the DNA sequences of pHLsec_Siglec-7_Fc, pTW5sec_FED, and pTW5sec_CRD were confirmed by DNA sequencing, competent *E. coli* DH5alpha cells were transformed with the plasmids for the maxipreparation of plasmid DNA. The mixture was incubated on ice for 30 minutes and then plated on a SOB agar plate containing 100 µg/ml ampicillin. The plate was incubated O.N. at 37°C.

The 20 µl tip used for picking the selected colony for colony PCR was used to inoculate 5 ml of sterile LB medium containing 100 µg/ml of ampicillin in 50 ml tube for the small-scale isolation of plasmid DNA. Bacteria grew O.N., at 220 rpm at 37°C. After culture growth, the cells were harvested (45 minutes, at 4°C and 4700 rpm), the supernatant discarded, and the pellet purified using the NucleoBond Xtra Midi/Maxi commercial kit.

8.1.2.2 Transfection with HEK293S cells

The recombinant proteins were expressed in suspension-adapted HEK293S GnTI- cells following a high-density transient transfection protocol. HEK293S GnTI- cell line was used to express Siglec-7 with homogenous Man5GlcNAc2 glycosylation. For each transfection, a total of 800 µg of DNA (1 µg per 1 million cells) was prepared in PBS-TK buffer. The transfection mix was comprised of 88% expression plasmid, 10% pTW5sec_P27, and 2% pTW5sec_aFGF.

HEK293S GnTI⁻ cells were previously cultivated, and their cell number and viability were properly assessed. The cells were centrifuged for 5 min at $95 \times g$ and the resulting supernatant was discarded. The cell pellet was resuspended in 34 ml of EXCELL293 medium, pre-tempered at 37°C , and transferred into a glass cell culture bottle. The DNA transfection mix was filtered through a $0.22 \mu\text{m}$ filter into the glass bottle containing the resuspended cells.

The DNA samples were filtered and transferred into a glass cell culture bottle containing HEK293S cells, together with linear polyethyleneimine (DNA:IPEI weight ratio of 1:4). After incubating the cells for 1.5 hours, 1.6 ml of 0.5 M valproic acid and antibiotics were added to the volume completed to 400 ml with EXCELL293 cell culture medium. After six days, the cells were harvested (at $3900 \times g$ for 30 minutes), and the supernatant was filtered and diluted with PBS buffer in a 1:1 ratio.

8.1.2.3 Purification of recombinant Siglec-7

The proteins were purified by immobilized metal affinity chromatography (IMAC) followed by size-exclusion chromatography (SEC). The fractions eluted from the SEC were quantified using the NanoDrop spectrophotometer and assessed by SDS-PAGE. Regarding the IMAC, the column was equilibrated with PBS buffer, the supernatant was loaded into the column and the flow through was collected. After the medium loading, the column was washed with PBS buffer to elute unbound proteins, while the protein of interest was eluted by addition of elution buffer (50 mM Na_2HPO_4 , 300 mM NaCl, 10 mM NaN_3 , 250 mM imidazole, pH 7.0).

Regarding the SEC, a Superdex 200 10/300 GL column was used. The protein was eluted in hepes buffer, pH 7.5.

8.1.2 Recombinant expression of Siglec-like adhesins SLBR-B, -H and -N in *E. coli* (related to Chapters V and VI)

The Siglec and Unique domains of Siglec-like adhesins were expressed as GST fusion proteins. The expression plasmid pGEX was kindly provided by Prof. Barbara A. Bensing (San Francisco, USA). Competent *E. coli* BL21(DE3) cells were transformed with 60 µg/µl plasmids. The bacterial culture was firstly grown in Luria Broth Medium using 50 µg/ml ampicillin at 37°C and then induced with 1 mM isopropyl-1-thio-d-galactopyranoside (IPTG) overnight at 18°C, when A600 reached 0.6 nm. Cells were harvested by centrifugation (20 min, 7500 rpm) and resuspended in phosphate saline buffer (PBS) pH 7.4, containing 1M DTT and AEBSF inhibitor of proteases. Cells were disrupted by sonication and harvested by centrifugation to obtain the soluble protein. The proteins were purified using a Glutathione Sepharose 4B column (GE Healthcare) and eluted with 10 mM GSH and 10 mM DTT in 50 mM Tris-HCl, pH 8.0.

The eluted GST-SLBR-H and GST-SLBR-B proteins were dialyzed against PBS pH 7.4 for ligand-based NMR experiments.

Regarding GST-SLBR-N, after removing the reducing agents using a High Trap Desalting 26/10, GST tag was cleaved after an incubation O.N. with 1 µg/µl Factor Xa protease. SLBR-N was purified using a HiLoad 16/60 Superdex 75 pg in PBS pH 7.4.

Then, SLBR-N was expressed in M9 minimal medium as both ¹⁵N and ¹³C¹⁵N labelled protein, by adding 2.8 g of ¹⁵NH₄Cl and 6 g of ¹³C-enriched glucose (in 2L of media). For the expression of the triple labeled protein the Silantes medium (²H¹³C¹⁵N enriched) was used. The protocol of expression and purification of the labeled protein was the same used for the unlabeled protein.

8.2 NMR analysis

8.2.1 SLBR-N protein assignment and titration (related to Chapter VI)

Solution NMR spectra were recorded at 298 K on Bruker's Avance™ NEO 1.2 GHz spectrometer, equipped with a triple resonance TXO cryo-probe. Protein samples were in water buffered solution (20 mM sodium phosphate, pH 7.5, 0.02% NaN₃). The ²H¹³C¹⁵N SLBR-N was prepared 340 μM in the shigemi tube, in 100 μl of PBS buffer. 2D ¹H¹⁵N TROSY-HSQC 3D TROSY NMR experiments were acquired for the protein assignment: HNCO, HNcaCO, HNCA, HNcoCA, CBCAcoNH, HNCACB. For the study on the protein-ligand interaction, a 2D ¹H¹⁵N TROSY-HSQC experiment was recorded on 200 μM of ¹⁵N SLBR-N alone and then titrated by adding 3'SL_n at different concentrations (12.5, 25, 100, 200, 400, 800 and 1600 μM). Data acquisition and processing were performed with TOPSPIN 4.1.1 software and CARA program for the protein assignment.¹³⁸

8.2.2 NMR ligand-based techniques

The NMR experiments were recorded on a Bruker AVANCE NEO 600-MHz equipped with a cryo-probe and data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in phosphate saline deuterated buffer (10 mM Na₂HPO₄, 2.7 mM KCl, 137 mM NaCl, 10 mM NaN₃), pH 7.4 at 298 K. [D4](trimethylsilyl)propionic acid, sodium salt (TSP, 10 μM) was used as internal reference.

8.2.2.1 STD NMR

Protein/ligand molar ratios varied from 1:20 to 1:100. STD NMR experiments were acquired with 32 k data points and zero-filled up to 64 k data points prior to processing at saturation time of 2s. The protein resonances were selectively irradiated using 40 Gauss pulses with a length of 50 ms, setting the off-resonance pulse frequency at 40

ppm and the on-resonance pulses in aromatic and aliphatic region. An excitation sculpting with gradient pulses (esgp) was applied for the suppression of water signals. When broad signals of the receptor were detected in the on-resonance spectra, 20 ms spin lock pulse was applied. The %STD displayed in the ligands' epitope maps were obtained by the ratio of the STD signals ($I_0 - I_{\text{sat}}$) and each relative peak intensity of the unsaturated reference spectrum (off-resonance, I_0). The highest STD signal was set to 100% and all the other STD were normalized to this value.

Regarding the interactions with h- and m-CD22 (Chapter III), the construction of STD build up curves was performed by fitting the saturation time data (ranging from 0.8 to 5 s) to a mono-exponential equation of the form: $\text{STD} = \text{STD}_{\text{max}}[1 - \exp(-k_{\text{sat}}t)]$, where STD is the STD signal intensity of a given proton at a saturation time t , STD_{max} is the asymptotic maximum of the curve, and k_{sat} is the observed saturation rate constant measuring the speed of STD build-up. The STD_{fit} value was derived by the slope of the STD build-up curve at a saturation time of 0. After calculating both STD_{fit} and K_{sat} values, all the signals were normalized to the largest STD_{fit} , giving $\text{STD}_{\text{epitopes fit}}$.

8.2.2.2 Tr-NOESY/ROESY analysis

Protein/ligand molar ratios varied from 1:10 to 1:20. Homonuclear 2D ^1H - ^1H NOESY/ROESY experiments were carried out by using data sets of 2048x512 points and mixing times of 600-700 ms for the free states and of 300 ms for the bound states. Proton – proton cross relaxation rates were measured integrating the NOE/ROE cross peaks of interest. The raw data of each cross peak were normalized against the corresponding diagonal peak and the ^1H - ^1H distances were calculated using the following equation: $r_{ij} = r_{\text{ref}} \sqrt[6]{\frac{\sigma_{\text{ref}}}{\sigma_{ij}}}$ (Chapter II, eq. [2.4]).

Regarding the construction of NOE build up curves (Chapter III), mixing times were chosen ranging 50 and 800 ms. The raw data of each cross peak were normalized using the decay of the corresponding diagonal signal as a reference that was fitted to

an exponential decay function and extrapolated to an intensity of 100% at zero mixing time. After the normalization, the intensities were plotted against the mixing times in the build-up curves in the following exponential equation:

$$f(x) = A * (1 - \exp(-b * x)) * (\exp(-c * t)) \quad [8.1]$$

with A, b and c being adjustable parameters. The initial slope at 0 mixing time (σ) was used to calculate the experimental distances (r_{ij}) by employing the isolated spin pair approximation and using the distance H1-H5 of the *N*-acetylglucosamine residue as a reference. The experimental error in the calculation of proton-proton distances was estimated to be $\pm 10\%$.

8.2.2.3 WaterLOGSY

For the molecular binding between 3'SL_n and SLBR-H (Chapter V) waterLOGSY experiments were performed on the free and bound states using a conventional 1D NOE-ePHOGSY pulse sequence provided in the Bruker library (ephogsygpn0.2)²⁰¹ with mixing time of 2 s. The bulk water was selectively irradiated by 1D NOESY and the solvent suppression was modified into a double pulsed field gradient (DPFG) perfect-echo.²⁰²

8.2.2.4 CPMG

The cpmgz pulse program based on spin-echo pulse sequence was used for the acquisition of CPMG experiments (Chapters IV, V, VII). The signal relaxation was measured for 100 ms and the experiments were acquired with 256 scans. The recycle delay was set to 2 s and fixed echo time was chosen at 3 ms. For the calculation of T₂ relaxation time, CPMG experiments at different mixing times were acquired. The spectra were loaded in the Dynamics Center 2.7.1 that fitted the data by the following equation:

$$f(t) = I_0 * \exp (-t/T) \quad [8.2]$$

A confident level of 95% was achieved.

8.2.2.5 DOSY

Pulsed-Field Gradient (PFG)-NMR experiments were carried out for investigating the interactions between Siglec-7-CRD and disialyl pentaose type 1 (Chapter IV). The stimulated echo sequence with bipolar gradient pulses and one spoil gradient (stebpgp1s1d from Bruker library) with a longitudinal eddy current delay was used. Once optimized the experimental parameters including the echo delay ($\Delta = 150$ ms) and the gradient duration ($\delta = 1.2$ ms), the sequence was run as 2D NMR experiment with a linear gradient G incremented, in 16 steps, from 2% to 95% of the maximum gradient amplitude (5.35 G cm^{-1}). A Fourier transformation and baseline correction of 1D ^1H spectra (F2 dimension) were applied.

8.3 Biophysical techniques

8.3.1 Fluorescence analysis (related to Chapters III-V)

The quenching fluorescence titration curves were achieved by using a Fluoromax-4 spectrofluorometer (Horiba, Edison, NJ, USA). The measurements were acquired at 10°C , upon excitation at 285 nm and recording the emission spectra in the range of 295–600 nm. The slit widths were chosen at 5 nm for the excitation and 10 nm for the emission wavelength. A quartz cuvette with a path length of 1 cm was used. A fixed concentration of SLBR-N was chosen at $0.25 \mu\text{M}$ in 1.2 mL PBS buffer (pH 7.4) and titrated by adding small amounts of ligand until the protein saturation. The binding curve was obtained by plotting $\Delta F/F_0$ values versus ligand concentration. Data were analyzed by using Origin, using the following equation:

$$\frac{\Delta I_f}{I_0} = \frac{\Delta I_{max}}{I_0} X_{FY} \quad [8.3]$$

where ΔI_f was the fluorescence intensity change upon addition of the ligand and ΔI_{max} the maximal fluorescence intensity change, F was fluorophore, Y the interference specie and $X_{FY} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, with $a=[F]tKb$, $b=1+[Y]tKb$, $c=[Y]tKb$ (Kb was the association constant).

For the evaluation of the binding interactions, in particular of the dissociation constant K_D (Chapter V), data were analyzed by non-linear regression equation considering one site- specific binding model:

$$Y = \frac{B_{max} * X}{K_D + X} \quad [8.4]$$

where Y was the fluorescence intensity change at the maximum wavelength, X was the ligand concentration, and B_{max} was the maximum specific binding.

8.3.2 Alpha screen assay (related to Chapter III)

Preliminary experiments were performed to optimize the experimental conditions of the Alpha Screen. The concentrations of beads were kept constant (20 $\mu\text{g/mL}$ final concentration of each bead), by varying only the concentration of the tagged protein and the tracer in a final assay volume of 40 μL . Excitation of donor beads was at 680 nM. Samples were measured at 520-620 nm in EnspireTM Alpha (Perkin-Elmer). After chosen the best concentrations of the tagged protein and tracer, corresponding to 0.3 nM and 3 nM, respectively, the displacement assay was performed by using as buffer assay PBS containing 0.5% BSA. Fc-tagged human CD22 protein (10 $\mu\text{L/well}$) was incubated with 6'SLN-polyacrylamide-biotin (10 $\mu\text{L/well}$, GLycoNz) and a final concentration range (1, 3, 4, 6, 10, 15, 30, 100, 300 μM) of glycans (10 $\mu\text{L/well}$) for 1h at 25 °C. Then, anti-mouse IgG Acceptor beads (10 $\mu\text{L/well}$, final concentration 20 $\mu\text{g/mL}$) were added to the plate and incubated for 60 min. Streptavidin donor beads

(10 μ L/well, final concentration 20 μ g/mL) in buffer assay were also added to the plate. The alpha counts were normalized according to the following equation: $Y = (X - \text{background}) * 100 / (Z - \text{background})$, X = [alpha counts in the presence of each ligand], Z = [alpha counts in the absence of ligands], background = [alpha counts in the absence of ligands and 6'-SLN-PAA-biotin]. The dose-response curves were obtained by fitting the data in GraphPad Prism8, using the Log(inhibitor) vs normalized response – variable slope formula. The experiments were performed in triplicate.

8.3.3 Nano Differential Scanning Fluorimetry (related to Chapter IV)

The instrument was Prometheus NT.48 and the measurements were performed with the support by colleague Celeste de Sousa Santos Abreu (Charles University, Prague). Prior to the measurements, the samples were centrifuged for 10 min at 10,000 \times g. A concentration of 1 mg/ml of each protein was exposed to a temperature gradient that led to their complete unfolding from 20°C to 95°C with 2°C increments per minute. Data were analyzed using the software programs PT.ThermControl and PR.Stability analysis.

8.3.4 Dynamic light scattering (related to Chapter IV)

Dynamic light scattering (DLS) was measured using Zetasizer Ultra, with technical support by colleague Celeste de Sousa Santos Abreu (Charles University, Prague). The method employed was one measurement of protein size followed by one measurement of multiangle dynamic light scattering at 25°C. Before the analysis, samples were centrifuged for 10 min at 10,000 \times g. 40 μ l of 1 mg/ml protein sample were pipetted into an ultra-low volume ZEN2112 quartz cuvette. Data were processed using the ZS Xplorer Software.

8.3.5 Analytical ultracentrifugation (Chapter IV)

Sedimentation velocity experiments acquired using analytical ultracentrifuge ProteomeLab XL-I were performed by RNDr. Ondřej Vaněk (Charles University, Prague). Proteins alone, ligands alone and their mixture were prepared at different concentrations and molecular ratio in HEPES buffer, pH 7.4, used as reference. Samples' concentrations were calculated using Lambert Beer's Law, with the assumptions of 1.2 cm optical pathlength and 0.8 as the desired absorbance value. The samples were loaded into an ultracentrifuge cell with optically clear windows, in turn put in the An50-Ti rotor into the ultracentrifuge chamber. Analyses were performed at 45,000 rpm at 20°C, with 200 absorbance scans taken in 4 min intervals at 280 nm wavelength. Buffer density, protein partial specific volumes, and particle dimensions were predicted by SEDNTERP software:²⁰³ $\rho(20^\circ\text{C}) = 1.00561$, $\eta(20^\circ\text{C}) = 0.010258$. The data were analyzed by SEDFIT software²⁰⁴ using the continuous sedimentation coefficient distribution $c(s)$ model. Figures were prepared in GUSI 1.4.2.²⁰⁵

8.4 *In silico* analysis

8.4.1 Homology modeling (related to Chapter III)

The homology modeling of m-CD22 (Chapter III) was performed by PhD Rosa Ester Forgione. The sequence encoding for V-set and C2 set domains of m-CD22 (Uniprot: NP_033975.3) was obtained from NCBI (<http://www.ncbi.nlm.nih.gov>) and aligned to h-CD22 template (PDB: 5VKJ) using BLAST²⁰⁶ (Basic Local Alignment Search Tool), showing above 58% of sequence identity. Then, the target template alignment was submitted to the SWISS-MODEL²⁰⁷ server to achieve the 3D structure of m-CD22. The quality of the model was evaluated by using PROCHECK²⁰⁸ web server.

8.4.2 Molecular mechanics

MM studies were carried out using Maestro software. The adiabatic maps were built for each disaccharide connected by a glycosidic linkage, defined by the torsion angles Φ (H1-C1-O-CX') and Ψ (C1-O-CX'-HX'). MM3* force field included in MacroModel at a dielectric constant of 80 was used for the calculations. For each disaccharide both Φ and Ψ dihedral angles were varied incrementally using a grid step of 18 degrees. The corresponding flexible maps were drawn as 2D contours plots using the graphical tools of MacroModel tool.

8.4.3 Molecular docking

Docking calculations were performed with AutoDock4 and the analysis of the docking poses was performed by AutoDockTools (ADT). A 3D grid was chosen with a volume necessary to cover the ligand and the key binding amino acids of the target protein. A distance-dependent dielectric constant and the original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock were used. A total of 200 runs using Lamarckian genetic algorithm was used to sample different ligand conformations, by randomly changing all the torsion angles and overall orientation of the molecule, with a population size of 100, and 250000 energy evaluations. ADT reports how many docked conformations were read and allows to visualize the docked conformations grouped according to the clustering performed at the end of the AutoDock calculation. The conformations were ordered by docked energy, from lowest to highest.

8.4.4 Molecular dynamics (Chapters IV-VII)

The proteins structures were firstly pre-processed within Maestro Protein Preparation Wizard in Maestro program. The sugars were built by using carbohydrate builder utility of the glycam website²⁰⁹; the torsion angles were chosen by following the values from the molecular mechanics calculations. Before the MD analysis, all the complexes studied in the thesis were minimized using Sander in Amber tools. Dynamics of 100

ns and 500 ns (for the disialylated core 2 *O*-glycans with SLBR-N, Chapter VI) were performed by using the CUDA²¹⁰, implementation of PMEMD in Amber 18 software²¹¹. The force fields to assign atom types and charges of the proteins and the sugars were AMBER ff14SB and GLYCAM-06j-1, respectively. 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 10 Å (15 Å for the ligands alone). Under periodic boundary conditions, the smooth particle mesh Ewald method was used to compute long-range electrostatic interactions, with a grid spacing set to 1 Å. The system was minimized and slowly heated from 0 to 300 K by applying a restriction to the solute. Then, temperature was kept at 300 K removing the restrains for the system equilibration. Coordinates were ranked to acquire 10000 structures of the progression of the dynamics. Trajectories were visualized by VMD program²¹² and analyzed using the ptraj module within AMBER18. The analysis of the clusters with respect to the ligand RMSD was calculated using K-mean algorithm implemented in ptraj module in order to choose the best poses along the simulation. The analysis of contacts to determine hydrogen bonds was calculated using the CPPTRAJ module in AMBER 18, setting the distance cut-off at 3 Å and the angle cut-off at 135°. No torsional restrictions were applied along the dynamics, except for the sTa-Thr ligand, where a restraint to the peptide dihedral angle (O-CB-CA-N) was applied to keep its value around 60 degrees. The non-standard atom types/residues were parametrized. A mol2 file was generated in antechamber program and a frcmod file was created using parmchk2 for the force field parameters of the ligand. Once parametrized, pdb file of ligand was built and the prmtop and inpcrd files were generated accordingly by using tLEaP module of AMBER 18 package. This approach was used for the parametrization of D-FucNAcN residue in the hexasaccharide of *F. nucleatum* OPS (Chapter IV) and Rbo chain in wall teichoic acids (Chapter VII).

8.4.5 CORCEMA-ST (related to Chapter III)

The PDB of the complexes obtained from docking calculations were analyzed in CORCEMA-ST implemented in Matlab software. In the input file, the following parameters were set dependently on the complex under study: protein and ligand concentrations used in the STD NMR experiments, saturation time of the STD (2 s), dissociation constant (in the range of 10^{-4} – 10^{-6} M), k_{on} (10^{-8} L mol⁻¹ s⁻¹) and the STD values calculated for each proton of the ligand. After running the calculation, the R-NOE factor was provided, meaning the goodness of the 3D model.

APPENDIX

APPENDIX

Molecular binding of mAb 2C7 recognizing a mimotope of *Neisseria gonorrhoeae* LOS

1.1 Introduction

Neisseria gonorrhoeae (*Ng*) is one of the most diffuse human pathogens worldwide. Generally sexually transmitted and occasionally by eye contact (conjunctivitis), *Ng* can cause several diseases, including urethritis in men and cervicitis in women, locally invasive (epididymitis and salpingitis) and disseminated gonococcal infections (DGI), and, in rare cases, can lead to meningitis and endocarditis.²¹³ Due the antibiotic resistance (*e.g.* ceftriaxone) developed by gonococcal strains, different therapies for the prevention and treatment of the disease, such as effective vaccines, are needed.²¹⁴ The Lipooligosaccharide (LOS) (see the structure on Chapter I, §1.2.1) on *Ng* outer membrane represent a potential target for vaccine development.²¹⁵ However, bacteria can undergo phase variation, a mechanism of antigenic escape that causes an alteration of proteins or carbohydrates on bacterial surface thus impairing host immune response;²¹⁶ therefore, vaccine development targeting bacterial envelope components like LPS is challenging. In the last years, different vaccines against *Ng* have been developed, *e.g.* intravaginal vaccine in female mice, able to elicit a memory response resulting in protection from the bacterium, but ineffective in triggering bactericidal activity.

Generally, the antigenic determinant of LOS is the oligosaccharide portion (OS). In gonococcal LOS, the core oligosaccharides are characterized by different lengths (figure 1.1); different regions (epitopes) of LOS core region are recognized by specific monoclonal antibodies (mAbs).²¹⁷ In some cases, cross-reactivity of mAb can occur, since they are structurally similar to human glycosphingolipids (GLSs), this preventing *Ng* immune recognition. On the other hand, antigenic determinants on LOS of *Ng* which do not cross react with human GSL antigens exist, such as the lactose

target linked to Hep II, called 2C7 epitope, widely conserved among gonococcal species (Figure 1.1).^{218,219}

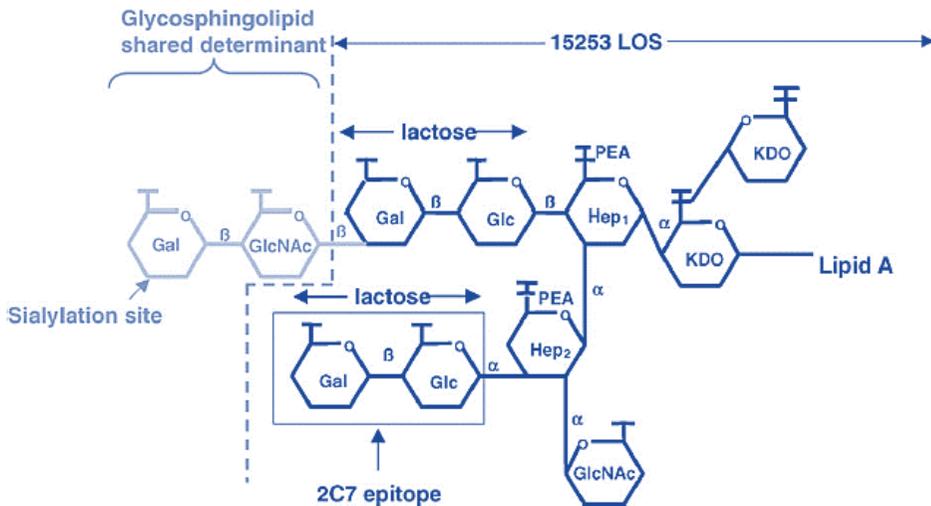


Figure 1.1. Structure of gonococcal lipooligosaccharide (LOS). Gonococcal LOS contains three oligosaccharide (OS) chains attached to lipid A via KDO: two OSs are linked to Hep I and one is linked to Hep II by means of various glycosyltransferases. Sialylation to galactose (Gal) residues can occur via α 2-6 or α 2-3 linkages. PEA is referred to phosphoethanolamine.

It was observed that the absence of Hep II linked lactose (therefore the complete 2C7 epitope) acutely reduces gonococcal infection in the mouse cervico/vaginal colonization model.²²⁰ Therefore, 2C7 epitope may be an important virulence factor that enhances or may be required for survival and productive infection in humans. Although IgtG gene, the glycosyltransferase that allows the attachment of the lactose to the Hep II, can undergo phase variation, almost all gonococcal species show the same 2C7 epitope, resulting immunogenic in natural infection and thus a good candidate as antigenic target.²²¹

Recently, the synthesis of peptides that mimicry the OS structure (mimotopes) has been developed²²⁶; 87% of mice immunized with a mimotope recognized by mAb 2C7

was evaluated, demonstrating the ability to produce anti-LOS antibodies and to trigger bactericidal activity. An optimized 2C7-LOS mimic, called TMCP2, was designed as a peptide circularized by a covalent thioether bond containing a multiple antigen peptide (MAP) to allow stability and avidity toward mAb 2C7 (figure 1.2).

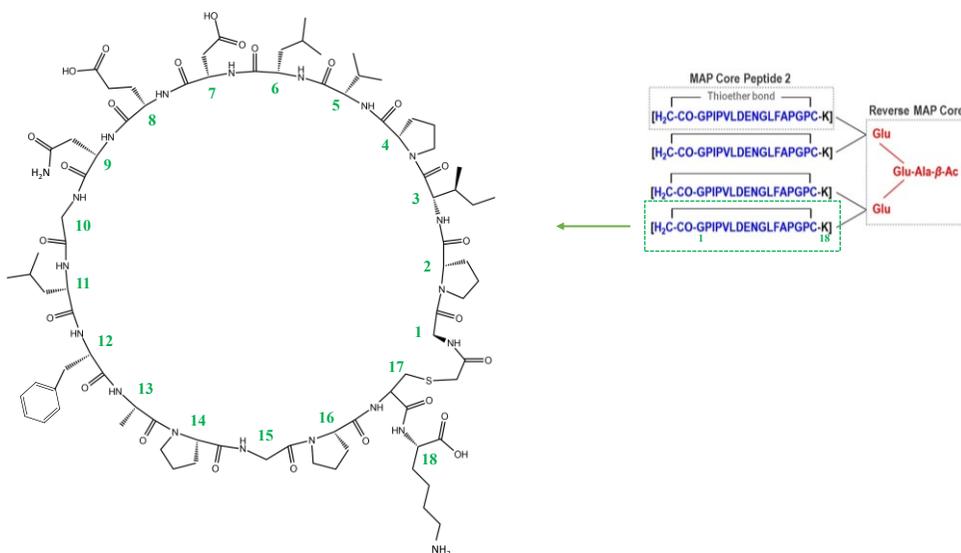


Figure 1.2. Structure of 2C7-LOS peptide mimic TMCP2. The gonococcal vaccine candidate was synthesized as tetrapeptide anchored to a lysine backbone linked with a multiple antigen peptide (MAP) to increase the avidity toward mAb and to simulate the cross-reactivity of LOS to cognate receptors on B cells. The stability of the peptide was given by the presence of covalent thioether bond that allowed the formation of a circular peptide. The single circularized branch of TMCP2 (highlighted by green square) is called CP2.

In collaboration with Prof. Ram Sanjay and Prof. Peter Rice (University of Massachusetts), we characterized the structural requirements for the interactions between 2C7 mAb and the anti-gonococcal vaccine candidate from the molecular viewpoint.

1.2 Binding analysis between mAb 2C7 and TMCP2

Binding experiments were explored between mAb 2C7 and the tetrameric form of CP2, the tetrapeptide TMCP2 (see the structure in figure 1.2).

The presence of signals in the STD NMR spectrum (figure 1.3) and, in particular, the differences between the ^1H and the STD NMR spectra were a clear indication of binding between the antibody and the tetrapeptide.

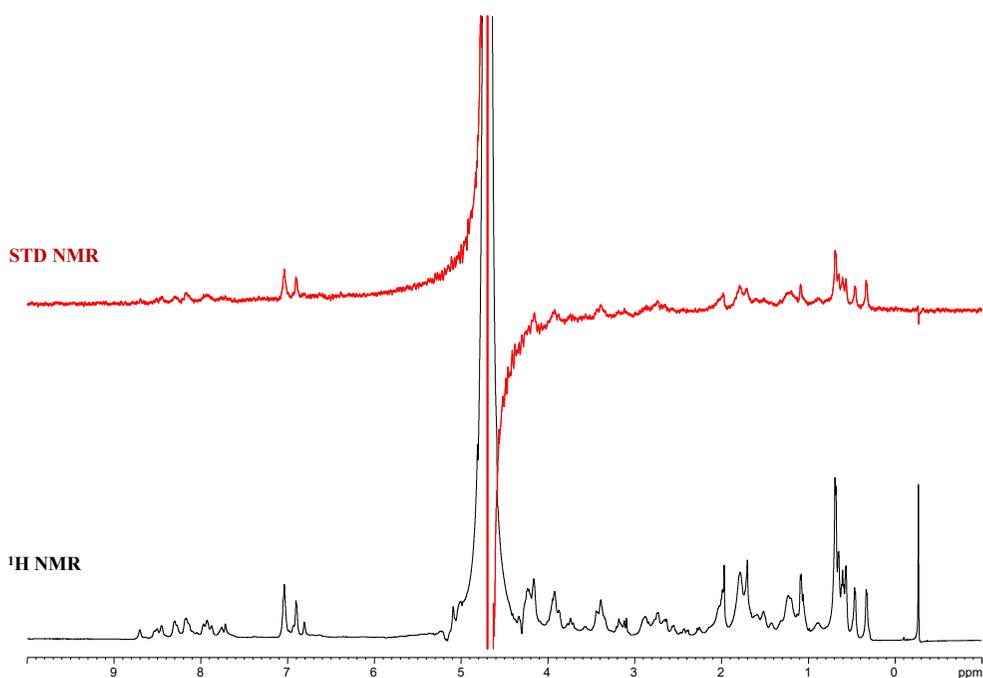


Figure 1.3. STD NMR experiments acquired on mAb 2C7 and tetra-peptide TMCP2 with a molar ratio of 1:5 and temperature of 283K.

The high number of signals in the spectra did not allow neither to distinguish among the different branches (single CP2 peptides) nor to obtain an epitope mapping. However, CH_3 protons of Ile and Leu at both 6 and 11 positions in the aliphatic region

as well as the aromatic (meta and para) protons of Phe showed clear STD signals (figure 1.3).

The diffusion coefficient of TMPC2 in solution was altered upon binding to mAb, another indication of the molecular peptide-antibody interaction (figure 1.4). Indeed, DOSY NMR experiments in the free and bound states were acquired to detect the diffusion properties of molecules in solution. Diffusion decays were fitted as described in Chapter II (§ 2.1.5) and diffusion coefficient values were obtained accordingly (figure 1.4).

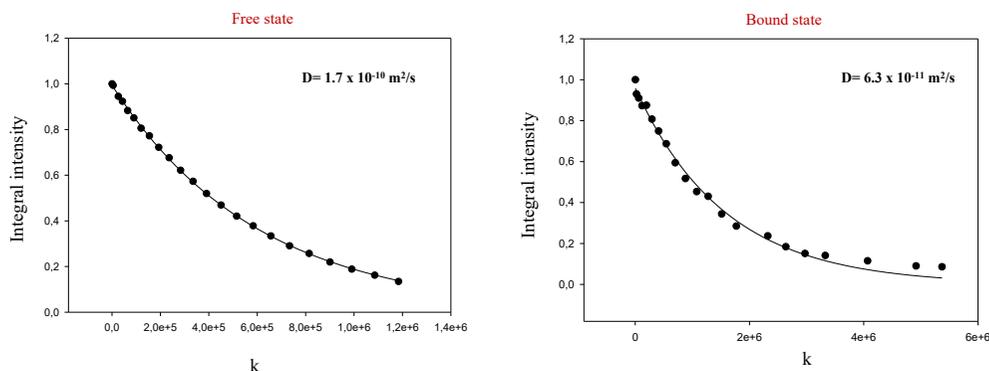


Figure 8.4. DOSY NMR experiments of TMCP2 in the free (left) and bound (right) states. The variation in these values reflected a different hydrodynamic radius of the peptide in the absence (1.4 nm) and in the presence of the antibody (3.8 nm).

1.3 Binding analysis between mAb 2C7 and CP2

To better define the details of the binding, we investigated the single circularized peptide CP2 (figure 1.2) in interaction with mAb 2C7. The assignment of this peptide was studied by means of homo- and hetero-nuclear 2D NMR experiments. Then, STD NMR, WaterLOGSY and CPMG experiments were carried out to achieve molecular details of the interaction.

To unveil information on the interactions between mAb 2C7 and the circularized peptide CP2, STD NMR spectra were acquired (figure 1.5). Comparing the STD (red spectrum) and the corresponding off-resonance (black spectrum) several enhancements of peptide protons were detected, indicative of a good recognition of CP2 by mAb 2C7, while other protons were likely excluded from the binding pocket and therefore not recognized. Changes in intensities and multiplicity of signals were observed in the STD spectrum, indicative of a selective binding with 2C7. Signals of amino acid protons of CP2 resonating in isolated regions of the spectrum were highlighted in the STD spectrum and indicated on the peptide structure (figure 1.5).

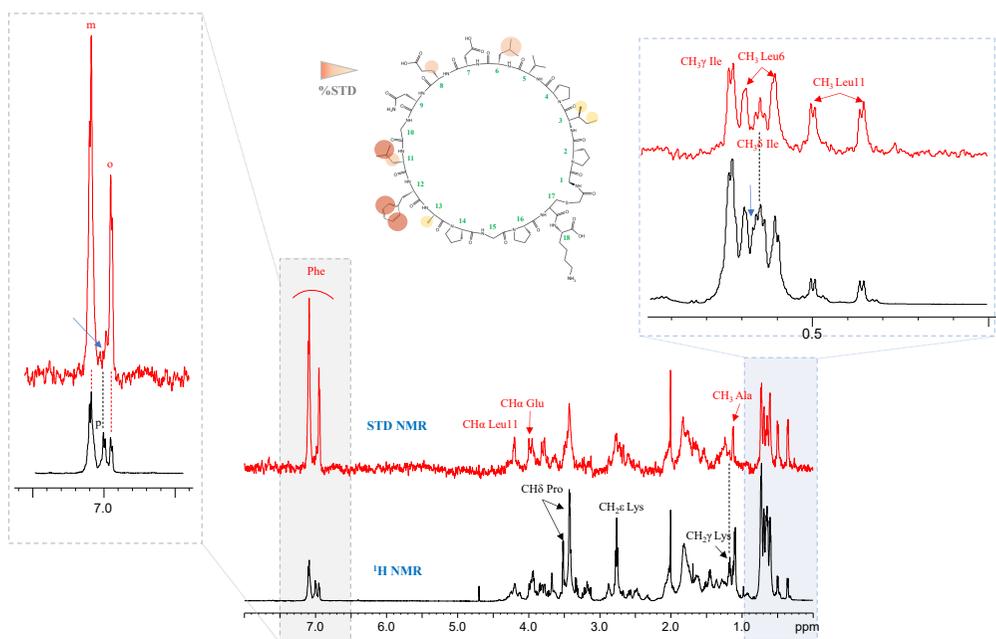


Figure 1.5. STD NMR experiments on mAb 2C7 and peptide CP2 with a molecular ratio of 1:50 and temperature of 283 K. The STD signals in the on-resonance (red) corresponded to the peptide protons interacting with mAb 2C7. Comparing the STD spectrum with the off-resonance (black), changes in multiplicity were observed, as indicated by blue arrows in the aromatic and aliphatic regions. The position of aromatic protons belonging to Phe were indicated as meta (m), orto (o), para (p).

In particular, Phe and Leu at position 11 of CP2 were the main residues involved in the interaction; notably, the proton in para did not show any STD signal (δ at 6.99 ppm, figure 1.5). CH α Glu as well as CH $_3$ Leu at position 6 also provided STD responses. Moreover, the multiplet at around 0.645 ppm observed in the off-resonance became a triplet in the STD NMR spectrum, meaning a selective binding of mAb 2C7 and CH $_3$ Ile.

Moreover, the molecular binding between mAb 2C7 and CP2 mediated by water molecules was investigated by waterLOGSY NMR experiments (figure 1.6). As described in Chapter II (§ 2.1.4), changes in the peak phase passing from the free to the bound state were indicative of binding. Interestingly, the positive signals in waterLOGSY acquired on CP2 bound to the antibody (figure 1.6) corresponded to those detected in the STD spectrum (figure 1.5), meaning that most of peptide protons involved in the binding with IgG were also mediated by surrounding water molecules.

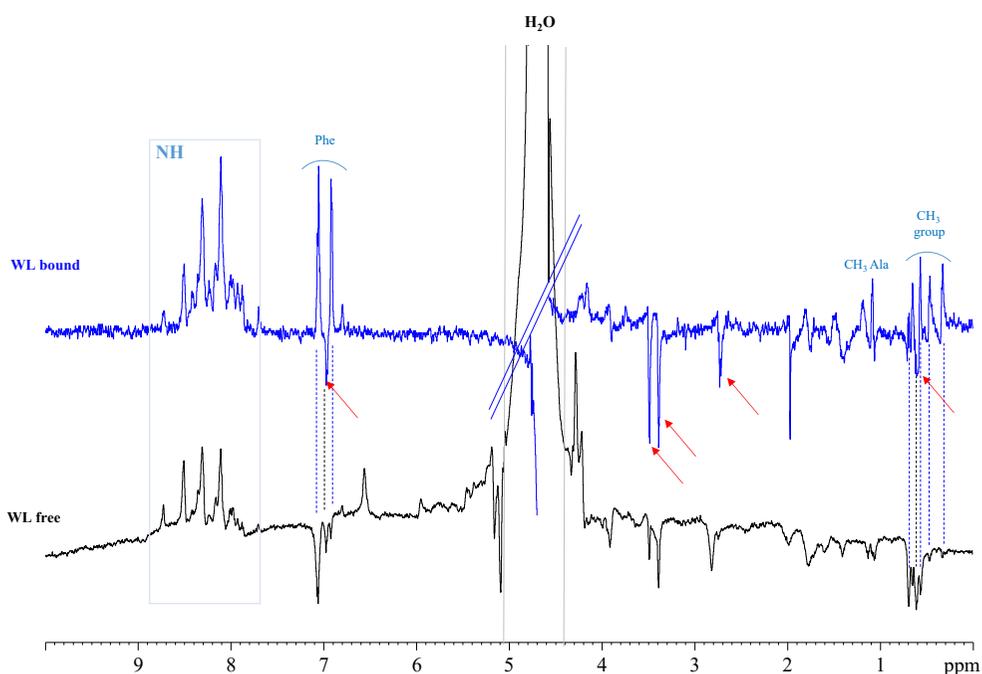


Figure 1.6. WaterLOGSY NMR experiment on mAb 2C7 and peptide CP2 with a molecular ratio of 1:20 and temperature of 283 K. In WL of CP2 bound to mAb (blue), the negative signals represented the protons not involved in the interactions. The positive responses corresponded to the STD signals (figure 8.4). Red arrows indicated the signals that remained negative in the bound state; interestingly those protons did not give STD response.

Relaxation experiments performed through 1D CPMG experiments (see Chapter II, § 2.1.5) supported and completed the binding data. Indeed, the interactions between mAb 2C7 and CP2 were further assessed by calculating differences in transverse relaxation time (T_2) between the free and bound states (figure 1.7).

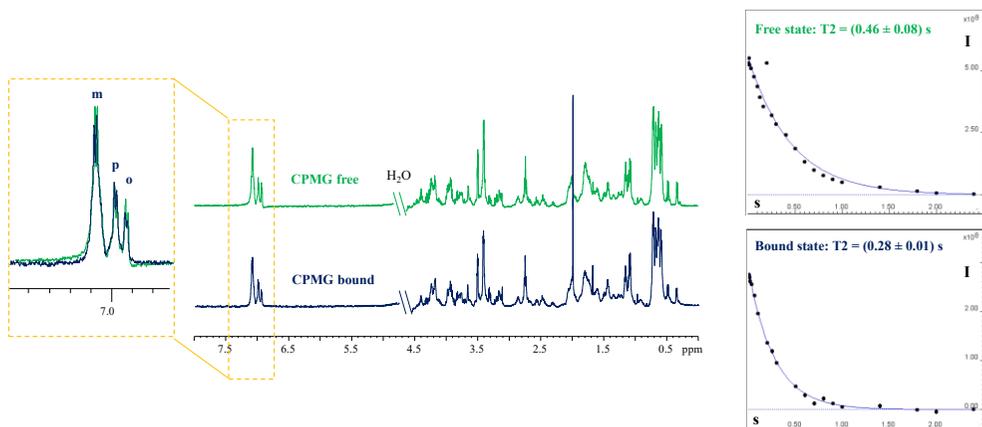


Figure 1.7. CPMG NMR experiments on mAb 2C7 and CP2 at molecular ratio of 1:50 and temperature 283K. Comparing the CPMG spectra acquired on CP2 in absence of the antibody (green) and when bound to mAb 2C7 (blue), a decrease of some signals was detected. Interestingly, signals with lower CPMG in the bound state corresponded to those giving STD enhancements. An example was shown by the aromatic region: the central signal at 6.9 ppm belonging to para-CH Phe displayed the same intensity in the free and bound states. The signals' integration at different mixing time were fitted in mono-exponential curves (see Chapter VIII) that provided the relaxation time values (right panel) with confident level of 95%.

Comparing the CPMG experiments of the peptide alone and in the mixture with the antibody, a decrease of signals in the bound state was observed. A quantitative

analysis of T_2 relaxation time obtained by integrating proton signals in CPMG acquired at different mixing times supported the binding results. The values were fitted in mono-exponential curves to provide T_2 values in the free and bound states (figure 1.7).

1.4 Discussion

The immunogenic activity and the easy and economic production of TMCP2, a peptide mimic of LOS 2C7 epitope, satisfies criteria for a good gonococcal vaccine candidate against gonorrhoea.

Since the mechanism of recognition between TMCP2 and the monoclonal antibody mAb for 2C7 LOS epitope has not been studied from a molecular viewpoint, we analyzed (further experiments currently ongoing) the molecular binding process NMR experiments demonstrating the occurrence of the interaction between mAb 2C7 and TMCP2. Given the high number of protons in the tetrameric mimotope, the single CP2 peptide branch was also investigated.

Overall, STD, WaterLOGSY and CPMG NMR experiments confirmed the molecular recognition of peptide CP2 by mAb 2C7. Interestingly, the protons giving positive signals in the STD and WaterLOGSY corresponded to those that decreased in CPMG experiments. We also considered the binding modes between the single Fab of mAb 2C7 and the peptide. Our results were fully comparable to the full IgG antibody, meaning that the contribution to the binding was mediated only by the Fab branch (data not shown).

Tr-NOESY and conformational analyses on TMCP2 peptide upon binding to mAb 2C7 is currently ongoing. Finally, we plan to evaluate the NMR binding of mAb 2C7 to *Ng* LOS together with competition experiments with TMCP2 mimotope.

BIBLIOGRAPHY

- 1 D.M. Oswald, B.A. Cobb. “Emerging glycobiology tools: A renaissance in accessibility” *Cellular immunology*, 333 (2018) 2–8.
- 2 G.W. Hart. “Thematic Minireview Series on Glycobiology and Extracellular Matrices: Glycan Functions Pervade Biology at All Levels” *J. Biol. Chem.* 288 (2013) 6903–6903.
- 3 A. Fernandez-Tejada, F. J. Cacada, J. Jimenéz-Barbero. “Glycans in medicinal chemistry: An underexploited resource” *ChemMedChem* 10 (2015) 1291-1295.
- 4 P. Valverde, A. Ardà, N. C. Reichardt, J. Jiménez-Barbero, A. Gimneo. “Glycans in drug discovery” *Med Chem Commun* 10 (2019) 1678.
- 5 “Relevance of glycans in the interaction between T lymphocyte and the antigen presenting cell”
W. Gómez-Henao, E. Patricia Tenorio, F. R. Chávez Sanchez, M. Cuéllar Mendoza, R. Lascurain Ledezma, E. Zenteno.” Relevance of glycans in the interaction between T lymphocyte and the antigen presenting cell” *INTERNATIONAL REVIEWS OF IMMUNOLOGY* 40, 4 (2021) 274–288.
- 6 T. H. Mogensen “Pathogen recognition and inflammatory signaling in innate immune defenses” *Clin. Microbiol. Rev.*, 22 (2009) 240–273.
- 7 S. S. Pinho, C. A. Reis. “Glycosylation in cancer: mechanisms and clinical implications” *NATURE REVIEWS* 15 (2015) 540-555.
- 8 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* 6, 4 (2019) 3113-3120.
- 9 R. K. Yu, Y. T. Tsai, T. Ariga and M. Yanagisawa. “Structures, biosynthesis, and functions of gangliosides—An overview” *J. Oleo Sci.*, 60, 10 (2011) 537–544.

-
- 10 A. Varki and T. Angata. "Siglecs- the major subfamily of I-type lectins." *Glycobiology* 16 (2006) 1R–27R.
- 11 M. E. Taylor and K. Drickamer. *Introduction to glycobiology*. 3rd edition. Oxford university press Inc., New York (2011).
- 12 K. Talabnin, C. Talabnin, *et al.* "Increased expression of the high-mannose M6N2 and NeuAc3H3N3M3N2F tri-antennary *N*-glycans in cholangiocarcinoma" *Oncol Lett.* 15, 1 (2018) 1030–1036.
- 13 M. Nakano, S. K. Mishra, Y. Tokoro, K. Sato, K. Nakajima, Y. Yamaguchi, N. Taniguchi, Y. Kizuka. "Bisecting GlcNAc is a general suppressor of terminal modification of *N*-glycan" *Mol Cell Proteomics Papers in Press.* 2 (2019).
- 14 B. Adamczyk, T. Tharmalingam, P. M. Rudd. "Glycans as cancer biomarker" *Biochim. Biophys. Acta* 1820 (2012) 1347–1353.
- 15 A. Varki, R.L. Schnaar, R. Schauer. "Sialic Acids and Other Nonulosonic Acids" *Essentials of Glycobiology* 3rd ed. Cold Spring Harbor Laboratory Press (2017) 179–195.
- 16 A.N. Samraj, O.M. Pearce, H. Laubli, A.N. Crittenden, A.K. Bergfeld, K. Banda, C.J. Gregg, A.E. Bingman, P. Secrest, S.L. Diaz, N.M. Varki, A. Varki. "A red meat-derived glycan promotes inflammation and cancer progression" *Proc Natl Acad Sci USA* 112, 2 (2015) 542–547.
- 17 V. S. Mahajan, S. Pillai. "Sialic acids and autoimmune disease" *Immunological reviews*, 269, 1 (2016) 145–161.
- 18 K. Kawanishi, C. Dhar, R. Do, N. Varki, P. Gordts, A. Varki. "Human species-specific loss of CMP-N-acetylneuraminic acid hydroxylase enhances atherosclerosis

via intrinsic and extrinsic mechanisms” *Proc Natl Acad Sci USA* 116, 32 (2019) 16036–16045.

19 Y.-C. Chang, V. Nizet. “The interplay between Siglecs and sialylated pathogens” *Glycobiology* 24, 9 (2014) 818–825.

20 A. Varki. “Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins” *Nature* 446 (2007) 1023–1029.

21 T. Angata, A. Varki. “Chemical Diversity in the Sialic Acids and Related α -Keto Acids: An Evolutionary Perspective” *Chem. Rev.* 102 (2002) 439–470.

22 R. Schauer, P. Johannis, J.P. Kamerling. “Exploration of the Sialic Acid World” *Adv Carbohydr Chem Biochem.* 75, (2018) 1-213.

23 A. J. Cagnoni, J. M. Pérez Sáez, G. A. Rabinovich, K. V. Mariño. “Turning-Off Signaling by Siglecs, Selectins, and Galectins: Chemical inhibition of Glycan-Dependent interactions in Cancer” *Frontiers in oncology*, 6 (2016) 109.

24 F. Li, J. Ding. “Sialylation is involved in cell fate decision during development, reprogramming and cancer progression” *Protein Cell* 10, (2019) 550–565.

25 A. Varki, M.E. Etzler, R.D. Cummings, J.D. Esko *et al.* “Discovery and Classification of Glycan-Binding Proteins” (Eds.), *Essentials of Glycobiology*. (2nd ed.). Cold Spring Harbor Laboratory Press (2009).

26 A. Molinaro, O. Holst, F. Di Lorenzo, *et al.* “Chemistry of lipid A: at the heart of innate immunity” *Chem Eur J* 21 (2015) 500–19.

27 S. Brown, J. P. Santa Maria Jr, S. Walker. “Wall Teichoic Acids of Gram-Positive Bacteria” *Annu Rev Microbiol.* 67 (2013)

28 M. Rajagopal, S. Walker. “Envelope Structures of Gram-Positive Bacteria” *Curr Top Microbiol Immunol.* 404 (2017) 1–44.

29 R. van Dalen, A. Peschel, N. M. va Sorge. “Wall Teichoic Acid in *Staphylococcus aureus* Host Interaction” *Trends in Microbiology* 28 (2020) 12.

30 C. Soliman, G.B. Pier, P.A. Ramsland. “Antibody Recognition of Bacterial Surfaces and Extracellular Polysaccharides” *Current Opinion in Structural Biology* 62 (2020) 48–55.

31 B. Krismer, C. Weidenmaier, A. Zipperer, A. Peschel. “The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota” *Nat Rev Microbiol.* 12, 15, 11 (2017) 675-687.

32 S. Lehar, T. Pillow, M. Xu, *et al.* “Novel antibody–antibiotic conjugate eliminates intracellular *S. aureus*” *Nature* 527 (2015) 323–328.

33 G. Xia, T. Kohler, A. Peschel. “The wall teichoic acid and lipoteichoic acid polymers of *Staphylococcus aureus*” *International Journal of Medical Microbiology* 300 (2010) 148–154.

34 J.G. Swoboda, J. Campbell, T.C. Meredith, S. Walker. “Wall teichoic acid function, biosynthesis, and inhibition” *Chem Eur J* (2010).

35 D. Keinhörster, S. E. George, C. Weidenmaier, C. Wolz. “Function and regulation of *Staphylococcus aureus* wall teichoic acids and capsular polysaccharides” *International Journal of Medical Microbiology* 309 (2019) 151333.

36 V. Winstel, G. Xia, A. Peschel. “Pathways and roles of wall teichoic acid glycosylation in *Staphylococcus aureus*” *Int. J. Med. Microbiol.* 304, (2014) 215–221.

37 D. Gerlach, Y. Guo, C. De Castro, S.-H. Kim, K. Schlatterer, F.-F. Xu, C. Pereira, P.H. Seeberger, S. Ali, J. Codée, W. Sirisarn, B. Schulte, C. Wolz, J. Larsen, A. Molinaro, B.L. Lee, G. Xia, T. Stehle, A. Peschel, *Nature* 563, 7733 (2018) 705–709.

38 ANTIBODY STRUCTURE AND FUNCTION, THE STRUCTURE OF AN ANTIBODY IS RELATED TO ITS FUNCTION, Chapter 4

-
- 39 M. L. Chiu, D. R. Goulet, A. Teplyakov, G. L. Gilliland. “Antibody Structure and Function: The Basis for Engineering Therapeutics” *Antibodies* 8 (2019) 55.
- 40 K. Chen *et al.* “Immunoglobulin D enhances immune surveillance by activating antimicrobial, pro-inflammatory and B cell-stimulating programs in basophils” *Nat Immunol.* 10, 8 (2009) 889–898.
- 41 A. Laurence, M. Aringer. “Effector Mechanisms in Autoimmunity” *The Autoimmune Disease Sixth Edition*, (2020).
- 42 V. Bruno, G. Battaglia, F. Nicoletti. “The advent of monoclonal antibodies in the treatment of chronic autoimmune diseases” *Neurol Sci* 3 (2011) S283–S288.
- 43 J. Bazan, I. Calkosinski, A. Gamian. “Phage display—A powerful technique for immunotherapy” *Hum Vaccin Immunother.* 8, 12 (2012) 1817–1828.
- 44 L. Ledsgaard, M. Kilstrup, A. Karatt-Vellatt, J. McCafferty, A. H. Lausts. “Basics of Antibody Phage Display Technology” *Toxins (Basel)* 9, 10, 6 (2018) 236.
- 45 G. A. Rabinovich, Y. van Kooyk, B.A. Cobb. “Glycobiology of immune responses: Glycobiology of immune responses” *Ann. N. Y. Acad. Sci.* 1253 (2012) 1–15.
- 46 A. Varki. “Biological roles of glycans” *Glycobiology* 27, 1, (2017) 3–49.
- 47 A. Varki, R.D. Cummings, J.D. Esko. *et al.*, editors. “Glycan-Binding Proteins” *Essentials of Glycobiology* 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press (2015-2017).
- 48 M. S. Macauley, P. R. Crocker, J. C. Paulson. “Siglec-mediated regulation of immune cell function in disease.” *Nat. Rev. Immunol.* 14, 10, (2014) 653– 666.
- 49 P. R. Crocker, A. Varki. “Siglecs in the immune system” *Immunology* 103, (2001) 137- 145.

50 P.R. Crocker, “Siglecs in innate immunity” *Curr. Opin. Pharmacol.* 5 (2005) 431e437.

51 A. Varki, R.L. Schnaar, P.R. Crocker. “I-Type Lectins” (2017) In: Varki A, Cummings RD, Esko JD, *et al.*, editors. *Essentials of Glycobiology*. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017. 35.

52 P. R. Crocker, E.A. Clark, M. Filbin, S. Gordon, Y. Jones, J.H. Kehrl, S. Kelm, N. Le Douarin, L. Powell, J. Roder, R.L. Schnaar, D.C. Sgroi, K. Stamenkovic, R. Schauer, M. Schachner, T.K. van den Berg, P.A. van der Merwe, S.M. Watt, A. Varki. “Siglecs: a family of sialic-acid binding lectins” *Glycobiology*, 8 (1998) 2.

53 S. Pillai, I.A. Netravali, A. Cariappa, H. Mattoo. “Siglecs and immune regulation” *Annu. Rev. Immunol.* 30 (2012) 357e392.

54 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” *Mol. Cell* 1 (1998) 719e728.

55 K. L. Hudson, G. J. Bartlett, R. C. Diehl, J. Agirre, T. Gallagher, L. L. Kiessling, D. N. Woolfson. “Carbohydrate-aromatic interactions in proteins” *J. Am. Chem. Soc.* 137, (2015) 15152–15160.

56 H. Attrill, *et al.* “Siglec-7 undergoes a major conformational change when complexed with the α 2,8-disialylganglioside GT1b” *J. Biol. Chem.* 281, (2006) 32774–32783.

57 J. V. Ravetch, L.L. Lanier. “Immune inhibitory receptors” *Science* 290 (2000) 84–89.

58 T. Ulyanova, D.D. Shah, M.L. Thomas. "Molecular cloning of MIS, a myeloid inhibitory siglec, that binds protein-tyrosine phosphatases SHP-1 and SHP-2" *THE JOURNAL OF BIOLOGICAL CHEMISTRY* 276, 17 (2001) 14451–14458.

59 T. Avril, S.D. Freeman, H. Attrill, R.G. Clarke, P.R. Crocker. "Siglec-5 (CD170) can mediate inhibitory signaling in the absence of immunoreceptor tyrosine based inhibitory motif phosphorylation" *J. Biol. Chem.* 280 (2005) 19843e19851.

60 Z. Yu, M. Maoui, L. Wu, D. Banville, S. Shen. "mSiglec-E, a novel mouse CD33-related siglec (sialic acid-binding immunoglobulin-like lectin) that recruits Src homology 2 (SH2)-domain-containing protein tyrosine phosphatases SHP-1 and SHP-2" *Biochem. J.* 353 (2001) 483e492.

61 H. Arase, L.L. Lanier. "Specific recognition of virus-infected cells by paired NK receptors" *Rev. Med. Virol.* 14 (2004) 83e93.

62 P.R. Crocker, J.C. Paulson, A. Varki. "Siglecs and their roles in the immune system" *Nat. Rev. Immunol.* 7 (2007) 255e266.

63 R.H. Quarles. "Myelin-associated glycoprotein (MAG): past, present and beyond" *J. Neurochem.* 100 (2007) 1431e1448.

64 A.S. O'Neill, T.K. van den Berg, G.E. Mullen. "Sialoadhesin - a macrophage-restricted marker of immunoregulation and inflammation" *Immunology.* 138, 3 (2013) 198–207.

65 K. F. Bornhöfft, T. Goldammer, A. Rebl, S. P. Galuska "Siglecs: A journey through the evolution of sialic acid-binding immunoglobulin-type lectins" *Developmental & Comparative Immunology* 86 (2018) 219-231.

66 T. Angata, C.M. Nycholat, M.S. Macauley. "Therapeutic Targeting of Siglecs using Antibody- and Glycan-based Approaches" *Trends Pharmacol. Sci.* 36 (2015) 645–660.

-
- 67 T. Angata, E.H. Margulies, E.D. Green, A. Varki. "Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms" *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 13251e13256.
- 68 B. S. Bochner, N. Zimmermann. "Role of siglecs and related glycan-binding proteins in immune responses and immunoregulation" *J. Allergy Clin. Immunol.* 135, (2015) 598–608.
- 69 J. Lübbers, E. Rodríguez, Y. van Kooyk. "Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions" *Front Immunol.* 9 (2018) 2807.
- 70 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, 1 (2012) 37–48.
- 71 B. Khatua, S. Roy, C. Mandal. "Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens" *Indian J Med Res.* 138, 5 (2013) 648–662.
- 72 M. Bax, M.L. Kuijf, A.P. Heikema, *et al.* "Campylobacter jejuni lipooligosaccharides modulate dendritic cell-mediated T cell polarization in a sialic acid linkage-dependent manner" *Infect Immun.* 79, 7 (2011) 2681–2689.
- 73 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.
- 74 J.E. Hudak, S.M. Canham, C.R. Bertozzi, "Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion" *Nat. Chem. Biol.* 10 (2014) 69-75.
- 75 C. Jandus, K.F. Boligan, O. Chijioke, H. Liu, M. Dahlhaus, T. Demoulins, C. Schneider, M. Wehrli, R.E. Hunger, G.M. Baerlocher, H.U. Simon, P. Romero, C.

Munz, S. von Gunten. “Interactions between Siglec-7/9 receptors and ligands influence NK cell-dependent tumor immunosurveillance” *J. Clin. Invest.* 124, (2014) 1810-1820.

76 M.S. Macauley, P.R. Crocker, J.C. Paulson, “Siglec regulation of immune cell function in disease” *Nature reviews. Immunology* 14 (2014) 653-666.

77 N. Kawasaki, C. Rademacher, J.C. Paulson. “CD22 regulates adaptive and innate immune responses of B cells” *J. Innate Immun.* 3 (2011) 411–19.

78 J. Ereño-Orbea, *et al.* “Molecular basis of human CD22 function and therapeutic targeting” *Nat. Commun.* (2017) 8.

79 C. Di Carluccio, *et al.* “Characterization of the dynamic interactions between complex *N*-glycans and human CD22” *Chembiochem* 21 (2019) 129.

80 S. Han, B.E. Collins, P. Bengtson, J.C. Paulson. “Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking” *Nat. Chem. Biol.* 1 (2005) 93–97.

81 T. Angata. “Associations of genetic polymorphisms of Siglecs with human diseases” *Glycobiology* 24 (2014) 785–93.

82 Y. Hatta, N. Tsuchiya, M. Matsushita, M. Shiota, K. Hagiwara, K. Tokunaga. “Identification of the gene variations in human CD22” *Immunogenetics* 49 (1999) 280–86.

83 J. Ai, A. Advani. “Current status of antibody therapy in ALL” *Br. J. Haematol.* 168 (2015) 471–80.

84 J. Müller, L. Nitschke. “The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease” *Nat. Rev. Rheumatol.*, 10 (2014) 422–428.

-
- 85 N. Yamakawa, Y. Yasuda, A. Yoshimura, A. Goshima, P. R. Crocker, G. Vergoten, Y. Nishiura, T. Takahashi, S. Hanashima, K. Matsumoto, Y. Yamaguchi, H. Tanaka, K. Kitajima, C. Sato. “Discovery of a new sialic acid binding region that regulates Siglec-7” *Scientific Reports*, 10 (2020) 1.
- 86 S. Varchetta, *et al.* “Lack of siglec-7 expression identifies a dysfunctional natural killer cell subset associated with liver inflammation and fibrosis in chronic HCV infection” *Gut*. 65 (2016)1998–2006.
- 87 E. Brunetta, *et al.* “The decreased expression of siglec-7 represents an early marker of dysfunctional natural killer-cell subsets associated with high levels of HIV-1 viremia” *Blood* 114 (2009) 3822–30.
- 88 T. Avril, E.R. Wagner, H.J. Willison, P.R. Crocker. “Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on *Campylobacter jejuni* lipooligosaccharides” *Infect Immun* 74 (2006) 4133–4141.
- 89 D. Lamprinaki, P. Garcia-Vello, R. Marchetti, C. Hellmich, K.A. McCord, K.M. Bowles, M.S. Macauley, A. Silipo, C. De Castro, P.R. Crocker PR, N. Juge. “Siglec-7 Mediates Immunomodulation by Colorectal Cancer-Associated *Fusobacterium nucleatum* ssp. *animalis*” *Front. Immunol.* 12 (2021) 744184.
- 90 G. Whitney, S. Wang, H. Chang, K. Y. Cheng, P. Lu, X. D. Zhou, W. P. Yang, M. McKinnon and M. Longphre. “A new siglec family member, siglec-10, is expressed in cells of the immune system and has signaling properties similar to CD33” *Eur. J. Biochem.* 268 (2001) 6083–6096.
- 91 G. Y. Chen, N. K. Brown, P. Zheng and Y. Liu. “Siglec-G/10 in self-nonsel self discrimination of innate and adaptive immunity.” *Glycobiology* 24 (2014) 800–806.

92 G. Y. Chen, J. Tang, P. Zheng and Y. Liu. “CD24 and Siglec-10 selectively repress tissue damage-induced immune responses” *Science*, 323 (2009) 1722–1725.

93 E. Bandala-Sanchez, N. G. Bediaga, E. D. Goddard-Borger, K. Ngui, G. Naselli, N. L. Stone, A. M. Neale, L. A. Pearce, A. Wardak, P. Czabotar, T. Haselhorst, A. Maggioni, L. A. Hartley-Tassell, T. E. Adams and L. C. Harrison. “CD52 glycan binds the proinflammatory B box of HMGB1 to engage the Siglec-10 receptor and suppress human T cell function” *Proc. Natl. Acad. Sci. U. S. A.*, 115 (2018) 7783–7788.

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

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

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

97 V. Phongsisay. “The immunobiology of *Campylobacter jejuni*: Innate immunity and autoimmune diseases.” *Immunobiology*, 221 (2016) 535–54.

98 S. Patel, N. Mathivanan, A. Goyal. “Bacterial adhesins, the pathogenic weapons” *Biomedicine & Pharmacotherapy* 93 (2017) 763–771.

99 D. H. Stones, A. M. Krachler “Against the tide: the role of bacterial adhesion in

host colonization” *Biochemical Society Transactions* 44 (2016) 1571–1580.

100 A. A. Nash, R. Dalziel, J. R. Fitzgerald. “Attachment to and Entry of Microorganisms into the Body, in *Mims' Pathogenesis of Infectious Disease*” *Mims' Pathogenesis of Infectious Disease Sixth Edition* (2015).

101 P. Klemm, M. A. Schembri. “Bacterial adhesins: structure and function” *Int. J. Med. Microbiol.* 290 (2000) 27-35.

102 O. Itzhak, E. Bayer, A. Soman. “Bacterial adhesion” *The Prokaryotes: Human Microbiology* 107 (2013) 123.

103 K. A. Kline, S. Falker, S. Dahlberg, S. Normark, B. Henriques-Normark. “Bacterial adhesins in host microbial interactions” *Cell Host & Microbe* (2009) Elsevier Inc.

104 K. N. Hallstrom, B. A. McCormick. “Pathogenicity Islands: Structure, and Roles in Bacterial Pathogenesis” *Molecular Medical Microbiology (Second Edition)* (2015) 303-314.

105 L. F. Milles, K. Schulten, H. E. Gaub, R. C. Bernardi. “Molecular mechanism of extreme mechanostability in a pathogen adhesin” *Science* 359 (2018)1527–1533.

106 Y.-A. Que, P. Moreillon “Infective Endocarditis” *Nat. Rev. Cardiol.* 8, (2011) 322–336.

107 B.A Bensing, Q. Li, D. Park, C. B. Lebrilla, P.M. Sullam. “Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis” *Glycobiology*, 28, 8 (2018) 601–611.

108 M.O. Gayta'n, A.K. Singh, S.A. Woodiga, S.A. Patel, S.S. An, A. Vera-Ponce de Leon, *et al.* “A novel sialic acid-binding adhesin present in multiple species

contributes to the pathogenesis of Infective endocarditis” *PLoS Pathog* 17, 1 (2021) e1009222.

109 K. Werdan, S. Dietz, B. Löffler, S. Niemann, H. Bushnaq, R. E. Silber, G. Peters and U. Müller-Werdan. “Mechanisms of infective endocarditis: pathogen-host interaction and risk states” *Nat. Rev. Cardiol.*, 11, 1 (2014) 35–50.

110 S. Chamat-Hedemand, A. Dahl, L. Østergaard, M. Arpi, E. Fosbøl, J. Boel, L. B. Oestergaard, T. K. Lauridsen, G. Gislason, C. Torp-Pedersen and N. E. Bruun. “Prevalence of Infective Endocarditis in Streptococcal Bloodstream Infections Is Dependent on Streptococcal Species” *Circulation*, 2020, 142, 8 (2020) 720–730.

111 C. Del Giudice, E. Vaia, D. Liccardo, F. Marzano, A. Valletta, G. Spagnuolo, N. Ferrara, C. Rengo, A. Cannavo and G. Rengo. “Infective Endocarditis: A Focus on Oral Microbiota” *Microorganisms*, 9, (2021) 1218.

112 B. A. Bensing, L. Li, O. Yakovenko, M. Wong, K. N. Barnard, T. M. Iverson, C. B. Lebrilla, C. R. Parrish, W. E. Thomas, Y. Xiong and P. M. Sullam. “Recognition of specific sialoglycan structures by oral streptococci impacts the severity of endocardial infection” *PLoS Pathog.*, 15 (2019) e1007896.

113 T. M. Pyburn, B. A. Bensing, Y. Q. Xiong, B. J. Melancon, T. M. Tomasiak, N. J. Ward, V. Yankovskaya, K. M. Oliver, G. Cecchini, G. A. Sulikowski, M. J. Tyska, P. M. Sullam, T. M. Iverson. “A Structural Model for Binding of the Serine-Rich Repeat Adhesin SLBR-B to Host Carbohydrate Receptors” *PLoS Pathog.*, 7 (2011) e1002112.

114 R. Agarwal, B. A. Bensing, D. Mi, P. N. Vinson, J. Baudry, T. M. Iverson and J. C. Smith. “Structure based virtual screening identifies small molecule effectors for the sialoglycan binding protein SLBR-H” *Biochem. J.*, 477, 19 (2020) 3695–3707.

115 B.A. Bensing *et al.* “Selectivity and engineering of the sialoglycan-binding spectrum in Siglec-like adhesins” *BioXiV* (2019).

116 L. Deng, B. A. Bensing, S. Thamadilok, H. Yu, K. Lau, X. Chen, S. Ruhl, P. M. Sullam and A. Varki. “Oral Streptococci Utilize a Siglec-Like Domain of Serine-Rich Repeat Adhesins to Preferentially Target Platelet Sialoglycans in Human Blood” *PLoS Pathog.*, 10, 12 (2014) e1004540.

117 B. A. Bensing, Z. Khedri, L. Deng, H. Yu, A. Prakobphol, S. J. Fisher, X. Chen, T. M. Iverson, A. Varki and P. M. Sullam. “Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins” *Glycobiology*, 26, 11 (2016), 1221–1233.

118 Y. Narimatsu *et al.* “An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells” *Molecular Cell*, 75 (2019) 394-407.

119 C. M. Dobson. “Biophysical Techniques in Structural Biology”, *Annu. Rev. Biochem.* 88 (2019) 25–33.

120 M. A. Johnson, B. M. Pinto “NMR spectroscopic and molecular modeling studies of protein–carbohydrate and protein–peptide interactions” *Carbohydrate Research* 339 (2004) 907–928.

121 W. Becker, K. C. Bhattiprolu, N. Gubensäk, K. Zangger. “Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy” *ChemPhysChem*, 19 (2018) 895 – 906.

122 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.

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

124 B. Meyer, T. Peters, *NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors*, *Angew. Chemie*, 42 (2003) 8, 864-890.

125 B. Meyer, T. Peters. “NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors” *Angew. Chem. Int. Ed.* 42 (2003) 864–890.

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

127 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) 25, 6108–6117.

128 J. 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*, 16 (2010) 26, 7803-7812.

129 C. Raingeval, O. Cala, B. Brion, M. Le Borgne, R. E. Hubbard, I. Krimm “1D NMR WaterLOGSY as an efficient method for fragment-based lead discovery” *Glycobiology* 29 (2019) 2, 124–136.

-
- 130 C. Dalvit, P. Pevarello, M. Tatò, M. Veronesi, A. Vulpetti, M. Sundström. “Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water, *Journal of Biomolecular NMR*” 18 (2000) 65–68.
- 131 R. Huang, I. K.H. Leung. “Protein–Small Molecule Interactions by WaterLOGSY” *Methods in Enzymology*, Academic Press (2019) 477-500.
- 132 E. Rennella, A. Sekhar, L. E. Kay. “Self-Assembly of Human Profilin-1 Detected by Carr–Purcell–Meiboom–Gill Nuclear Magnetic Resonance (CPMG NMR) Spectroscopy” *Biochemistry*, 56 (2017) 692–703.
- 133 L. H. Lucas, C. K. Larive. “Measuring Ligand-Protein Binding Using NMR Diffusion Experiments” *Concepts Magn. Reson.*, 20A (2004) 24-41.
- 134 R. Marchetti, M. J. Dillon, M. N. Burtnick, M. A. Hubbard, M. T. Kenfack, Y. Blériot, C. Gauthier, P. J. Brett, D.P. AuCoin, R. Lanzetta, A. Silipo, A. Molinaro. “Burkholderia pseudomallei capsular polysaccharide recognition by a monoclonal antibody reveals key details toward a biodefense vaccine and diagnostics against melioidosis” *ACS Chem. Biol.*, 10 (2015) 10, 2295–2302.
- 135 T. R. Alderson, L. E. Kay. “NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function” *Cell* 184 (2021).
- 136 S. Giuntini, E. Balducci, L. Cerofolini, E. Ravera, M. Fragai, F. Berti, C. Luchinat. “Characterization of the Conjugation Pattern in Large Polysaccharide–Protein Conjugates by NMR Spectroscopy” *Angew. Chem.* 129 (2017) 15193 – 15197.
- 137 C. Fernandez, G. Wider. “TROSY in NMR studies of the structure and function of large biological macromolecules” *Current Opinion in Structural Biology* 13 (2003) 570–580.

138 <http://wiki.cara.nmr.ch/FrontPage>

139 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.

140 M. P. Williamson. “Chemical Shift Perturbation” *Modern Magnetic Resonance* (2018) 995-1012.

141 I. A. Vakser. “Protein-protein docking: from interaction to interactome” *Biophysical journal*, 107, 8 (2014) 1785–1793.

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

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

144 K. Crampon, A. Giorkallos, M. Deldossi, S. Baud, L. A. Steffanel. “Machine-learning methods for ligand–protein molecular docking” *Drug Discovery Today* 27, 1 (2022) 151-164.

145 W.F. de Azevedo, R. Dias. “Molecular Docking Algorithms” *Bioorg. Med. Chem.*, 16, 20 (2008) 9378-9382.

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

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

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

149 K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. González-Outeiriño, C.R. Daniels, B.L. Foley, R.J.J. Woods. “GLYCAM06: A generalizable biomolecular force field. Carbohydrates” *Comput. Chem.* 29, 4 (2008) 622–655.

150 O. Guvench, S. Mallajosyula, E. Raman, E. Hatcher, K. Vanommeslaeghe, T. Foster, F. Jamison, A.J. MacKerell, “CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling” *Chem. Theory Comput.* 7, 10 (2011) 3162–3180.

151 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling. “ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB” *J. Chem. Theory Comput.* 11, 8 (2015) 3696-3713.

152 D.J. Rigden. “Ab Initio Protein Structure Prediction” (ed.) *From Protein Structure to Function with Bioinformatics* (2009) 3-25.

153 D. S. Patel, R. Pendrill, S.S. Mallajosyula, G. Widmalm, A.D. MacKerell. “Conformational Properties of α - or β -(1→6)-Linked Oligosaccharides: Hamiltonian Replica Exchange MD Simulations and NMR Experiments” *J. Phys. Chem. B* 118, (2014) 2851–2871.

154 A.D. French. “Comparisons of rigid and relaxed conformational maps for cellobiose and maltose” *Carbohydrate Res.* 188 (1989) 206–211.

-
- 155 N.L. Allinger, M. Rahman, H. Lii. “A molecular mechanics force-field (Mm3) for alcohols and ethers” *J. Am. Chem. Soc.*, 112 (1990) 8293–8307.
- 156 A.D. Jr Mackerell. “Empirical force fields for biological macromolecules: overview and issues” *J. Comput. Chem.*, 25 (2004) 1584–1604.
- 157 S. Perez, E. Fadda, O.N. Makshakova. “Computational Modeling in Glycoscience” *Comprehensive Glycoscience*, 2nd edition (2020).
- 158 S. A. Hollingsworth, R.O. Dror. “Molecular Dynamics Simulation for All” *Neuron*, 99, 6 (2018) 1129–1143.
- 159 K. Vanommeslaeghe, O. Guvench, A.D. Jr MacKerell. “Molecular mechanics” *Current pharmaceutical design*, 20, 20 (2014) 3281–3292.
- 160 M. C. Payne, M. P. Teter, D. C. Allan, T. Arias, J. Joannopoulos. “Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients” *Rev. Mod. Phys* 64, 4 (1992) 1045.
- 161 W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein. “Comparison of simple potential functions for simulating liquid water” *J. Chem. Phys.* 79, 2 (1983) 926-935.
- 162 V. Jayalakshmi, N.R. Krishna. “Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand–Receptor Complexes” *J. Magn. Reson.* 155 (2002) 106–118.
- 163 J. Angulo, P.M. Nieto. “STD-NMR: application to transient interactions between biomolecules—a quantitative approach” *Eur. Biophys. J.* 40 (2011) 1357–1369.

164 S. Duan, J.C. Paulson. “Siglecs as Immune Cell Checkpoints in Disease” Annual review of immunology, 38 (2020) 365–395.

165 M. Hedlund, V. Padler-Karavani, N.M. Varki, A. Varki, A. “Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression” Proceedings of the National Academy of Sciences of the United States of America, 105 (2008) 48, 18936–18941.

166 Y. N. Malykh, R. Schauer, L. Shaw. N-Glycolylneuraminic acid in human tumours” Biochimie, 83 (2001) 7, 623–634.

167 A. Varki. “Are humans prone to autoimmunity? Implications from evolutionary changes in hominin sialic acid biology” Journal of autoimmunity 30 (2017) 134–142.

168 T. Angata. “Possible Influences of Endogenous and Exogenous Ligands on the Evolution of Human Siglecs.” Front. Immunol. 9 (2018) 2885.

169 Y. Tang, X. Zeng, J. Liang. “Surface Plasmon Resonance: An Introduction to a Surface Spectroscopy Technique” J Chem Educ. 87 (2010) 7, 742–746.

170 www.perkinelmer.com/alphatechnology

171 R. Marchetti, R. Lanzetta, I.C. Michelow, A. Molinaro, A. Silipo. “Structural Study of Binding of α -Mannosides to Mannan-Binding Lectins” Eur. J. Org. Chem. (2012) 5275–5281.

172 Y. Yuan, *et al.* “Investigation of Binding of UDP-Gal *f* and UDP-[3-F]Gal *f* to UDP-galactopyranose Mutase by STD-NMR Spectroscopy, Molecular Dynamics, and CORCEMA-ST Calculations” *J. Am. Chem. Soc.* 130 (2008) 3157–3168.

173 C. Thibaudeau, R. Stenutz, B. Hertz, T. Klepach, S. Zhao, Q. Wu, I. Carmichael, A. S. Serianni. “Correlated C-C and C-O bond conformations in saccharide

hydroxymethyl groups: parametrization and application of redundant ^1H - ^1H , ^{13}C - ^1H , and ^{13}C - ^{13}C NMR J-couplings” *J. Am. Chem. Soc.* 126 (2004) 15668-15685.

174 W. Koźmiński, D. Nanz, D. “HECADE: HMQC- and HSQC-Based 2D NMR Experiments for Accurate and Sensitive Determination of Heteronuclear Coupling Constants from E.COSY-Type Cross Peaks” *J. Magn. Reson.* 124 (1997) 383–392.

175 S. Jo, Y. Qi, W. Im. “Preferred conformations of *N*-glycan core pentasaccharide in solution and in glycoproteins” *Glycobiology* (2015).

176 R. Marchetti, *et al.* “*Burkholderia pseudomallei* Capsular Polysaccharide Recognition by a Monoclonal Antibody Reveals Key Details toward a Biodefense Vaccine and Diagnostics against Melioidosis” *ACS Chem. Biol.* 10 (2015) 2295–2302.

177 A. Guillot, M. Dauchez, N. Belloy, J. Jonquet, L. Duca, B. Romier, P. Maurice, L. Debelle, L. Martiny, V. Durlach, S. Baud, S. Blaise. “Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans” *Scientific reports* 6 (2016) 35666.

178 S. Jo, Y. Qi, W. Im. “Preferred conformations of *N*-glycan core pentasaccharide in solution and in glycoproteins” *Glycobiology* 26 (2016) 1, 19–29.

179 W. Nishima, N. Miyashita, Y. Yamaguchi, Y. Sugita, S. Re J. “Effect of bisecting GlcNAc and core fucosylation on conformational properties of biantennary complex-type *N*-glycans in solution. *The journal of physical chemistry*” *Phys. Chem. B* 116 (2012) 8504–8512.

180 T. L. Holland, L. M. Baddour, A. S. Bayer, B. Hoen, J. M. Miro and V. G. Fowler Jr. “Infective Endocarditis” *Nat Rev Dis Primers* 2 (2017) 16059.

181 Y. Q. Xionga, B. A. Bensing, A. S. Bayer, H. F. Chambers, P. M. Sullam. “Role of the serine-rich surface glycoprotein GspB of *Streptococcus gordonii* in the pathogenesis of infective endocarditis” *Microb Pathog.* 2008, 45, 4, 297–301.

-
- 182 Y. Takahashi, E. Takashima, K. Shimazu, H. Yagishita, T. Aoba, K. Konishi. “Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by *Streptococcus gordonii* DL1” *Infect Immun.* 74, 1 (2006) 740-743.
- 183 D. Takamatsu, B. A. Bensing, H. Cheng, G. A. Jarvis, I. R. Siboo, J. A. López, J. McLeod Griffiss and P. M. Sullam. “Binding of the *Streptococcus gordonii* surface glycoproteins SLBR-B and SLBR-H to specific carbohydrate structures on platelet membrane glycoprotein Iba” *Molecular Microbiology* 58, 2 (2005) 380–392.
- 184 G.C. Hansson. “Mucins and the Microbiome” *Annual Review of Biochemistry* 89 (2020) 769-793.
- 185 R. K. Yu, Y.T. Tsai, T. Ariga, M. Yanagisawa. “Structures, biosynthesis, and functions of gangliosides—An overview” *J Oleo Sci.* 60 (2011) 10, 537–544.
- 186 F.-Q. Wen, A.A. Jabbar, D.A. Patel, T. Kazarian, L.A. Valentino. “Atherosclerotic Aortic Gangliosides Enhance Integrin-Mediated Platelet Adhesion to Collagen, Arteriosclerosis, Thrombosis, and Vascular Biology” 19 (1999) 519–524.
- 187 N. Sasaki and M. Toyoda. “Vascular Diseases and Gangliosides” *Int. J. Mol. Sci.* 20 (2019) 6362.
- 188 B. Krismer, C. Weidenmaier, A. Zipperer, A. Peschel. “The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota.” *Nat Rev Microbiol.* 15, 11 (2017) 675-687.
- 189 S. Lehar, T. Pillow, M. Xu, *et al.* “Novel antibody–antibiotic conjugate eliminates intracellular *S. aureus*” *Nature* 527 (2015) 323–328.

-
- 190 C. Weidenmaier, A. Peschel. “Teichoic acids and related cell wall glycopolymers in Gram-positive physiology and host interactions” *Nat Rev Micro.* 6 (2008) 276–287.
- 191 E. Boldock, *et al.* “Human skin commensals augment *Staphylococcus aureus* pathogenesis” *Nature Microbiol.* 3 (2018) 881–890.
- 192 K. Kurokawa, K. Takahashi, B.L. Lee. “The staphylococcal surface-glycopolymer wall teichoic acid (WTA) is crucial for complement activation and immunological defense against *Staphylococcus aureus* infection” *Immunobiology* 221, 10 (2016) 1091-101.
- 193 E. Tacconelli *et al.* “Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis” *Lancet Infect Dis.* 18, 3 (2018) 318-327.
- 194 N. Mistretta, M. Brossaud, F. Telles, V. Sanchez, P. Talaga, B. Rokbi. “Glycosylation of *Staphylococcus aureus* cell wall teichoic acid is influenced by environmental conditions” *Sci Rep.* 9, 1 (2019) 3212.
- 195 D. Raafat, M. Otto, K. Reppschläger, J. Iqbal, S. Holtfreter. “Fighting *Staphylococcus aureus* Biofilms with Monoclonal Antibodies” *Trends Microbiol.* 27, 4 (2019) 303-322.
- 196 K. Takahashi, K. Kurokawa, P. Moyo, D.-J. Jung, J.-H. An, L. Chigweshe, E. Paul, B.L. Lee. “Intradermal immunization with wall teichoic acid (WTA) elicits and augments an anti-WTA IgG response that protects mice from methicillin-resistant *Staphylococcus aureus* infection independent of mannose-binding lectin status” *PLoS ONE* 8, 8 (2013) e69739.
- 197 S. M. Lehar, T. Pillow, M. Xu, L. Staben, K.K. Kajihara, R. Vandlen, L. DePalatis, H. Raab, W.L. Hazenbos, J. Hiroshi Morisaki, J.VKim, S. Park, M. Darwish, B.-C. Lee, H. Hernandez, K.M. Loyet, P. Lupardus, R. Fong, D. Yan, C.

Chalouni, E. Luis, Y. Khalfin, E. Plise, J. Cheong, J.P. Lyssikatos, M. Strandh, K. Koefoed, P.S. Andersen, J.A. Flygare, M. Wah Tan, E.J. Brown, S. Mariathasan. “Novel antibody-antibiotic conjugate eliminates intracellular *S. aureus*”, *Nature* 527, 7578 (2015) 323–328.

198 R. Fong, K. Kajihara, M. Chen, I. Hotzel, S. Mariathasan, W.L.W. Hazenbos, P.J. Lupardus. “Structural investigation of human *S. aureus*-targeting antibodies that bind wall teichoic acid. *MAbs*” 10, 7 (2018) 979-991.

199 S. Ali, A. Hendriks, R. Dalen, T. Bruyning, N. Meeuwenoord, H.S. Overkleeft, D.V. Filippov, G.A. Marel, N.M. Sorge, J.D.C. Codée. “(Automated) Synthesis of Well-defined *Staphylococcus Aureus* Wall Teichoic Acid Fragments” *Chem. Eur. J.*, 27, 40 (2021) 10461–10469.

200 R. van Dalen, M.M. Molendijk, S. Ali, K.P.M. van Kessel, P. Aerts, J.A.G. van Strijp, C.J.C. de Haas, J.D.C. Codée, N.M. van Sorge. “Do not discard *Staphylococcus aureus* WTA as a vaccine antigen” *Nature*, 572, 7767 (2019) E1–E2.

201 C. Dalvit. “Homonuclear 1D and 2D NMR Experiments for the Observation of Solvent – Solute Interactions” *J. Mag. Reason.*, B112 (1996) 282 – 288.

202 F. De Biasi, D. Rosa-Gastaldo, X. Sun, F. Mancin, F. Rastrelli. “Nanoparticle-Assisted NMR Spectroscopy: Enhanced Detection of Analytes by Water-Mediated Saturation Transfer” *J. Am. Chem. Soc.* 141 (2019) 4870–4877.

203 www.jphilo.mailway.com

204 P. Schuck. “Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling” *Biophys. J.* 78 (2000) 1606–1619.

205 C.A. Brautigam. “Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data” *Methods Enzym.* 562 (2015) 109–133.

206 <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

207 A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton. “Jalview Version 2-a multiple sequence alignment editor and analysis workbench” *Bioinformatics* 25 (2009) 1189-1191.

208 R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton. “PROCHECK - a program to check the stereochemical quality of protein structures” *J. App. Cryst.* 26 (1993) 283-291.

209 www.glycam.com

210 A.W. Goetz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, and R.C. Walker. “Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born” *J. Chem. Theory Comput.* 8 (2012) 1542-1555.

211 D.A. Case, et al. 2021, Amber 2021, University of California, San Francisco.

212 D.R. Roe, T. E. Cheatham. “PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data” *J. Chem. Theory Comput.*, 9,7 (2013) 3084-3095.

213 P. A. Rice, W. M. Shafer, S. Ram, A. E. Jerse. “*Neisseria gonorrhoeae*: Drug Resistance, Mouse Models, and Vaccine Development” *Annu. Rev. Microbiol.* 71 (2017) 665–86

-
- 214 S. Gulati, B. Zheng, G.W. Reed, X. Su, A.D. Cox, F. St. Michael, J. Stupak, L.A. Lewis, S. Ram, P.A. Rice. “Immunization against a Saccharide Epitope Accelerates Clearance of experimental Gonococcal Infection” *PLoS Pathog* 9 (2013) 8, e1003559.
- 215 H. Cheng, Z. Yang, M. M. Estabrook, C. M. John, G. A. Jarvis, S. McLaughlin, J. McLeod Griffiss. “Human Lipooligosaccharide IGG That Prevents Endemic Meningococcal Disease Recognizes an Internal Lacto-N-neotetraose Structure” *THE JOURNAL OF BIOLOGICAL CHEMISTRY* 286 (2011) 51, 43622-43633.
- 216 S. Gulati, J. Shaughnessy, S. Ram P.A. Rice. “Targeting Lipooligosaccharide (LOS) for a Gonococcal Vaccine” *Front. Immunol.* 10 (2019) 321.
- 217 J. Ngampasutadol, P.A. Rice, M.T. Walsh, S. Gulati. “Characterization of a Peptide Vaccine Candidate Mimicking an Oligosaccharide Epitope of *Neisseria Gonorrhoeae* and Resultant Immune Responses and Function” *Vaccine* 24 (2006) 2, 157–170.
- 218 E.T. O’Connor, K.V. Swanson, H. Cheng, K. Fluss, J.M. Griffiss, D.C. Stein. “Structural Requirements for Monoclonal Antibody 2-1-L8 Recognition of *Neisseria* Lipooligosaccharides” *Hybridoma* 27 (2008) 2, 71–79.
- 219 R. Yamasaki, H. Koshino, S. Kurono, Y. Nishinaka, D.P. McQuillen, A. Kume, S. Gulati, P. A. Rice. “Structural and Immunochemical Characterization of a *Neisseria gonorrhoeae* Epitope Defined by a Monoclonal Antibody 2C7; the Antibody Recognizes a Conserved Epitope on Specific Lipo-oligosaccharides in Spite of the Presence of Human Carbohydrate Epitopes” *THE JOURNAL OF BIOLOGICAL CHEMISTRY* 274 (1999) 51, 36550–36558.

220 S. Ram, S. Gulati, L.A. Lewis, S. Chakraborti, B. Zheng, R.B. DeOliveira, G.W. Reed, A.D. Cox, J. Li, F. St. Michael, J. Stupak, X.-H. Su, S. Saha, C.S. Landig, A. Varki, P.A. Rice. “A Novel Sialylation Site on Neisseria Gonorrhoeae Lipooligosaccharide Links Heptose II Lactose Expression with Pathogenicity” *Infect Immun* 86 (2018) 8.

221 S. Gulati, M.W. Pennington, A. Czerwinski, D. Carter, B. Zheng, N.A. Nowak, R.B. DeOliveira, J. Shaughnessy, G.W. Reed, S. Ram, P.A. Rice. “Preclinical efficacy of a lipooligosaccharide peptide mimic candidate gonococcal vaccine” *mBio* 10 (2019) e02552-19.

RINGRAZIAMENTI

Il dottorato è stata un'esperienza di vita unica ed un percorso che mi ha arricchito sia dal punto di vista scientifico sia personale. Durante questi tre anni di studio/lavoro, convegni, scuole e lunghi periodi fuori casa, ho avuto il piacere di condividere ansie, pianti e anche risate con tante persone che sento di dover ringraziare.

Innanzitutto, coloro che mi hanno dato la possibilità di intraprendere questo percorso e di raggiungere questo importante traguardo: i miei supervisori. Ringrazio Tony, per aver creduto in me sin da quando ero solo una studentessa triennale, spronandomi a dare sempre il massimo ed insegnandomi ad essere determinata ed ambiziosa. Ringrazio Alba, una persona che ho imparato a conoscere col tempo. Al di là dei suoi (fondamentali) insegnamenti scientifici, ho sentito il suo sostegno e la sua vicinanza nei momenti in cui avevo più bisogno, e per me questo conta molto.

Ringrazio Roberta, per la fiducia che ha riposto in me. Abbiamo condiviso giornate intere di lavoro, esaurimenti vari, pianti e risate, con lei non servono tante parole per capirsi... Grazie per tutto!

Ringrazio la Prof.ssa Lombardi, coordinatrice di dottorato e mia relatrice, sempre disponibile e premurosa nei miei riguardi, con la quale ho condiviso fruttuosi e piacevoli incontri per la discussione dei progressi dei miei progetti.

Ringrazio il prof. Marco Fragai del dipartimento di chimica di UniFi ed il Prof. Ondřej Vaněk del dipartimento di biochimica dell'Università Karlova per l'ospitalità e l'accoglienza nei rispettivi gruppi di ricerca di Firenze e Praga, stupende città che porto nel cuore.

Ringrazio tutti i ragazzi del Centro di Risonanze Magnetiche (CERM) di Sesto Fiorentino, persone meravigliose che mi hanno fatto sentire parte del gruppo fin dal primo giorno; in particolare, Domenico e Bora, con i quali ho condiviso la maggior parte del mio tempo, instaurando una vera amicizia.

Grazie a Celeste, il mio punto di riferimento nel laboratorio di Praga, come le dicevo sempre: "My good ghost"!

Ringrazio i miei colleghi del gruppo SSCN, soprattutto chi, con un sorriso o una pausa caffè ha reso migliori le mie giornate... In particolare, Maria, Angela, Molly e Ferran, senza i quali il laboratorio non sarebbe lo stesso; Rosa e Miguel soprattutto per il supporto scientifico.

Ringrazio le amiche che mi sono state vicino durante questo percorso, soprattutto quelle storiche, sulle quali sapere di poter sempre contare, Marica, Alessia e Annarella, perché, sebbene gli impegni personali ci portino a non vederci né sentirci quotidianamente, siamo consapevoli del bene che ci lega.

Ringrazio Rita e Mariagrazia, conosciute proprio durante l'inizio del mio dottorato, per avermi fatto sentire parte della loro famiglia e per essere presenti in questo importante traguardo raggiunto.

Ringrazio la mia stupenda famiglia, mia madre, mio padre, Federica e la nonna Luisa, per credere costantemente in me, sostenere le mie scelte ed essere sempre dalla mia parte. Se sono arrivata fin qui, il merito è dei miei genitori, dei loro sacrifici, della loro fiducia e del loro amore. Per me sono un esempio di vita.

Infine, la persona che sento di dover ringraziare più di tutti è Antonio, sebbene sia difficile esprimere con delle parole quello che ha fatto e continua a fare ogni giorno per rendermi felice. Lo ringrazio per essere stato presente ad ogni mio crollo emotivo ed in ogni piccola e grande difficoltà incontrata, per essermi stato vicino anche quando ci separavano chilometri di distanza, per il tempo speso per me nonostante i suoi impegni personali. Ha sempre rispettato le mie scelte, anche quelle non condivise, ed ha creduto nelle mie potenzialità molto più di quanto non lo facessi io. Sono felice di condividere la mia vita con lui.

Grazie a tutti!
