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

## Department of Chemical Sciences

neapōlis

## Ph.D. School in Chemical Sciences – Cycle XXXVI

# Analysis of the molecular basis behind the recognition mechanisms of prokaryotic glycans by the host immune system

Angela Marseglia

Supervisor

Prof. Roberta Marchetti

Examiner

Prof. Gerardino D’Errico

UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Università e della Ricerca

PON
RICERCA E INNOVAZIONE
2014 - 2020

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”

UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Università
e della Ricerca

PON
RICERCA
E INNOVAZIONE
2014 - 2020

This PhD project is a collaboration between “University of Naples Federico II” and “GlaxoSmithKline Biologicals SA”.

GSK

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **3'SLn** | 3'Sialylactosamine |
| **6'SLn** | 6'Sialylactosamine |
| **AA** | Acetylated Alditols |
| **AMG** | Acetylated Methyl Glycosides |
| **AMR** | Antimicrobial Resistance |
| **APC** | Antigen-Presenting Cell |
| **BCA** | Bicinchoninic Acid |
| **BCR** | B-Cell Receptor |
| **CAP** | Community-Acquired Pneumonia |
| **CCA** | Cholangiocarcinoma |
| **CD** | Cluster of Differentiation |
| **CD** | Cytoplasmic Domain |
| **CDAP** | 1-Cyano-4-Dimethylaminopyridinium Tetrafluoroborate |
| **CF** | Cystic Fibrosis |
| **CMAH** | Cytidine Monophosphate-N-acetylneuraminic Acid Hydroxylase |
| **CORCEMA-ST** | COmplete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer |
| **COSY** | COrrelation SpectroscopY |
| **CPS** | Capsular PolySaccharide |
| **CRD** | Carbohydrate Recognition Domain |
| **$CRM_{197}$** | Cross Reacting Material 197 |
| **CS** | Chondroitin Sulfate |
| **CSM** | Chemical Shift Mapping |
| **CSP** | Chemical Shift Perturbation |
| **CTL** | C-Type Lectin |
| **CTLD** | C-Type Lectin Domain |
| **DAMP** | Danger-Associated Molecular Pattern |
| **DC** | Dendritic Cell |
| **DC-SIGN** | Dendritic Cell-Specific Intercellular Adhesion Molecule-3-Grabbing Non-Integrin |
| **DDM** | N-Dodecyl-β-D-Maltoside |
| **DIC** | Differential Interference Contrast |
| **DNA** | DeoxyriboNucleic Acid |
| **DS** | Dermatan Sulphate |

| | |
|---|---|
| **ECD** | ExtraCellular Domain |
| **EDC** | 1-Ethyl-3-[3-Dimethylaminopropyl]Carbodiimide Hydrochloride |
| **EDTA** | EthyleneDiamineTetraacetic Acid |
| **ELISA** | Enzyme-Linked ImmunoSorbent Assay |
| **EPS** | ExoPolySaccharide |
| **ER** | Endoplasmic Reticulum |
| **FC** | Flow Cell |
| **Fc** | Fragment Crystallizable |
| **Fuc** | Fucose |
| **GAG** | GlycosAminoGlycan |
| **Gal** | Galactose |
| **Gal*f*** | Galactofuranose |
| **GalNAc** | N-Acetylgalactosamine |
| **GBP** | Glycan-Binding Protein |
| **GC-MS** | Gas Chromatography-Mass Spectrometry |
| **GDP-Man** | Guanosine Diphosphate-Mannose |
| **Glc** | Glucose |
| **GlcA** | Glucuronic Acid |
| **GlcN** | Glucosamine |
| **GlcNAc** | N-Acetylglucosamine |
| **GM** | Gut Microbiota |
| **GPI** | Glycosylphosphatidylinositol |
| **GPI-AP** | Glycosylphosphatidylinositol-Anchored Protein |
| **HAI** | Health care-Associated Infection |
| **Hep** | Heptose |
| **HIV** | Human Immunodeficiency Virus |
| **HMBC** | Heteronuclear Multiple Bond Correlation |
| **HS** | Heparin/Heparan Sulphate |
| **HSQC** | Heteronuclear Single Quantum Correlation |
| **IE** | Infective Endocarditis |
| **IgG** | Immunoglobulin G |
| **IL** | Interleukin |
| **IPTG** | Isopropyl β-D-1-Thiogalactopyranoside |
| **Kdo** | 3-Deoxy-D-Manno-Oct-2-Ulopyranosonic Acid |
| **KL** | K Locus |
| **KS** | Keratan Sulphate |
| **LB** | Luria-Bertani |

| | |
|---|---|
| **Leg** | Legionaminic acid |
| **LeY-PAA** | Polyacrylamide-Based Multivalent Lewis Y Antigen |
| **LOS** | LipoOligoSaccharide |
| **LPS** | LipoPolySaccharide |
| **LTA** | LipoTeichoic Acid |
| **MAMP** | Microbe-Associated Molecular Pattern |
| **Man** | Mannose |
| **ManNAc** | N-Acetylmannosamine |
| **MD** | Myeloid Differentiation |
| **MD** | Molecular Dynamics |
| ***Mge*** | *Mycoplasma genitalium* |
| **MGL** | Macrophage Galactose-type Lectin |
| **MHC** | Major Histocompatibility Complex |
| **MM** | Molecular Mechanics |
| ***Mpn*** | *Mycoplasma pneumoniae* |
| **MS** | Mass Spectrometry |
| **MurNAc** | N-Acetylmuramic Acid |
| **MW** | Molecular Weight |
| **ND** | Neck Domain |
| **Neu5Ac** | N-Acetylneuraminic acid |
| **Neu5Gc** | N-Glycolylneuraminic acid |
| **NHS** | N-Hydroxy-Succinimide |
| **NLR** | Nucleotide-Binding Oligomerization Domain-Like Receptor |
| **NMR** | Nuclear Magnetic Resonance |
| **NOESY** | Nuclear Overhauser Effect SpectroscopY |
| **NuPAGE** | SDS-PolyAcrylamide Gel Electrophoresis at neutral pH |
| **OD** | Optical Density |
| **OM** | Outer Membrane |
| **OS** | Core Oligosaccharide |
| **OSR1** | OligoSaccharide type R1 |
| **PBS** | Phosphate Buffered Saline |
| **PCP** | Phenol-Chloroform-Petroleum ether |
| **PDB** | Protein Data Bank |
| **PG** | PeptidoGlycan |
| **PID** | Pelvic Inflammatory Disease |
| **PMAA** | Partially Methylated and Acetylated Alditols |
| **PRR** | Pathogen Recognition Receptor |

| | |
|---|---|
| **QM** | Quantum Mechanics |
| **R-form LPS** | Rough-form LipoPolySaccharide |
| **Rha** | Rhamnose |
| **RMSD** | Root Mean Square Deviation |
| **RMSF** | Root Mean Square Fluctuation |
| **RNA** | RiboNucleic Acid |
| **ROESY** | Rotating frame Overhauser Effect SpectroscopY |
| **RU** | Resonance Unit |
| **RU** | Repeating Unit |
| **SAMP** | Self-Associated Molecular Pattern |
| **SARS-CoV-2** | Severe Acute Respiratory Syndrome CoronaVirus 2 |
| **SCA** | Semicarbazide |
| **SDS-PAGE** | Sodium Dodecyl Sulphate-PolyAcrylamide Electrophoresis Gel |
| **SEC** | Size Exclusion Chromatography |
| **SEC-HPLC** | Size Exclusion Chromatography-High Performance Liquid Chromatography |
| **S-form LPS** | Smooth-form LipoPolySaccharide |
| **Sia** | Sialic acid |
| **SIGLEC** | Sialic acid-binding Immunoglobulin-type Lectin |
| **SLBR** | Siglec-Like Binding Region |
| **SNFG** | Symbol Nomenclature For Glycans |
| **SPR** | Surface Plasmon Resonance |
| **sTa-Thr** | Sialyl-T-Antigen |
| **STD NMR** | Saturation Transfer Difference NMR |
| **STORM** | STochastic Optical Reconstruction Microscopy |
| **TCR** | T-Cell Receptor |
| **TFA** | Trifluoroacetic Acid |
| **TLR** | Toll-Like Receptor |
| **TMD** | Transmembrane Domain |
| **TNF** | Tumor Necrosis Factor |
| **TOCSY** | TOtal Correlation SpectroscopY |
| **tr-NOESY** | transferred-NOESY |
| **TROSY** | Transverse Relaxation Optimized SpectroscopY |
| **UC** | UltraCentrifugation |
| **UDP-GlcNAc** | Uridine-Diphosphate-N-Acetylglucosamine |
| **WTA** | Wall Teichoic Acid |

# SECTION I – INTRODUCTION

# Chapter 1 – Molecular host-guest interactions in living systems

Host-guest interactions are fundamental molecular relationships that underpin many essential processes within living organisms, especially at eukaryotes-bacteria interface. Bacteria constitute an essential part of our ecosystem, contributing to the balance and sustainability of the environment; some of these microorganisms are harmless or even beneficial to their host, but some others can pose a significant threat to the global health by causing infections and diseases. Not by chance, the interactions between microbial guest and human host can be both pathogenic and symbiotic, including, as examples, the recognition of microbial pathogens by the host immune system[1] or the beneficial interaction between humans and bacterial members of the gut microbiota, a collection of microorganisms which colonizes the human gastrointestinal tract, developing a mutually beneficial relationship with the host.[2]

The interactions of invading microbes with their hosts are influenced, to an important degree, by the pattern of glycoconjugates and glycan-binding proteins (GBPs) that each express. On the microbe side, complex glycans, exposed on bacterial cells, are known to be Microbe Associated Molecular Patterns (MAMPs), with key roles in the regulation of the host immune response, mediating, as example, all stages of infections. On the host side, specific receptors, known as Pathogen Recognition Receptors (PRRs), are designated to detect potentially harmful pathogens through the recognition of peculiar cell wall motifs, initiating an inflammatory reaction.

Understanding, at a molecular level, the basis of the complex interplay between glycoconjugates and their glycan receptors has profound implications for the design and development of novel therapeutics and diagnostics.

Despite significant advancements in medicine, bacterial infections and their secondary effects remain one of the leading causes of human mortality. Notably, a concerning and persistent increase in antibiotic resistance across many bacterial species represents one of the key contemporary challenges to public health. Among these antimicrobial resistance species, ESKAPE have recently attracted attention. ESKAPE is the acronym of six classes of Gram-positive and Gram-negative bacterial pathogens: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacter* species. These pathogens are common causes of severe, life-threatening hospital-acquired infections, particularly in critically ill and immunocompromised patients.[3] In this situation, the analysis of bacterial glycoconjugates for the development of new antimicrobial strategies become necessary.[4]

In this context, this thesis aims to unravel the molecular processes governing host-guest interactions. The focal point lies in the comprehensive analysis of molecular recognition events involving bacterial glycans such as capsular polysaccharides (CPS), lipopolysaccharides (LPS) and lipooligosaccharides (LOS) from different Gram-negative bacteria and their interaction with human proteins, with a specific emphasis on the C-type lectin DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin), but, also the interaction of eukaryotic N- and O-glycans with bacterial adhesins, as those isolated from *Mycoplasma* species. By shedding light on these molecular interactions, a deeper understanding of different physiological and pathological processes will be provided.

## 1.1 Glycans and their relevance in biological functions

Glycans, also known as carbohydrates or sugars, represent one of the four fundamental classes of biomolecules in living systems, alongside nucleic acids, proteins, and lipids. These complex molecules exhibit considerable structural diversity and possess an information content several times higher than other biological biomolecules. Although historically, the field of glycans has been underexplored, compared to lipids, proteins, and nucleic acids, glycoscience field has experienced rapid growth in recent years.[5]

Glycans cover the surfaces of all cell types, from fungi to mammals. In prokaryotes, plants or fungi, glycans conjugated with proteins or lipids constitute the cell wall, while in humans and animals, they form the glycocalyx, a dense and complex array of glycoconjugates with variable composition and structure[6] that represents the first contact point in any kind of host-guest interaction, either symbiotic or pathogenic.

Glycan's biological functions can be divided in three categories:

- structural roles in and outside the cells;
- energy metabolism, including nutrient storage;
- information carriers as molecular patterns recognized by endogenous or exogenous GBPs.[7] As example, bacterial glycan-based signatures, the so called MAMPs, can be recognized by the eukaryotic host through innate immune receptors, known as PRRs. These PRRs contribute to initial recognition of bacterial glycans, thus providing an early defense mechanism against bacterial infections.(Figure 1.1)[8]

Apart from their involvement in controlling immune responses, glycans are directly involved in the etiology of many diseases, including bacterial and viral infections, cancer and autoimmune diseases.[9]

Considering the growing evidence of the crucial roles played by carbohydrates in numerous biological events, such as cell differentiation, homing to specific tissues, cell adhesion, cell recognition, microbial pathogenesis and immunological recognition, there has been an increasing attention on the study of the structure, conformation, and dynamics of glycans, in the last few decades. Despite the complexity, great efforts have been made to decipher the so-called "glycome", thus, paving the way for the design of sugar-based drugs to be employed not only as therapeutics but, also, as prevention, in the field of vaccines.

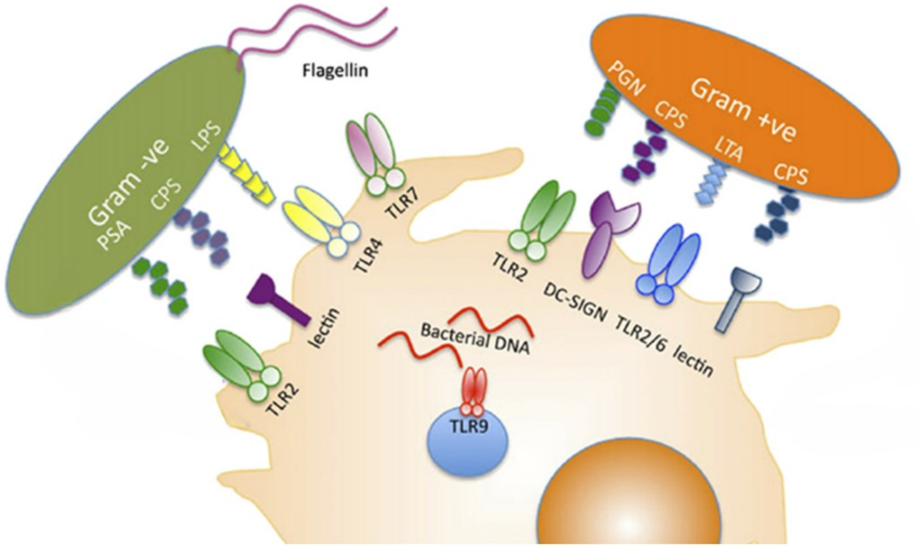


**Figure 1.1:** Illustration of host interaction with Gram-positive and Gram-negative bacteria.

The intricate nature of glycans arises from the numerous possibilities in which monosaccharides, the building units of glycans, can be linked to each other. They can assume linear or branched structures, with glycosidic linkages in either α and β configuration.[10] Furthermore, glycans may also feature substitutions with functional groups like acetyl, phosphate and sulfate, or can be found attached to lipids (glycolipids), peptides (glycopeptides) and proteins (glycoproteins), due to different glycosylation patterns, depending on the specific bond between protein and glycan (Figure 1.2).[11-13] Thus, even a limited number of sugar monomers can create thousands of complex glycans.

This variety allows cells to encode extensive information within these structures. However, it also presents a significant challenge in investigating their structure, conformation, and understanding the role that glycosylation plays in biological phenomena.

## 1.2 Eukaryotic glycoconjugates

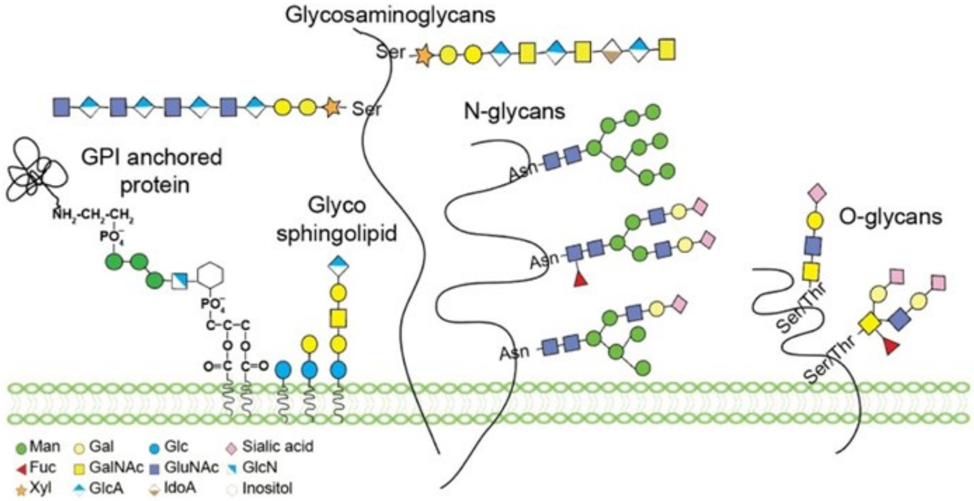


**Figure 1.2:** Schematic representation of eukaryotic glycans, using SNFG symbols for monosaccharides.

As mentioned above, complex glycans, covalently linked to lipids or proteins, represent the most abundant class of biomolecules exposed on membrane surfaces of all cells.

The *glycosphingolipids* are ubiquitous membrane components, composed of a glycan portion linked to a ceramide, which is composed of sphingosine and fatty acids, and possess crucial roles in signal transduction events, cell-cell interactions, and cell adhesion. Gangliosides are an example of glycosphingolipids; these molecules are constituted at least by three sugars, containing a sialic acid moiety, and are mostly abundant in the brain (Figure 1.3).[14]

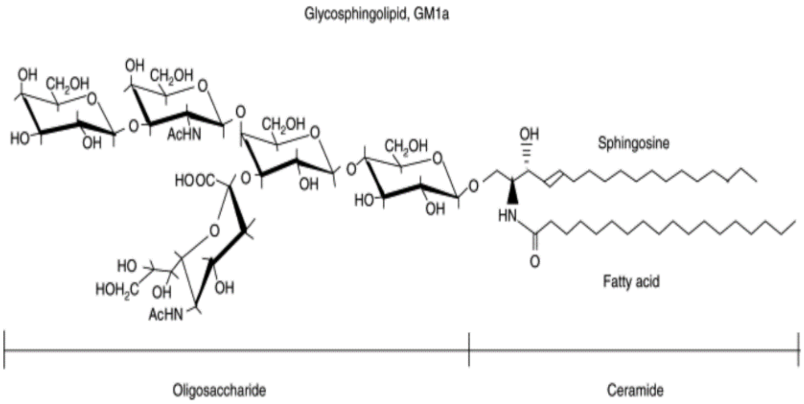


**Figure 1.3:** Chemical structure of the ganglioside GM1a, as example of glycosphingolipid: the oligosaccharide moiety is attached to the ceramide containing sphingosine and fatty acids.

Another class of glycolipids is composed by the *glycoglycerolipids*, characterized by an alkaline-labile glycerol backbone. Glyceroglycolipids are not widely distributed in vertebrates, but 3-sulfated galactosyldiacylglycerol (seminolipid) is the most abundant glycolipid in the testis and sperm (Figure 1.4).

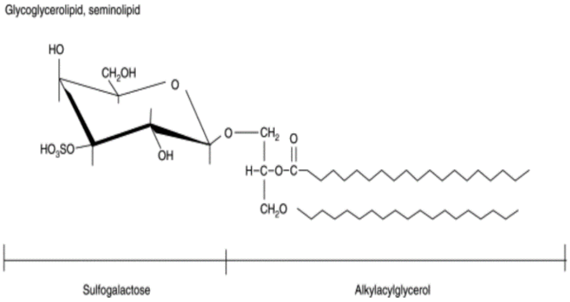


**Figure 1.4:** Chemical structure of the seminolipid in which the sulfogalactose is attached to the alkylacylglycerol.

Moreover, in many eukaryotes, a small functionally different group of proteins is attached to the cell surface by a *glycosylphosphatidylinositol (GPI) anchor*. GPI-anchored proteins (GPI-APs) share a common core structure consisting of ethanolamine-$PO_4$-6Manα1-4GlcNα1-6*myo*-inositol-1-P-lipid with variable additional glycans, fatty acid esters, and phosphoethanolamine groups (Figure 1.5). GPI anchors impart to their attached proteins the ability to be shed from the cell surface in soluble form through the action of cellular or serum GPI-cleaving enzymes.[15]

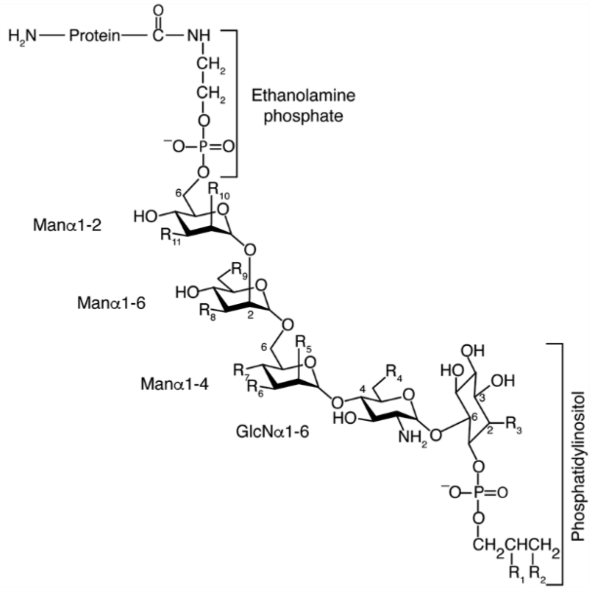


**Figure 1.5:** General structure of GPI anchors attached to proteins. Heterogeneity in GPI anchors is derived from various substitutions of this core structure, represented here as R groups. ($R_1$=fatty acid or OH; $R_2$=fatty acid or alkyl or alkene chain, in some cases the lipid may also be a ceramide rather than a glycerolipid; $R_3$=OH or fatty acid; $R_4$= OH or aminoethylphosphonate in *T. cruzi*; $R_5$,$R_9$= ethanolamine phosphate or OH; $R_6$, $R_7$, $R_8$, $R_{10}$, $R_{11}$= carbohydrate substituents or OH) (adapted from [15]).

As mentioned earlier, many proteins on the external surface of mammalian cells contain carbohydrates. Indeed, protein glycosylation is the most prevalent post-translational modification in mammals. This process begins in the endoplasmic reticulum (ER) and Golgi apparatus, while the final processing occurs in the cis-, medial-, and trans- Golgi compartments.[16] Three different types of glycoproteins come up containing *N-glycans*, *O-glycans* and *proteoglycans*. In general, glycosylation begins with a lipid dolichol pyrophosphate on the ER membrane, which serves as starting point for co-translational modification to form the oligosaccharide chain. In the N-glycosylation process, the initial phase involves anchoring two

GlcNAc and five Man residues to the dolichol lipid, derived from specific nucleotide sugars, UDP-GlcNAc and GDP-Man, on the cytoplasmic side of the ER membrane. This dolichol-based lipid with a $GlcNAc_2Man_5$ oligosaccharide is then flipped to the luminal side of the ER membrane. The oligosaccharide chain continues to grow in the ER lumen, until a chain of $Glc_3Man_9GlcNAc_2$ is formed. This 14-sugar residue glycan is, then, transferred to an Asn within the ER lumen, based on a consensus sequence Asn-X-Ser/Thr (X can be any aa except Pro), by an oligosaccharyl-transferase. In the trimming stage, the N-glycan's core region undergoes modifications by removing three Glc residues and one Man from the chain, facilitating the glycoprotein's translocation to the Golgi apparatus. In the Golgi region, the glycoprotein undergoes further modifications through specific glycosidases and glycosyltransferases, depending on the biological functions it will perform. This confers heterogeneity to the N-glycan structures, classified into three main groups: *oligomannose-, complex-* and *hybrid-type N-glycans*. All share a common chitobiose core (Figure 1.6), consisting of three Man and two GlcNAc, with one of the GlcNAc residues covalently linked to the Asn.[17] In particular, oligomannose-type N-glycans, possess many Man residues in their branching structures and they have been recently associated with tumor progression in cholangiocarcinoma (CCA) and bile duct cancer.[18] Complex-type N-glycans can have up to five branches, initiated by GlcNAc, and are highly located in kidneys and brain.[19] Finally, hybrid N-glycans, with roles in post-translational modification of neuronal glycoproteins and embryogenesis,[20] possess in one arm Man1-6 exclusively linked to Man residues, in the other arm containing Man1-3, the organization is similar to the complex-type N-glycans.

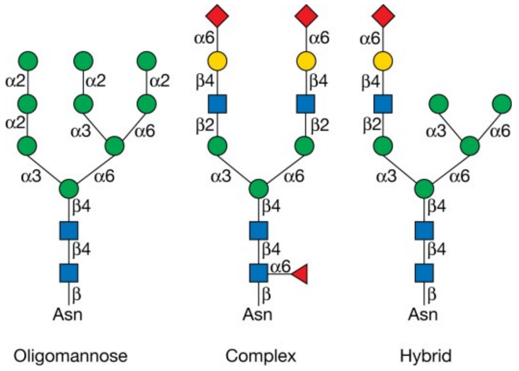


**Figure 1.6:** SNFG representation of the different types of N-Glycans with a common chitobiose core structure (extracted from ref.[17]).

Unlike N-glycans, O-glycans biosynthesis does not require a precursor to transfer the oligosaccharide chain to the protein. Indeed, as post-translational modification occurring in Golgi apparatus, various enzymes sequentially attach single sugars to the protein, synthetized in the ER, without a consensus sequence, starting from the first GalNAc monosaccharide linked to a Ser or Thr residue. Among O-glycans, *mucins* are considered the primary class (Figure 1.7). They represent the major constituents of mucus, located on the epithelial cells of gastrointestinal, genitourinary, and respiratory tracts, providing hydration and protection against adverse environmental factors. However, conditions such as cancer, inflammatory bowel disease, congenital disorders of glycosylation, or hypersecretory bronchial and lung diseases can alter their properties. Additionally, sialylated O-glycans are interesting due to their role in interacting with bacterial adhesins to colonize and infect host cells.[21]

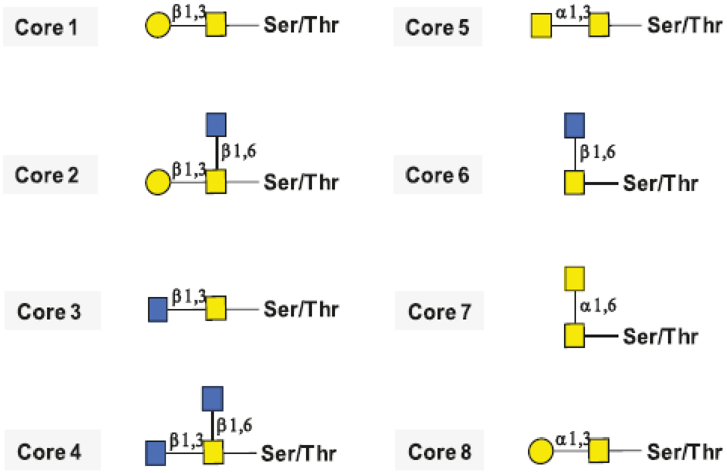


**Figure 1.7:** Core structures of mucin-type glycans. In the O-glycosylation, as post-transductional modification, various enzymes connect the first GalNAc to Ser or Thr residues, with no consensus sequence (extracted from [22]).

Finally, proteoglycans are glycoproteins that incorporate chains of O-linked glycosaminoglycans (GAGs), linked to specific Ser residues of proteins through a tetrasaccharide structure. GAGs are long polysaccharides containing disaccharides of GlcNAc or GalNAc along with Gal or uronic acids like GlcA or its epimerized form iduronic acid. GAGs exhibit functional diversity and are classified into three categories based on the specific disaccharide structure: (i) *dermatan sulfate* (DS) and *chondroitin sulfate* (CS), (ii) *heparin/heparan sulfate* (HS) and (iii) *keratan sulfate* (KS). Interesting to underline that KS is the one that can be linked to N- or O-glycans.[23]

### 1.2.1 Sialic acids

Sialic acids (Sias) are a family of nine-carbon keto monosaccharides, generally present in vertebrates and usually found as terminal residues of the glycan chains of glycoproteins and glycolipids.[24]

The nomenclature of these molecules comes out from their first isolations in salivary mucins by G. Blix in 1936 (“sialic acid” from the Greek word “saliva”) and in brain glycolipids by E. Klenk in 1941 (“neuraminic acid” from “neurons”).[25]

Over 60 different Sias have been identified, which differ for structural modification of the sugar backbone. However, all nonulosonic acids are peculiar due to the presence of many molecular components in one monosaccharide like a carboxylic acid, an exocyclic side containing 3-carbon atoms, and, often, one or more acylated amino groups. The anomeric C-2 position is generally α-(2,3)- or α-(2,6)-linked to a hydroxyl group of a Gal or GalNAc residue. While, the other positions can be subjected to modifications, like sulfation, methylation or phosphorylation, generating further complexity in this monosaccharides' family (Figure 1.8).[25]

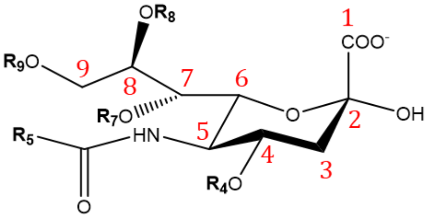
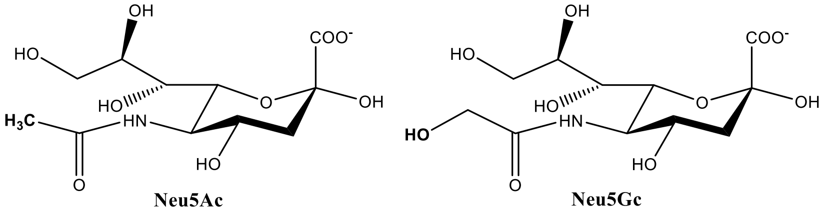


**Figure 1.8:** Sugar backbone of Sias. Some variations can occur as indicated: at physiological pH, the carboxylate gives the negative charge of Sia; R4 = H; -acetyl; Fuc; Gal. R5 = Amino; *N*-acetyl; *N*-glycolyl; hydroxyl; *N*-glycolyl-*O*-acetyl; *N*-glycolyl-*O*-methyl. R7 = H; -acetyl. R8 = H; -acetyl; -methyl; -sulfate; Sia; Glc. R9 = -H; -acetyl; -lactyl; -phosphate; -sulfate; Sia.

In mammalian cells, the predominant sialic acid is the N-acetylneuraminic acid (Neu5Ac) (Figure 1.9). It serves as a ligand for animal and bacterial lectins, antibodies and enzymes, playing a crucial role in mediating bacterial and viral infections, tumor growth, metastasis, and contributing to various biological processes, including immunology, microbiome regulation, cell signaling, reproduction and the biology of the nervous system. [26] A modification of the 5-N-acetyl group of Neu5Ac, resulting in the addition of a hydroxyl group, leads to the formation of N-glycolylneuraminic acid (Neu5Gc), the prevalent sialic acid in chimpanzee but absent in humans, due to a mutation in the CMAH (Cytidine monophospho-N-acetylneuraminic acid hydroxylase) gene, responsible for its synthesis (Figure 1.9).[27] Despite its absence in humans, Neu5Gc can be obtained from the diet, especially through the consumption of red meat. Its accumulation in tissues may lead to inflammation and contribute to the progression of cancer.[28]



**Neu5Ac** **Neu5Gc**

**Figure 1.9:** Chemical structure of Neu5Ac and Neu5Gc, predominant Sias in mammals.

Due to their widespread distribution and structural diversity, Sias can modulate a diverse array of physiological and pathological processes. Indeed, Sias on the surfaces of mammalian cells are considered potent markers of "*self*" and, upon recognition of specific GBPs (cf. §1.5), play pivotal roles in regulating the immune system. [29, 30]

Worth noting, several pathogens have developed the capability to capture Sias from the eukaryotic host, in order to incorporate them into their glycoconjugates, or to synthetize *de novo* monosaccharides structurally related to Sias (Figure 1.10)[31]. Indeed, covering their surfaces with this type of sugars, many pathogenic bacteria can mimic self-associated molecular patterns (SAMPs), eluding host immune response, often exploiting the action of inhibitory host immune proteins.

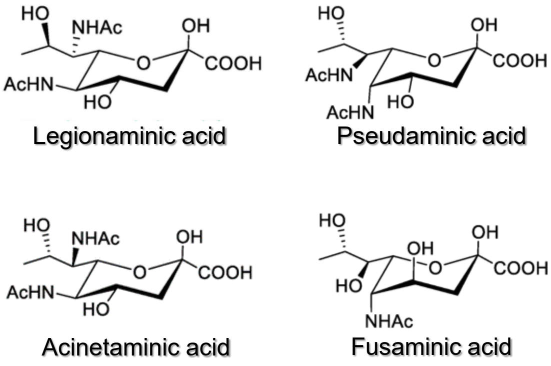


**Figure 1.10:**Structures of different nonulosonic acids, structurally related to Sias.

Delving into the intricate world of glycoconjugates extends beyond the exploration of eukaryotic systems, encompassing a critical examination of bacterial glycoconjugates. This broader perspective is imperative due to the pivotal roles these structures play in different physiological and pathological processes. The comprehensive study of both eukaryotic and bacterial glycoconjugates is essential for unravelling the complexities of host-guest interactions, immune responses, and disease mechanisms. Therefore, in the following paragraphs, a detailed explanation of bacterial glycoconjugates will be provided to fully grasp the spectrum of glycoscience and its implications for health and disease.

## 1.3 Bacterial glycoconjugates

The characteristics of bacterial cell walls and their glycoconjugates can vary within the same bacterial species and, of course, between different classes of bacteria; for example, different glycoconjugates are exposed on the surface of Gram-positive and Gram-negative bacteria (Figure 1.11). This bacterial classification is based on a procedure called Gram staining, firstly developed by H. C. Gram in 1882, and permits to designate two different classes of bacteria: [32]

- **Gram-positive bacteria**: cell wall as a single, relatively thick and uniform layer, containing peptidoglycan (more than 60% of the total cell wall), polysaccharides and/or teichoic acids (WTAs), or teichuronic acid, as major components;
- **Gram-negative bacteria**: thinner cell wall distinctly layered, with an outer membrane (OM), mainly composed of lipopolysaccharides (LPS), capsular polysaccharides (CPS), phospholipids, proteins, lipoproteins and peptidoglycan (less than 10% of the total cell wall).[33]

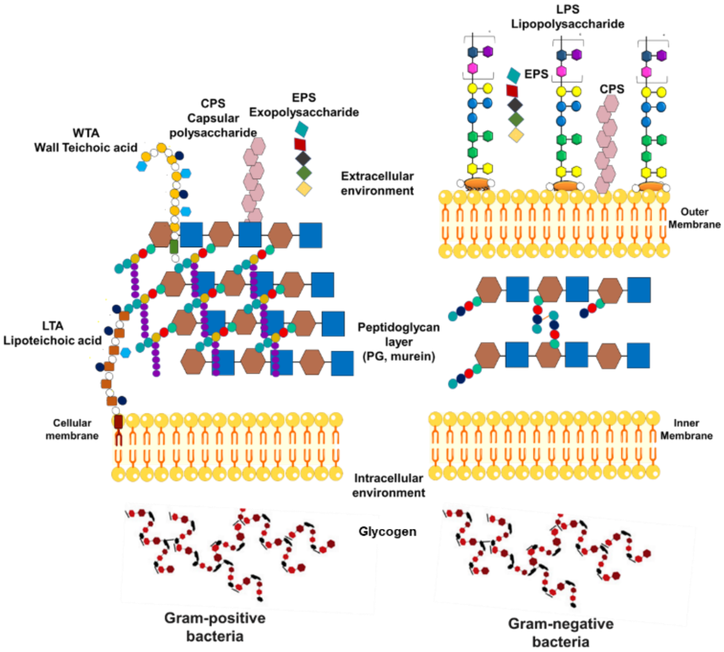


**Figure 1.11:** Schematic representation of Gram-positive and Gram-negative bacteria cell walls. Gram-negative bacteria are characteristic for having two different cell membranes, the inner and the outer membrane, both separated by a thin PG layer. The OM can display LPS or LOS, anchored to the membrane through the lipid A moiety, CPSs and/or EPS. Instead, Gram-positive bacteria have only one cell membrane, which is covered by a thick PG layer and that is usually decorated with TA anchored to the membrane, known as lipoteichoic acids (LTAs) or covalently bound to the PG layer called wall teichoic acids (WTAs). In addition, also CPS and EPS can decorate the extracellular environment as happens with Gram-negative bacteria.

A brief description of the main bacterial cell wall glycoconjugates, including (i) *cell-wall associated compounds* like peptidoglycan (PG) or teichoic acids (WTAs), (ii) *secreted exopolysaccharides* (EPS), and (iii) *surface-located compounds* like capsular polysaccharides (CPS), lipopolysaccharides (LPS) and lipooligosaccharides (LOS), is reported below:[34, 35]

- *PG*

PG or murein is an heteropolymer composed of alternating ß-1,4-linked residues of GlcNAc and MurNAc with a peptide portion attached to MurNAc and containing a mixture of both L- and D- amino acids. This intricate structure provides strength and stability to bacterial cell walls.[36] PG is recognized by the host as a marker of infection. Indeed, as MAMP, it is recognized by TLR2 and other PRRs, like NLRs, triggering the innate immune response.[37]

- *WTAs*

The bacterial cell wall is enforced and functionalized by different carbohydrate-based polymers of diverse chemical composition. Characteristically found in the Gram-positive bacterial cell wall are the teichoic acids (TAs), zwitterionic glycopolymers containing phosphodiester-linked polyol repeat units, crucial for bacterial survival and interaction with the environment. Among them, the wall teichoic acids (WTAs), are synthetized on a disaccharide bactoprenyl phosphate carrier on the inner leaflet of the cell membrane by the action of specific transferases and eventually transported to the outer leaflet where they are cross-linked to PG.[38] The common structure of WTAs is composed of (1→5)-D-ribitol-5-phosphate (RboP) or (1→3)-L-α-glycerol-phosphate (GroP) repeats that are linked to a common disaccharide linker, that is highly conserved across bacterial species and that is constituted of ManNAc-(β-1→4)-GlcNAc-1-phosphate. In addition, the RboP WTAs backbone can be decorated with monosaccharides and/or with cationic D-alanine esters, which are added on the intracellular and extracellular side of the membrane, respectively. In *Staphylococcus aureus*, specific glycosyltransferases, TarM, TarS and TarP, attach a GlcNAc residue to the C3 or C4 hydroxyl groups of RboP residues, either in an α- and β- configuration. Thus, each glycosyltransferase modifies WTAs differently, resulting in distinct WTAs glycosylation patterns that impact immune recognition and clearance of Gram-positive bacteria.[39] D-Alanine modifications modulate the polymers' overall charge and play an important role in the defence mechanisms of the bacterium, for example, in the protection against cationic antimicrobial peptides,[40] and can also have major implications for WTAs uptake, processing and presentation by immune cells.[41] Importantly, the D-Alanine residues are introduced by an as-yet unidentified enzyme. WTAs molecules can extend through and beyond the bacterial cell surface, allowing direct interactions with the environment.

- *EPS*

EPS are gel-like substances rich in high-molecular-mass polysaccharides, loosely attached to cell surfaces, and often released in substantial quantities into the surrounding environment. Regarding their chemical structures, some are neutral macromolecules, the majority exhibits polyanionic characteristics due to the presence of amino sugars, uronic acids or ketal-linked pyruvate. These polyanionic polymers, featuring abundant acetyl groups, demonstrate an ability to bind various metal cations such as $Fe^{2+}$, $Zn^{2+}$, $Cu^{2+}$ and $Co^{2+}$, indicating a potential role in concentrating metal ions in the microenvironment around the cell. EPS play a crucial role in the formation of the biofilm matrix, contributing to biochemical interactions between bacteria and surrounding cells, providing resistance to chemical and physical treatments.[42]

Finally, surface located glycoconjugates such as CPS, LPS and LOS serve as species-specific MAMPs. The remarkable diversity in their structures renders them ideal ligands for establishing specific interactions, whether symbiotic or pathogenic, with the host. A more detailed description of these glycoconjugates will be presented in the next subsection, as they are the focal points of the present thesis.

- *CPS*

CPS are peculiar glycoconjugates, located on the surface of Gram-positive and Gram-negative bacteria. They are high molecular weight polysaccharides, attached to the OM by lipids covalently linked to their reducing end, and, then, inserted in the lipid bilayer of the membrane.[43] As example, the structure of the conserved reducing terminal glycolipid found in the CPS of *E. coli* and *N. meningitidis* is reported in Figure 1.12.

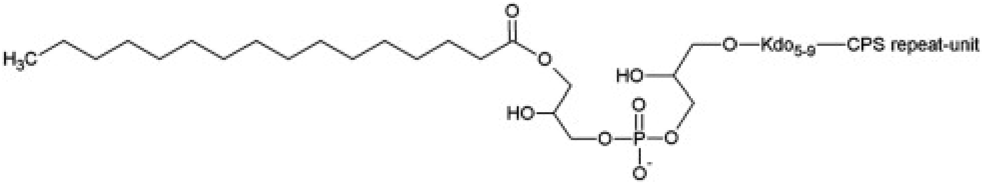


**Figure 1.12:** Structure of the conserved reducing terminal glycolipid found in CPS from *E. coli* and *N. meningitidis*. Some variation in the acyl chain can occur at either position 1 or 2 of the glycerol backbone (the phosphate is position 3), depending on the isolate (adapted from ref.[43]).

The structure of CPS consists of negatively charged and highly hydrophilic long-chain polysaccharides, with linear or branched configuration,[44] viewed by microscopy as an extensive layer surrounding the cell (Figure 1.13).[45]

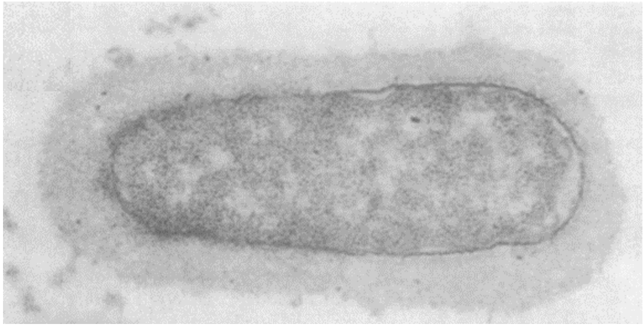
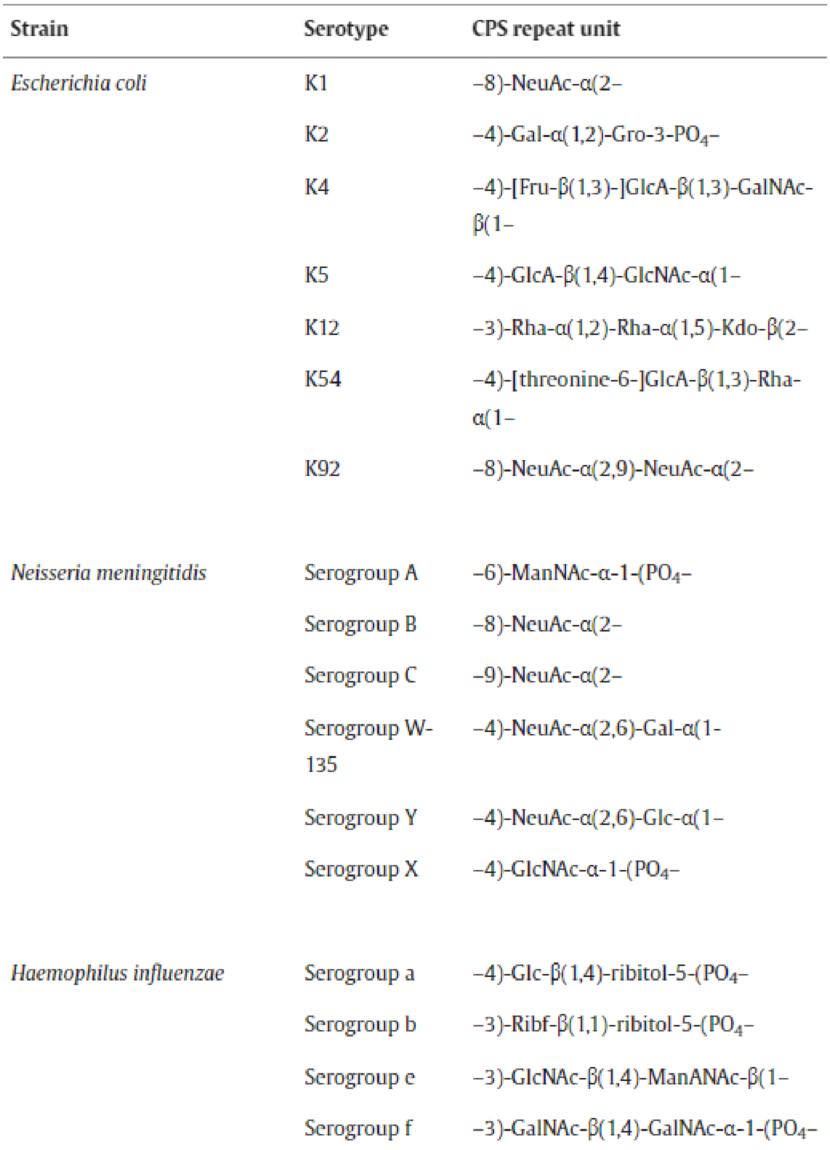


**Figure 1.13:** Electron micrograph of ultrathin section from *E.coli* 20025 (04:K12) (from ref.[45]).

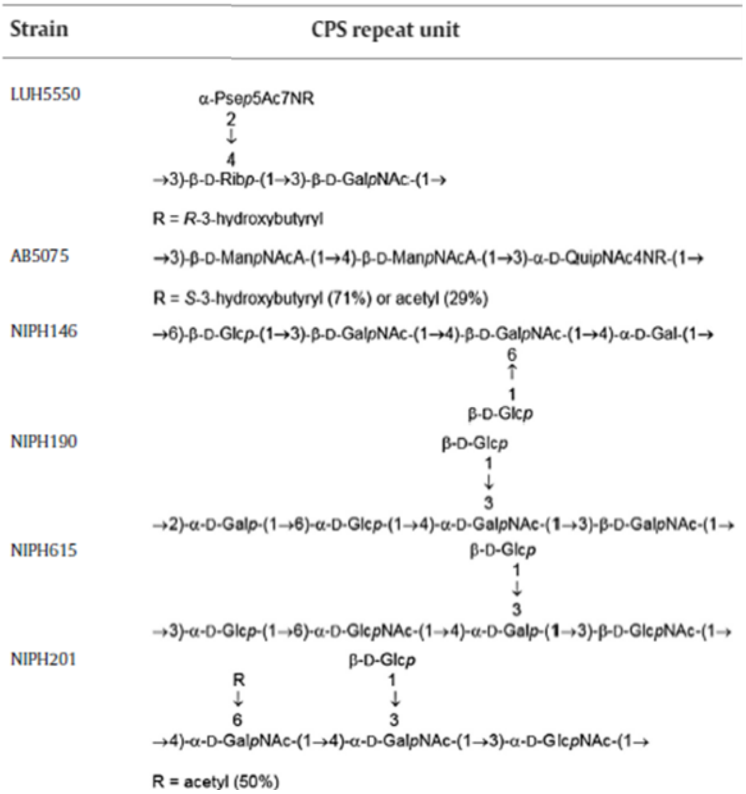
In one bacterial species it is possible to have strains with structurally different CPS, resulting in different serogroups or serotypes; in Table 1 different capsular serotypes and serogroups present in different strains of *E. coli*, *N. meningitidis* and *Haemophilus influenzae* are reported as examples,[43] in Table 2 examples of different structures of CPS repeating units from *A. baumannii*.

**Table 1:** Examples of CPS repeating units from encapsulated pathogenic bacteria (adapted from [43]).



| Strain | Serotype | CPS repeat unit |
|---|---|---|
| *Escherichia coli* | K1 | –8)-NeuAc-α(2– |
| | K2 | –4)-Gal-α(1,2)-Gro-3-$PO_4$– |
| | K4 | –4)-[Fru-β(1,3)-]GlcA-β(1,3)-GalNAc-β(1– |
| | K5 | –4)-GlcA-β(1,4)-GlcNAc-α(1– |
| | K12 | –3)-Rha-α(1,2)-Rha-α(1,5)-Kdo-β(2– |
| | K54 | –4)-[threonine-6-]GlcA-β(1,3)-Rha-α(1– |
| | K92 | –8)-NeuAc-α(2,9)-NeuAc-α(2– |
| *Neisseria meningitidis* | Serogroup A | –6)-ManNAc-α-1-($PO_4$– |
| | Serogroup B | –8)-NeuAc-α(2– |
| | Serogroup C | –9)-NeuAc-α(2– |
| | Serogroup W-135 | –4)-NeuAc-α(2,6)-Gal-α(1- |
| | Serogroup Y | –4)-NeuAc-α(2,6)-Glc-α(1– |
| | Serogroup X | –4)-GlcNAc-α-1-($PO_4$– |
| *Haemophilus influenzae* | Serogroup a | –4)-Glc-β(1,4)-ribitol-5-($PO_4$– |
| | Serogroup b | –3)-Ribf-β(1,1)-ribitol-5-($PO_4$– |
| | Serogroup e | –3)-GlcNAc-β(1,4)-ManANAc-β(1– |
| | Serogroup f | –3)-GalNAc-β(1,4)-GalNAc-α-1-($PO_4$– |

**Table 2:** Examples of CPS repeating units from *A. baumannii* (adapted from [46]).



| Strain | CPS repeat unit |
|---|---|
| LUH5550 | α-Psep5Ac7NR<br>2<br>↓<br>4<br>→3)-β-D-Rib*p*-(1→3)-β-D-Gal*p*NAc-(1→<br>R = *R*-3-hydroxybutyryl |
| AB5075 | →3)-β-D-Man*p*NAcA-(1→4)-β-D-Man*p*NAcA-(1→3)-α-D-Qui*p*NAc4NR-(1→<br>R = *S*-3-hydroxybutyryl (71%) or acetyl (29%) |
| NIPH146 | →6)-β-D-Glc*p*-(1→3)-β-D-Gal*p*NAc-(1→4)-β-D-Gal*p*NAc-(1→4)-α-D-Gal-(1→<br>6<br>↑<br>1<br>β-D-Glc*p* |
| NIPH190 | β-D-Glc*p*<br>1<br>↓<br>3<br>→2)-α-D-Gal*p*-(1→6)-α-D-Glc*p*-(1→4)-α-D-Gal*p*NAc-(1→3)-β-D-Gal*p*NAc-(1→ |
| NIPH615 | β-D-Glc*p*<br>1<br>↓<br>3<br>→3)-α-D-Glc*p*-(1→6)-α-D-Glc*p*NAc-(1→4)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→ |
| NIPH201 | β-D-Glc*p*<br>R 1<br>↓ ↓<br>6 3<br>→4)-α-D-Gal*p*NAc-(1→4)-α-D-Gal*p*NAc-(1→3)-α-D-Glc*p*NAc-(1→<br>R = acetyl (50%) |

CPS are major virulence factors and protect bacteria from different environmental pressures, such as the immune system during the host infection, preventing the activation of the alternate complement pathway. Interesting to underline that encapsulated bacteria are not only pathogens, but also *non*-pathogenic and commensal ones. Indeed, due to their hydrophilic nature, capsules can also protect them from desiccation or can block the infection from bacteriophages.[44]

In the early studies conducted between 1923 and 1929, Avery and Heidelberger at the Rockefeller Institute discovered the immunogenicity of the CPS of *pneumococcus*.[47] Today, this knowledge is harnessed for the development of glycoconjugate vaccines. Various strategies are employed, including *covalent* linkages between activated CPS conjugated, randomly or not, to carrier proteins, to stimulate a T-cell-dependent immune response against the glycan moiety.[48] Additionally, *non-covalent* methods are also employed for vaccine development, such as CPS entrapped in cross-linked proteins[49] or liposomal encapsulation of CPS and proteins.[50] The use of glycoconjugate vaccines could reduce the use of antibiotics, especially in the current era, in which antibiotic-resistance is steadily increasing; not only, it could give protection by herd immunity also to not vaccinated people.[51]

- *LPS* and *LOS*

LPS is one of the main glycoconjugates present on the OM external leaflet of Gram-negative bacteria. Historically, it has been considered a menace for human health because, as bacterial endotoxin, it could lead to dramatic outcomes, like septic shock. More recently, some evidence have shown that LPS can also have positive implications; for example, it is necessary for the development and maturation of host immune system and for the depiction of dangerous microorganisms.[52]

The morphology of Gram-negative bacterial colonies can be smooth or rough, depending on the type of LPS coating their OM. LPS in the smooth form (S-form LPS) is composed by three regions structurally and functionally distinct: (i) a conserved glycolipid region called *Lipid A*, (ii) a short core oligosaccharide region (*core OS region*) and (iii) a surface-exposed O-polysaccharide, dubbed as *O-antigen,* that is the most variable region. When the O-antigen is absent, the LPS is in the rough form (R-form LPS) and can be also named LOS (Figure 1.14).[53]

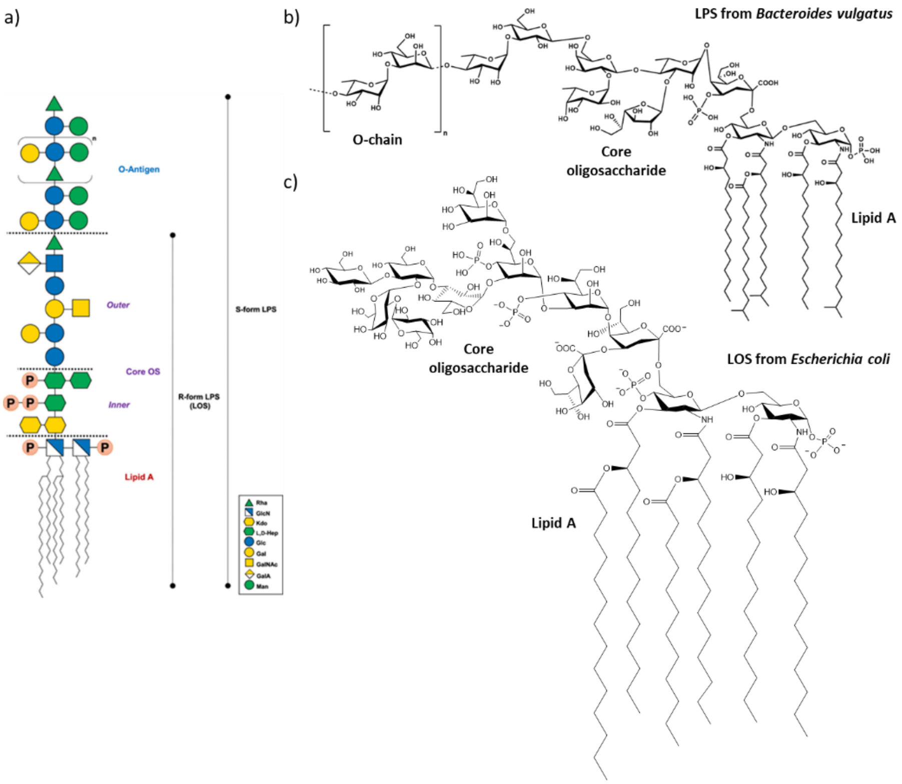
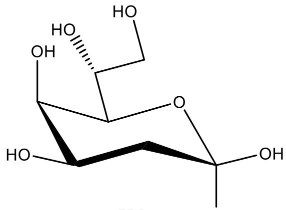


**Figure 1.14:** a) Cartoon representation of the general chemical structure of an LPS (extracted from ref.[52]). b) Chemical structure of LPS from *B. vulgatus*, reported as an example. c) Chemical structure of LOS from *E. coli*, reported as example.

The lipid A is responsible for the endotoxicity of the LPS. Its hydrophobic nature permits the anchoring to the OM. Structurally, the lipid A consists of two GlcN linked β-(1→6) each other, with a phosphate group at positions 1 and 4, and 3-hydroxy fatty acids attached to positions 2 and 3 through amide and ester linkages. The primary acyl chains can also undergo further modifications with secondary acyl chains.[54]

The core OS region, covalently linked to the lipid A, can be divided in inner and outer core. The inner core is the most conserved moiety among bacterial species, possessing peculiar sugars like heptoses or 3-deoxy-D-*manno*-oct-2-ulopyranosonic acid (Kdo) that is the sugar marker of Gram-negative bacteria (Figure 1.15). The outer core of the OS region, instead, has major variability and it is usually characterized by basic and neutral hexoses and hexosamines sugars.[55]



**Kdo**

**Figure 1.15:** Chemical structure of Kdo in β-configuration.

The O-antigen, absent in the LOS, is constituted by a species-specific variable number of oligosaccharide repeating units.[56]

In general, into the LPS structure of a bacterium, the lipid A moiety is responsible for the endotoxicity and pathogenicity, whereas core OS and O-antigen are implicated in the antigenicity and serotype-specific immunogenicity. Indeed, LPS as MAMPs, participate in host-guest interactions and upon recognition of host PRRs, stimulate the innate immune response. One of the better studied PRR complex recognizing LPS is host myeloid differentiation-2/Toll-like receptor 4 (MD-2/TLR4) receptor complex.[54]

## 1.4 Innate and adaptive immunity

The preceding section of this thesis elucidated the significance of bacterial glycoconjugates and their recognition by PRRs, emphasizing their role as distinctive MAMPs. This paragraph will focus on the crucial aspect of initiating the host immune response in the presence of a pathogen, delving into the functioning of the immune system during bacterial infections. It is important to note that this paragraph is intricately connected to the subsequent ones, particularly to the one dedicated to C-type lectins, proteins with a pivotal role in recognizing carbohydrates. The comprehensive understanding of these immune mechanisms is fundamental to unravel the intricate interplay between pathogens and the host immune system.

Humans have developed immune defence systems to eradicate harmful pathogens within the body. The human immune system can be viewed as a complex system with multiple layers of defence: (i) *external barriers*, which encompass both physical elements like skin, ciliated epithelia, and mucous membranes, as well as chemical components such as enzymes in secretions and stomach acids; (ii) *innate immunity* and (iii) *adaptive immunity* (Figure 1.16).[57]

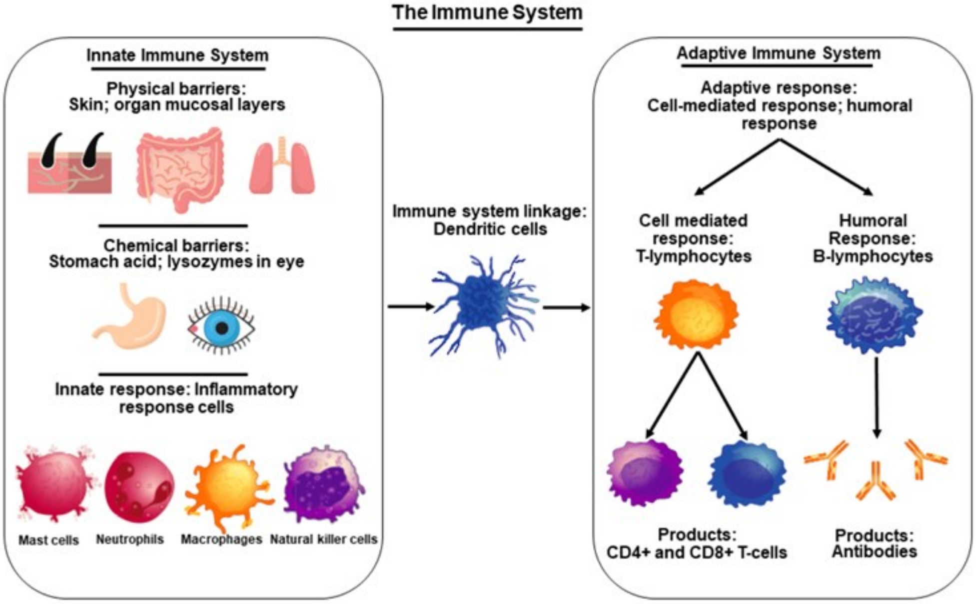


**Figure 1.16:** Overview of human immune system mechanisms (extracted from [58])

The innate immune response is the first and faster line of host defence against pathogenic microorganisms, consisting of a wide range of hematopoietic cells, including tissue-resident cells like macrophages and dendritic cells (DC), as well as mobile cells, such as neutrophils, eosinophils, and monocytes that circulate in the blood and lymph.[59] These cells are able to detect invading pathogens, recognizing their peculiar MAMPs, through specific PRRs, including[60] (i) *Toll-like receptors (TLRs)*, classified into different groups based on localization and recognition-type;[61] (ii) *nucleotide-binding oligomerization domain, (NOD)-like receptors (NLRs)*, cytoplasmic 'pathogen-sensing molecules' with the ability to recognize also danger-associated molecular patterns (DAMPs), [62, 63] and (iii) *lectins* (cf. §1.5.1).

Innate immunity primarily occurs by phagocytosis[64] thanks to the recruitment of effector cells to the infection site during the inflammatory response[65] upon the recognition of pathogens by *non*-lymphoid cells like endothelial cells or macrophages. Once recognized specific MAMPs, these cells release chemokines, which are small soluble proteins that act as chemotactic factors, attracting phagocytes from the bloodstream to the infection site.[66] Activated resident cells and phagocytes release cytokines, such as tumor necrosis factor (TNF)-α and interleukins (ILs), enhancing the phagocytic capabilities of innate immune cells. This increased secretion of cytokines and chemokines results in the recruitment of immune cells and plasma proteins to the infected tissue site, causing classic signs of inflammation (swelling, redness, pain, and heat). The inflammatory response not only brings antimicrobial agents to the infection site but, also, plays a crucial role in the tissue-repair.[67] This complex response possesses no specificity; a subsequent infection will trigger the same sequence of events.

The adaptive immune response, activated by innate immunity, is a more sophisticated system, with the peculiarity of having immunological memory, due to the production of memory T and B cells, lymphocytes responsible for the cellular and humoral immunity, respectively[68].

T cells express unique antigen-binding receptors called T-cell receptors (TCRs) and require antigen-presenting cells (APCs), usually DC, but also macrophages, fibroblasts and epithelial cells, to recognize a specific antigen. Also, APCs express proteins known as the major histocompatibility complex (MHC), class I or II in accordance with the localization, that displays fragments of antigens (peptides). This antigen presentation process stimulates T cells to differentiate into either cytotoxic T cells (CD8+ cells) or T-helper cells (CD4+ cells). CD8+ cells destroy target cells, CD4+ cells maximize the immune response.[68, 69] B cells, differently from T-cells, recognize antigens directly, with no need of APCs, thanks to the presence of antigen-binding receptors (BCRs) on their membrane. Their principal functions are the production of antibodies, by differentiation in plasma cells, and also, in some cases, they can act as APCs.[68]

## 1.5 Glycan binding proteins

Previously, an overview of the most important eukaryotic and bacterial glycans has been presented. The focus now shifts to the proteins that specifically recognize and interact with these glycans. This transition is essential as it unveils the second key actors in the elaborate host-guest interplay: GBPs. Their exploration will cross both human and bacterial realms. By exploring the diversity and specificity of GBPs, insights into the molecular dialogues that underlie crucial biological events will be provided. This comprehensive analysis is pivotal for advancing in the understanding of infection mechanisms, immune responses, and the development of targeted therapeutic strategies.

### 1.5.1 Exploring lectins: a comprehensive overview

The term "lectin" was first introduced in 1954[70] and it has consistently been employed to describe any previously identified or newly encountered proteins and glycoproteins capable of interacting with carbohydrates. Lectins are a class of GBPs, ubiquitous in nature, with roles in cell-cell interactions, signaling pathways, and innate immune responses against pathogens, acting also as APCs in the discrimination between *non-self*, altered-*self*, and *self*-biomolecular signatures. Given their ability to selectively and reversibly bind carbohydrates, they are deeply involved in the recognition of glycoconjugates exposed on the microbial surface. [71]

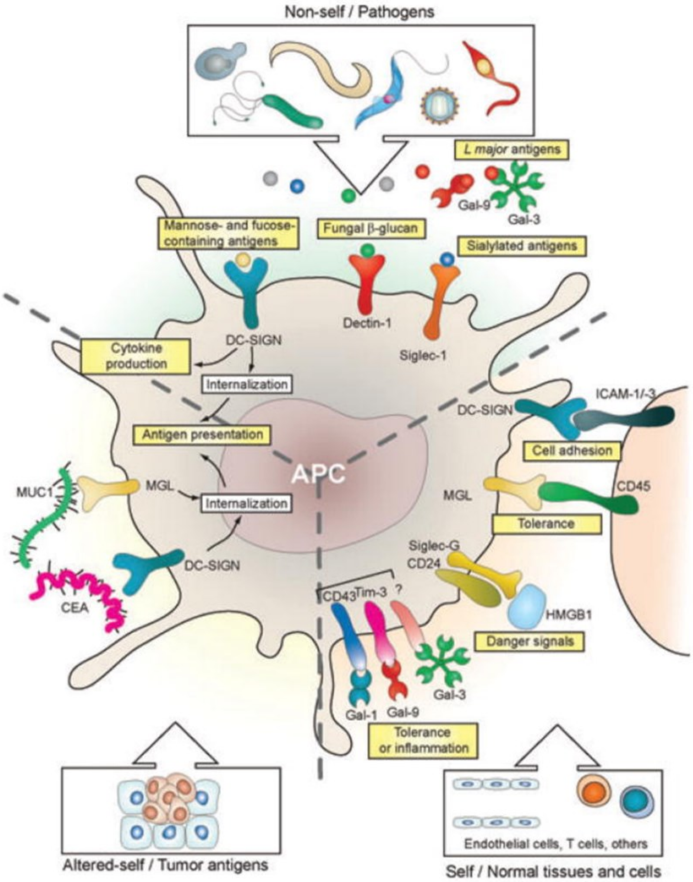


**Figure 1.17:** Representation of lectins functions as APCs (extracted from ref.[72]).

In vertebrates, lectins can be classified in two ways, based on their subcellular localization or their structural characteristics.

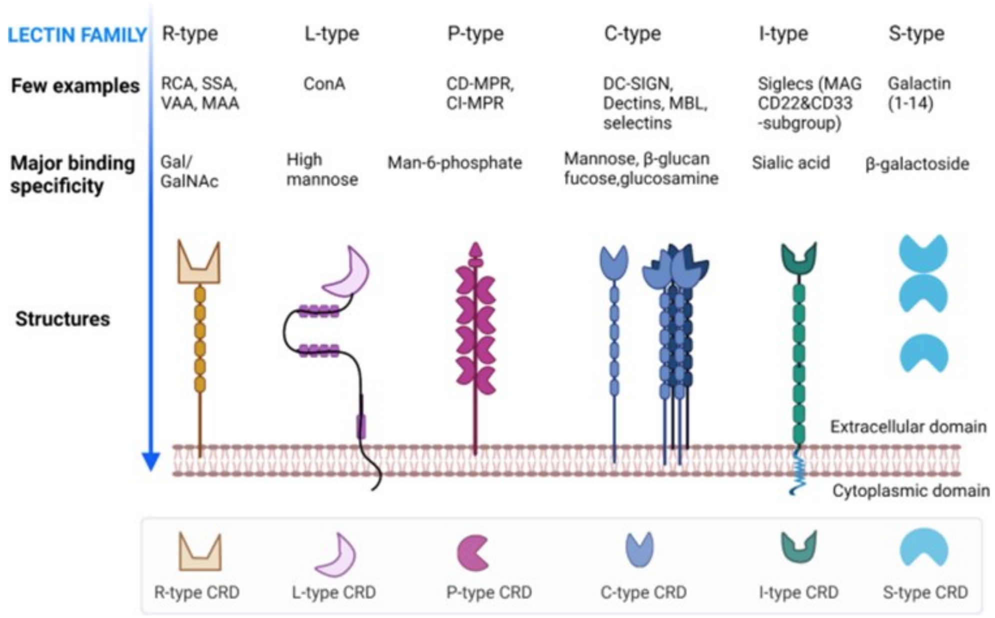
The classification by subcellular location distinguishes:

- **integral lectins**: structural membrane components;
- **soluble lectins**: freely circulating within intra- and intercellular fluids.[73]

The classification by structural characteristics consists of several different types of lectins, including some recently discovered (Figure 1.18):

- **C-type lectins**: $Ca^{2+}$ ions dependent to bind carbohydrates;

- **I-type lectins**: carbohydrate recognition domain (CRD) similar to immunoglobulins;
- **S-type lectins** (or galectins): thiol-dependent and specific to β-galactosides;
- **Pentranxins**: multimeric structures to form pentameric lectins;
- **P-type lectins**: specific to glycoproteins containing mannose 6-phosphate.[74]
- **Chitolectins**: binding-carbohydrates proteins, lacking of enzymatic activity, although their binding site resembles the catalytic Family 18 evolutionary group, to which they are affiliated;[75]
- **F-type lectins** (or fucolectins): recognition of α-L-Fuc residues, with unique carbohydrate- and calcium-binding sequences;[76]
- **F-box lectins:** characteristic for having protein-protein interaction motifs;[77]
- **Ficolins**: oligomeric lectins with role in innate immunity by recognizing carbohydrates in Gram-positive and Gram-negative bacteria;[78]
- **L-type lectins**: different from others due to the tertiary structure, composed of antiparallel β-sheets connected by short loops and β-bends, usually α-helices are not present;[79]
- **M-type lectins**: closely related to α-mannosidases, with no mannosidase activity;[80]
- **R-type lectins:** protein-UDP acetylgalactosaminyltransferases containing an R-type carbohydrate recognition domain;[81]
- **X-type lectins** (or intelectins); no typical lectin domain but contain a fibrinogen-like domain and a unique intelectin-specific region.[82]



| LECTIN FAMILY | R-type | L-type | P-type | C-type | I-type | S-type |
|---|---|---|---|---|---|---|
| Few examples | RCA, SSA, VAA, MAA | ConA | CD-MPR, CI-MPR | DC-SIGN, Dectins, MBL, selectins | Siglecs (MAG CD22&CD33 -subgroup) | Galactin (1-14) |
| Major binding specificity | Gal/ GalNAc | High mannose | Man-6-phosphate | Mannose, β-glucan fucose,glucosamine | Sialic acid | β-galactoside |

**Figure 1.18:** Cartoon representation of the structures of the most important lectins.

Studies have revealed that most lectins have been preserved across evolutionary lines, indicating their essential role in the sugar-recognition processes, pivotal for vital processes and development.[83] Particularly, the specificity and the affinity of the lectin-carbohydrate complex depends on the lectin involved, that can be very sensitive to the structure of the carbohydrate, such as the distinction between Man and Glc, and/or to the orientation of the anomeric substituent, whether it is an α or β anomer. Interactions between lectins and carbohydrates occur mainly through (i) *hydrogen bonds*, (ii) *van der Waals forces* (steric interactions) and (iii) *hydrophobic interactions*.[84] Indeed, *non*-covalent interactions play a crucial role in molecular recognition processes, defining the protein-ligand interface and contributing to the complex formation. In neutral carbohydrates, hydrogen bonds represent the dominant binding forces, due to the presence of hydroxyl groups. These latter can exhibit dual roles as hydrogen-bond donors and acceptors, providing cooperative hydrogen bonding, commonly observed in glycan-lectin complexes. In charged carbohydrates, salt bridges can also be established with charged residues of the protein, contributing to polar interactions.

Furthermore, coordination interactions with metal ions can also take place. This is particularly evident in the interaction between carbohydrates and lectins, especially C-type lectins. In this case, the binding is often coordinated by a divalent metal cation $Ca^{2+}$ which is stabilized in the binding site of the lectin.[85] In addition, stacking interactions, also called CH–π interactions, are also involved in the stabilization of the carbohydrate complexes with aromatic residues, further contributing to the stabilization of molecular conformations and facilitating the binding.[86]

#### 1.5.1.1 Structure and function of C-type lectins

Among all types of lectins, C-type lectins (CTLs) stand out as one of the largest and most diverse family found in animals. They are soluble or transmembrane proteins, divided into 17 subgroups according to the phylogeny and domain organization (Figure 1.19), with one or more C-type lectin-like domains (CTLDs).[87, 88] While these proteins were initially termed for their $Ca^{2+}$-dependent carbohydrate-binding capacity, there are CTLs where the presence of a calcium-binding site is not essential for carbohydrate interaction. A notable example is dectin-1, a cell-surface immune receptor for β-glucans, the primary structural components in fungal cell walls. The recognition of β-glucans by dectin-1 can initiate the phagocytosis of fungal pathogens and elicit protective inflammatory responses. Despite possessing a CTLD, its metal-binding site is uniquely positioned, and, notably, the binding of β-glucans by dectin-1 does not rely on metal ions. Nevertheless, studies indicate that $Ca^{2+}$ ion binding in this distinct position provides structural stability to the protein domains and may contribute to unidentified physiological functions.[89]

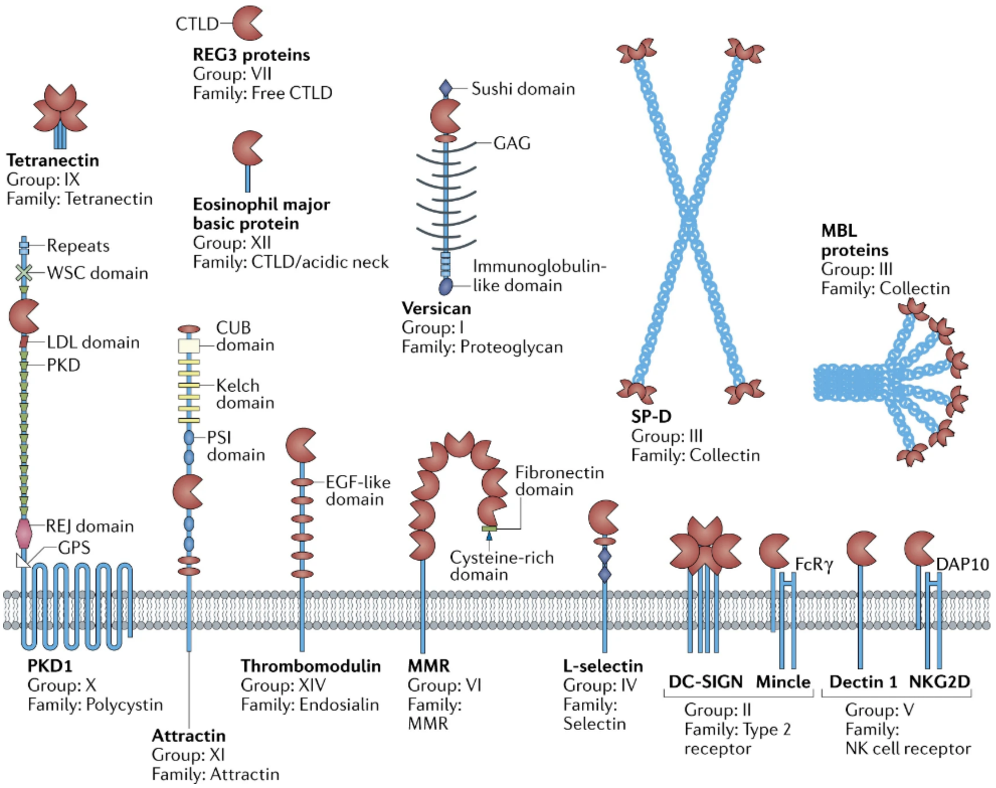


**Figure 1.19:** Cartoon representation of different soluble and membrane-bound C-type lectins (extracted from ref.[88]).

From the structural point of view, the carbohydrate-recognition domains (CRDs) of CTLs share some common secondary structure elements. The CRD is a region of 110–130 amino acid residues with a double-looped, two-stranded antiparallel ß-sheet formed by the amino- and carboxy-terminal residues connected by two α-helices and a three-stranded antiparallel ß-sheet. There are also two conserved disulfide bonds and up to four sites for binding $Ca^{2+}$, depending on the lectin occupancy. The sugar, the $Ca^{2+}$ ion in site 2 (always occupied being the locus for the carbohydrate binding) and amino acids within the CRD form a ternary complex where the residues composition confers the sugar specificity (Figure 1.20).[90]

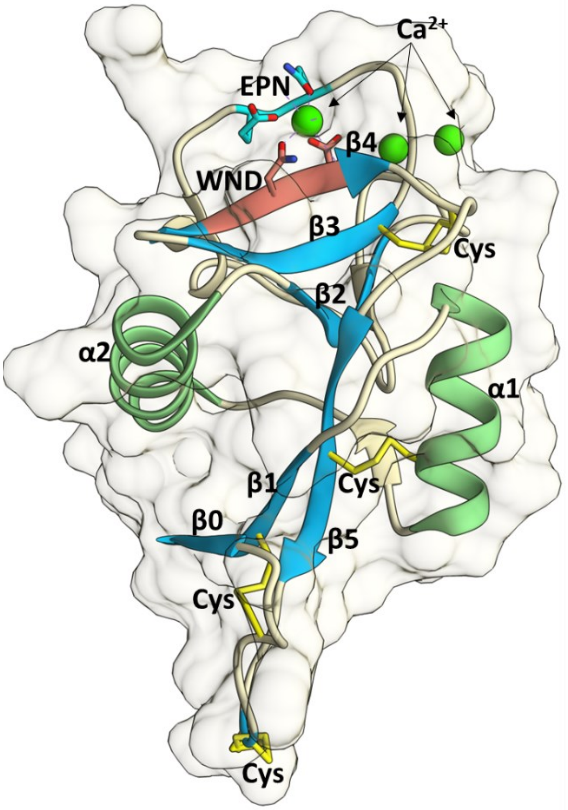


**Figure 1.20:** Structural organization of CRD, using DC-SIGN as example. Highlighted typical EPN and WND motifs, secondary structural elements, four Cys and $Ca^{2+}$ binding site locations, with two free coordination points for the attachment of the sugars, as described in the text (PDB:1SL4).

Interestingly, some tripeptides that bind sugars are well conserved, namely, the WND motif (Trp-Asn-Asp) on the β4-strand, the EPN (Glu-Pro-Asn) and QPD (Gln-Pro-Asp) motifs at the loop. Particularly, Man and Fuc binding specificities are associated to the EPN, while Gal and GalNAc to the QPD (Figure 1.21).[91] However, some new motifs have been discovered, like the QAP motif,[92] together with interactions with novel glycans,[72] which further complicate the analysis of the recognition of this specific and extensive class of proteins.

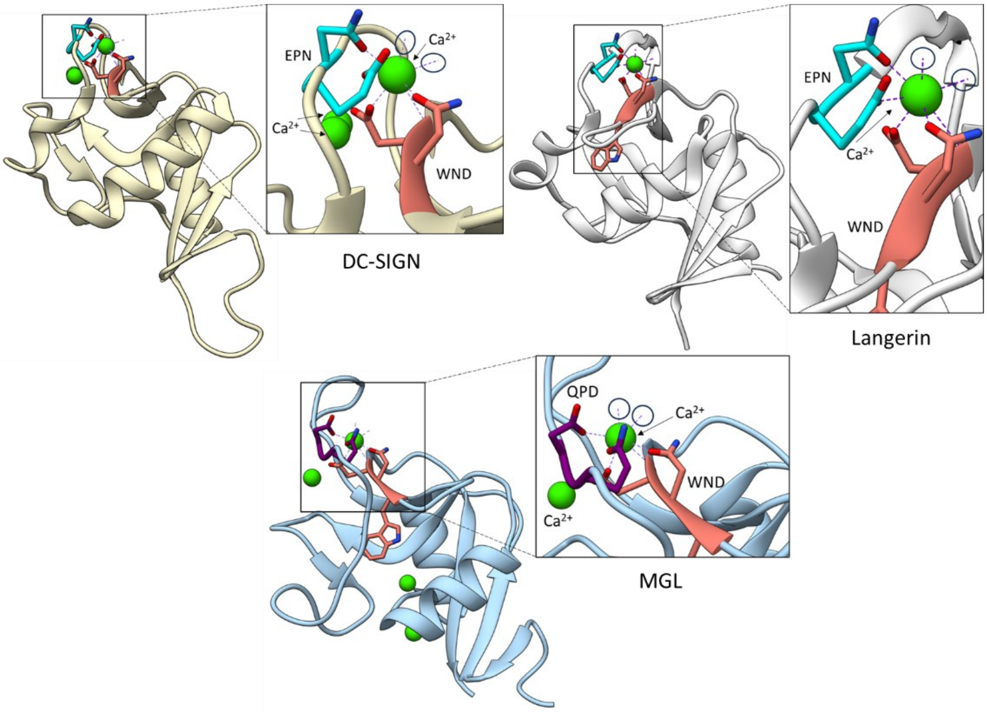


**Figure 1.21:** CRD of different CTLs. Highlighted in the squares the characteristic WND (in salmon), EPN (in cyan) and QPD (in violet) motifs, together with $Ca^{2+}$ binding site locations, in the circles the two free coordination points for the attachment of the sugars (PDB: 1SL4, 4N33, 6PY1).

From the functional point of view, generally CTLs are characterized not only by the critical role within innate and adaptive antimicrobial immune responses, but, also, in homeostasis and autoimmune diseases, due to their ability to recognize *self* and *non-self* ligands. It has been demonstrated their functions in cell-cell communication, adhesion, and cellular signaling; indeed, while numerous CTLs present on immune system cells are known to trigger intracellular signaling, this aspect of their function remains not completely understood. The different configurations of these receptors within the cell's plasma membrane underscore the existence of multiple distinct mechanisms through which sugar binding at the cell surface can instigate events on the cytoplasmic side of the membrane (Figure 1.22).[93] For example, the intracellular regions of two receptors, dectin-1 found on macrophages and DCs, and prolectin present on B cells, possess signaling motifs that enable the direct activation of Syk kinase. Activations initiated through dectin-1 signaling encompass processes like phagocytosis, the respiratory burst, and the generation of inflammatory cytokines such as TNF-α and IL-6. Yet, both DC-SIGN and MGL (macrophage galactose receptor) have been documented to influence the signaling pathways triggered by TLRs. Even if for MGL not specific binding partners have been proposed,[94] regarding DC-SIGN it is proposed that the adapter LSP1 interacts with its cytoplasmic domain and becomes phosphorylated through a pathway initiated by TLR4.[95] Regarding the immune response, CTLs are also involved in antigen presentation, by presenting antigens to MHC class II to activate CD4+ T cell response or to CD8+ through MHC class I.[72]

Based on these premises, the study of CTLs is crucial since this class of proteins mediate key recognition events, binding processes, and immune responses, serving as essential components in the host-guest interactions. A deeper understanding of these mechanisms lays the foundation for new therapeutic and diagnostic applications.

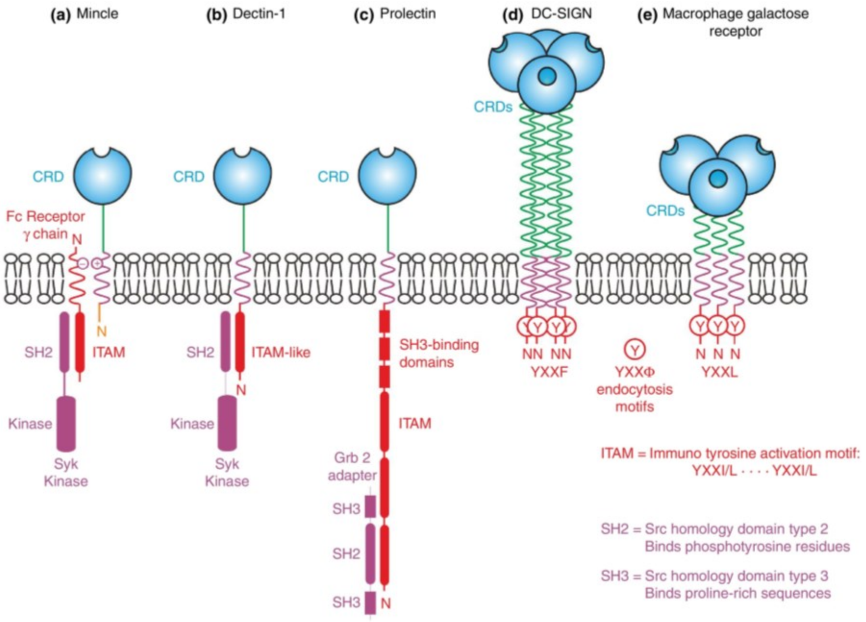


**Figure 1.22:** Examples of organization of different CTLs containing extracellular C-type CRDs and intracellular domains involved in signaling. (extracted from ref.[93]).

### 1.5.1.2 Key aspects of DC-SIGN

DC-SIGN is a notable C-type lectin receptor playing critical roles in the regulation of the immune responses.

DC-SIGN, also known as CD209 (Cluster of differentiation 209) is a type II transmembrane protein, expressed both in DCs and macrophages,[96] able to bind carbohydrates in a Ca$^{2+}$-dependent manner. Structurally, its extracellular domain (ECD) (MW 38845.8 Da) comprises an elongated *neck domain* (ND), useful for the tetramerization of the protein, that has tandem repeats of highly conserved 23 amino acids, with an N-terminal in the cytosol, followed by a C-terminal *C-type carbohydrate-recognition domain* (CRD), outside the membrane, which is the heart of the molecular recognition processes.[97] In detail, the CRD of DC-SIGN possesses a globular structure characterized by 12 β-strands, 2 α-helices and 3 disulphide bridges. Extending from the protein surface, a loop is part of the two Ca$^{2+}$-binding sites; one of these Ca$^{2+}$ sites is critical for maintaining the CRD conformation, while the other is crucial for directly coordinating carbohydrate structures.[96] The principal binding site, represented by the EPN motif (Glu347, Pro348 and Asn349) together with Glu354 and Asn365 residues, is essential for carbohydrate coordination and manages the specificity for the recognized ligands.[98] Mutations at these sites result in the loss of ligand binding. Then, there is a *transmembrane domain* (TMD) which anchors the protein onto the membrane and a *cytoplasmic domain* (CD), containing several internalization motifs (i.e. di-Leu motif) that suggest the role of DC-SIGN also as endocytic receptor (Figure 1.23).[99]

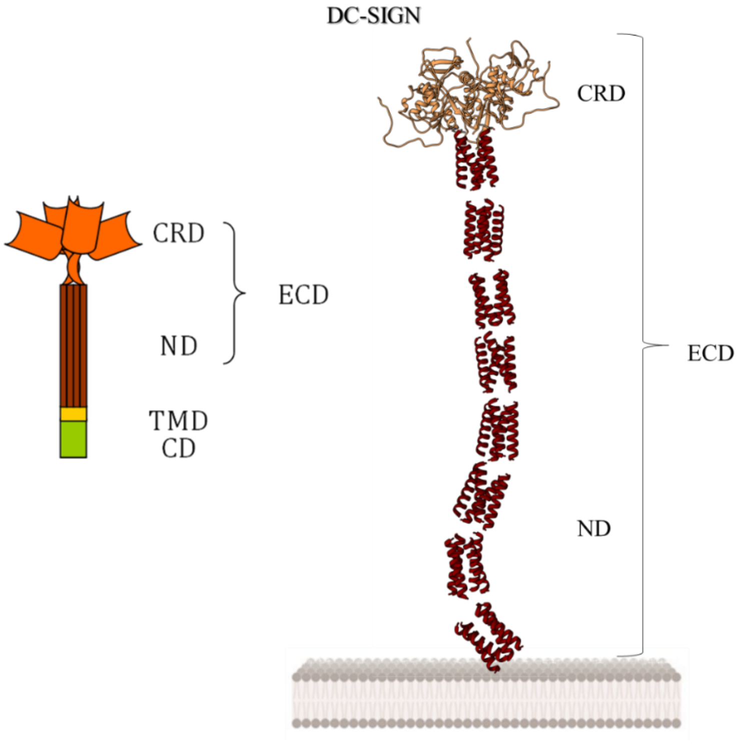


**Figure 1.23:** Organization of DC-SIGN structural domains.

As a CTL receptor, DC-SIGN possesses the capacity to differentiate among various ligands, enabling it to modulate TLRs signaling in either a pro-inflammatory or anti-inflammatory manner.[100] Notably, DC-SIGN is able to specifically recognize fucosylated glycans, including blood-type Lewis antigens (Lea, Leb, LeX, LeY, and sulfo-Lea), as well as mannose-expressing ligands,[101] which trigger TLR-induced cytokine responses. The pro-inflammatory response is associated with Man-containing structures, while the anti-inflammatory response is elicited through interactions with fucosylated residues. Both Man and Fuc ligands induce the secretion of IL-10, but the crucial distinction occurs on the impact on IL-6 and IL-12. Fucosylated ligands reduce LPS-induced IL-6 and IL-12 secretion, whereas mannosylated glycans enhance the production of both ILs. Consequently, a pro-inflammatory response is attributed to Man-expressing ligands, while an anti-inflammatory response is associated with Fuc-expressing ones.[102] However, recently, the ability to recognize GlcNAc or Gal residues by DC-SIGN has also emerged.[103, 104]

In conclusion, DC-SIGN plays a pivotal role in orchestrating the immune response by recognizing carbohydrate structures exposed on pathogens or self-glycoproteins. Its multifaceted functions encompass adhesion, migration, signaling, and antigen uptake/presentation, making it a central player in immune-related processes. Despite the expanding knowledge of DC-SIGN's biological roles, a comprehensive understanding of its interaction with bacterial glycan ligands remains elusive. The molecular details of this interaction are crucial for unraveling the intricacies of host-pathogen dynamics. Furthermore, DC-SIGN contributes to antigen uptake by antigen-presenting cells. Its role in sensing commensal bacteria promotes host-microbiota mutualism, preventing systemic inflammation. DC-SIGN's capacity for both tolerance and immune response modulation, depending on the context, raises intriguing possibilities for pathogen evasion and the development of therapeutic strategies.

Investigating C-type lectins, especially DC-SIGN, provides critical insights into microbial recognition, immunity, and host-microbiota interactions. This knowledge is instrumental not only for advancing our understanding of fundamental immunological processes but, also, for exploring potential therapeutic applications, such as in the context of inflammatory diseases related to the gut microbiota.

### 1.5.2 Bacterial glycan-binding proteins

As mentioned earlier, GBPs are largely involved in host-guest interactions and are present not only in eukaryotic cells, but also in prokaryotes. Bacterial lectins were firstly described in 1970 by N. Sharon and colleagues, in the context of bacterial attachment to the host cell for further colonization and infection. For this reason, bacterial lectins usually refer to bacterial adhesins, which bind specific host cell-surface glycoproteins, glycosphingolipids, or GAGs, by means of their CRDs.[105]

Adherence to skin or mucosal surfaces could be identified as a defining feature of the human microbiome. It stabilizes symbiosis ranging from a commensality relationship where the interaction benefits one of the symbiotic partners, while some other species involved neither gains benefit nor is harmed, to a mutualistic relationship with favourable outcomes for all the organisms involved. Although a vast majority of these microbes carry out critical functions for host physiology, they may be also the initial step in many infectious diseases, leading to both chronic and acute type inflammation.[106] The adhesion process can occur through the recognition of terminal sugars or internal carbohydrate motifs, triggering an intricate signal transduction pathway within the host cell, which may result in the activation of innate defence mechanisms or in the manipulation of cellular functions that aid bacterial colonization.[107] The different carbohydrate ligands on host cells that serve as receptors for bacterial attachment, determine the tropism of individual bacteria for specific host tissues, such as the skin, respiratory tract, or gastrointestinal tract.[108] Moreover, the nature of the adhesion can vary depending on the biochemical function. The binding with host can be weak and aspecific, establishing hydrophobic interactions, but also more stable and with higher affinity, through more specific interactions.[107]

Bacterial adhesins are typically found in pili or fimbriae, bacterial surfaces appendages composed of protein subunits with different lengths (0.2-2 μm) and diameters (2-10 nm), useful for their motility and for the attachment to host cells. As examples, some strains of E. *coli*, express pili that strongly adhere to P-blood group-related glycosphingolipids in the bladder epithelium, leading to urinary tract infections. Also, pathogenic *Salmonella* strains produce pili that aid in adherence to human intestinal cell mucosa, causing food poisoning and infectious diarrhea (Figure

1.24A).[109, 110] Alternatively, bacterial surface-anchored proteins, and so afimbrial adhesins, can also represent a critical colonization factor. For instance, the filamentous hemagglutinin (FHA) of *Bordetella pertussis* mediates the attachment of the bacterium to the ciliated epithelial cells of bronchi and trachea, causing local inflammation and the so called "whooping cough" disease. To note, FHA has been incorporated into modern pertussis vaccines for children, to prevent infection pointing out the role of adhesins also as targets for antibodies (Figure 1.24B).[111] Moreover, in *Burkholderia cenocepacia,* responsible for highly contagious pneumonia among cystic fibrosis (CF) patients with high rate of mortality, three soluble lectins BC2L-A, BC2L-B and BC2L-C have been identified as protagonists of their adhesion to host tissues using oligosaccharide-mediated recognition processes. For example, the smallest lectin BC2L-A specifically interacts with Man residues in LPS epitopes, in $Ca^{2+}$- dependant manner. Yet, two soluble lectins PA-IL and PA-IIL in *Pseudomonas aeruginosa* adhere to the host tissues recognizing galactosides and fucosides, contributing to chronic lung infections in CF patients.[112]

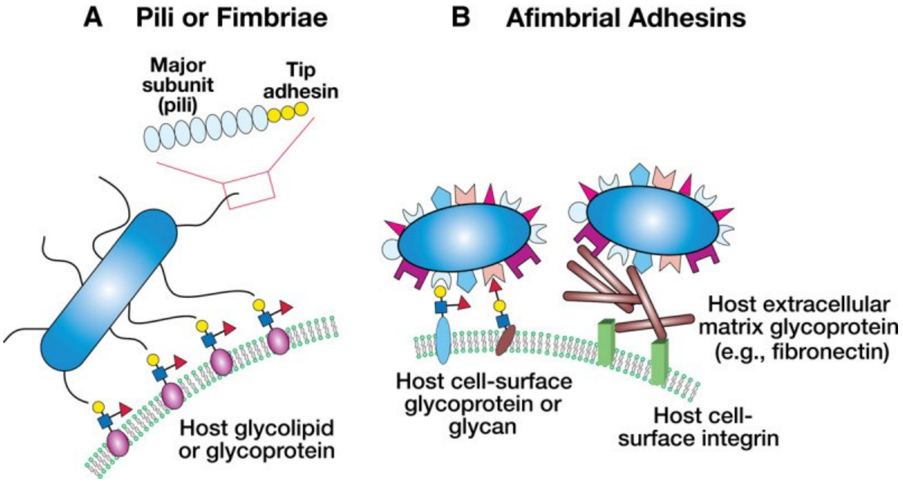


**Figure 1.24:** Examples of bacterial adherence to host cell surfaces, mediated by pili or fimbriae containing adhesins (A) and afimbrial adhesins (B) (extracted from ref.[108]).

Another important class of bacterial adhesins to mention is Siglec-like adhesins, which play a crucial role for the attachment of different *Streptococcus* strains to host tissues. SLBR-B and SLBR-H are two Siglec-like adhesins produced by *Streptococcus gordonii*, a Gram-positive oral commensal bacterium, responsible for infective endocarditis (IE) if present in the bloodstream.[113]. They possess typical "Siglec-like" binding regions (SLBRs), consisting of two conserved domains crucial for the host sialoglycans binding: a V-set Ig fold Siglec subdomain, closely resembling those found in mammalian Siglecs (sialic acid binding immunoglobulin-like lectins) in terms of topology and strand inserts, from this the name "Siglec-like adhesins"; a Unique domain, which isn't directly involved in carbohydrate interaction but may influence the conformation of the Siglec domain. Together these two domains are pivotal for bacterial recognition of host sialoglycans, containing terminal Neu5Ac-α-(2,3)-Gal.[114] These Siglec-like adhesins have attracted attention because, despite having a similar fold to mammalian Siglecs, the recognition processes function quite differently. Indeed, while the binding pockets of bacterial Siglec-like adhesins, like SLBR-B and SLBR-H, fully accommodate the sialylated glycans,[114] in the case of mammalian Siglecs, only Neu5Ac is principally recognized and accommodated in the binding pocket.[115]

So, the diversity and specificity of the bacterial adhesins and their interactions with host glycans underscore the complexity of host-pathogen relationships and the importance of understanding these processes in the context of infectious diseases.

#### 1.5.2.1 Cytoadhesins

Several *mycoplasma* species, including two human pathogens *Mycoplasma pneumoniae* (*Mpn*) and *Mycoplasma genitalium* (*Mge*), express adhesins interacting with specific sialylated glycans exposed on target host cell surfaces, thus mediating the bacterial adherence and causing mycoplasmas-associated infectious diseases.[116, 117] *Mpn,* an exclusively human parasite,[118] primarily affects the conducting airways, being the leading cause of upper and lower respiratory tract infections like community-acquired pneumonias (CAPs).[116, 119] On the other hand, *Mge* is a sexually transmitted bacterium that frequently co-infects the urogenital tract with other pathogens, such as HIV, *C. trachomatis*, and *N. gonorrhoeae*[120]. In particular, in women, it is responsible for several inflammatory reproductive tract syndromes like cervicitis,[121] whereas, in men, it is the leading cause of 30-40% cases of urethritis.[122]

*Mpn* and *Mge* cytoadherence and gliding motility are essential for colonization of the human tissues and are mainly mediated by a complex attachment organelle, consisting of a specialized tip structure with a central core of a dense rod-like filament composed of adhesins, interactive proteins, and adherence accessory proteins.[123] Essential for infectivity is the transmembrane adhesion complex called Nap[124] (Figure 1.25), with a tetrameric organization composed by heterodimers of the cytoadhesins P140 and P110 in case of *Mge*, and their homologs P1 and P40/P90 for *Mpn*.[125, 126] These cytoadhesins possess a similar domain organization, consisting of a large extracellular domain (the crown), a transmembrane helix and a short cytoplasmic tail (C-domain). The extracellular portion is divided into two subdomains: a small C-terminal domain and a N-terminal region constituted by a seven-bladed β-propeller, creating a crown structure, where the sialic acid binding site is located.[127]

Given the increasing emergence of antibiotic resistance documented in *mycoplasma* species, such as for *Mpn* and *Mge*, the development of alternative therapeutic strategies to counteract mycoplasmas-associated infections is urgently needed. Despite structural features of these main cytoadhesins have been already elucidated[124], molecular insights on the sialoglycans recognition and binding are still needed, in order to provide alternative anti-adhesion strategies, either by reducing the contact between host tissues and pathogens, and/or by prevention of adhesion of the infectious agent.

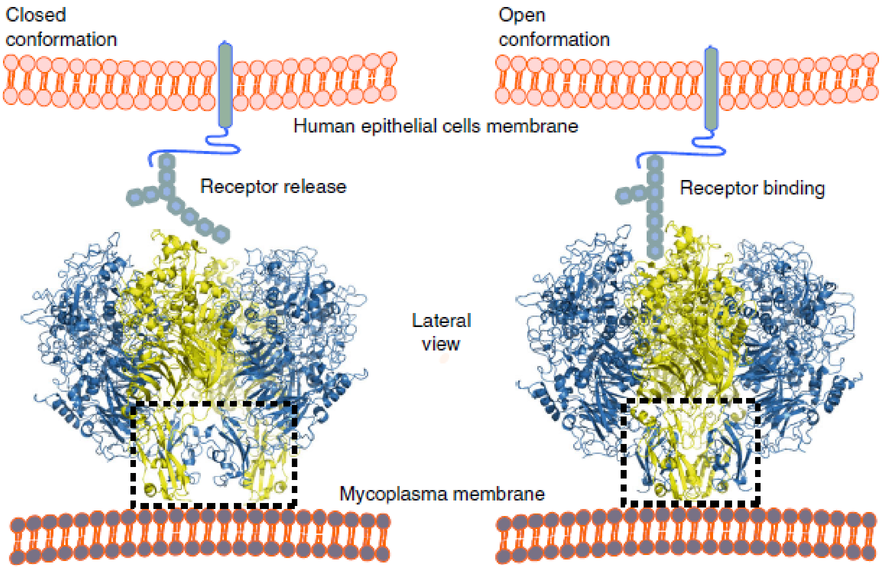


**Figure 1.25:** Structural rearrangements between "open" and "closed" conformations of NAP adhesion complex. (adapted from ref.[124])

## 1.6 Objectives of the thesis

This thesis is based on the understanding at a molecular level of the key events at the basis of host-guest interactions. To reach this aim, an in-deep knowledge of the structures and functions of bacterial and eukaryotic glycans is fundamental for further investigate their recognition from both human and bacterial lectins to depict the molecular mechanisms at the basis of pivotal biological processes, like cell-cell communication, bacterial adherence, modulation of immune responses, development of diseases and tumor progression. The investigation of the binding of different human and bacterial lectins to their selected ligands has been carried out by using different wet lab procedures, NMR spectroscopy, Mass Spectrometry and other different biophysical techniques like Fluorescence and Surface Plasmon Resonance together with computational approaches. In the chapters of the next section, the results obtained from the study of different systems will be shown.

In the context of the immune system regulation, either in a pro- and an anti-inflammatory response, after the expression and purification of the human lectin DC-SIGN, carried out during my secondment at Institut de Biologie Structurale in Grenoble (France), in the lab of Prof. F. Fieschi, the molecular basis of its recognition of the R1-type LOS of the pathogenic bacterium *E. coli* has been unveiled (Chapter III). In addition, the molecular recognition of the LPS of the commensal bacterium *B. vulgatus* by DC-SIGN, has been also explored (Chapter IV).

Human-microbe interactions have been also investigated in the case of *Mpn* and *Mge* bacterial adhesins. *Mpn* and *Mge* are two human pathogens which colonize the respiratory and urogenital epithelia, respectively. They express cytoadhesins, located on the bacterial surface, able to interact with the host, mediating the attachment to sialylated glycan receptors and triggering the bacterial infections. So, given the high incidence of pathologies related to *Mpn* and *Mge*, the absence of a vaccine for *Mpn* in contrast to other respiratory pathogens, and the rapid increase of antibiotic resistance of mycoplasma pathogens,[128] the recognition by the cytoadhesins P40/P90 in *Mpn* and

P110 in *Mge* of sialylated *N*- and *O*-glycans has been investigated (Chapter V), with the aim to find alternative anti-adhesion strategies.

Furthermore, the persistent increase in antibiotic resistance across many bacterial species represents one of the key contemporary challenges to public health. In this context, bacterial species belonging to the ESKAPE pathogens attract attention and among them, *A. baumannii*, responsible for various nosocomial infections, is object of the present study. It possesses the peculiar production of CPS which confer the resistance to adverse environmental conditions. In this thesis, the isolation, purification and the structural characterization of CPS from an *A. baumannii* clinical isolate will be shown, together with some preliminary results regarding its interaction with host immune receptors. Furthermore, its conformational behavior in the free state has been also evaluated through computational studies, after the parametrization of a peculiar sugar contained in its CPS structure. Yet, thanks to my stay at GSK - Siena (Italy), under the supervision of Dr. M.R. Romano and Dr. R. Adamo, *in vivo* studies have been performed to evaluate the immunogenic activity of the characterized CPS conjugated to the carrier protein $CRM_{197,}$ on mice (Chapter VI). Understanding these molecular intricacies can shed light on the mechanisms underlying the pathogen's virulence and provide insights into novel therapeutic strategies and vaccine development, critical for combating infections caused by this multidrug-resistant bacterium.

# Chapter 2 - Analytical techniques for investigating protein-glycan interactions

This chapter makes a journey from wet lab procedures useful for the isolation of the partners of protein-glycan interactions, delving into the principles of Mass Spectrometry (MS), Nuclear Magnetic Resonance (NMR) and diverse other biophysical techniques for the glycan structural characterization, and culminating in the domain of computational modeling to investigate the tridimensional behavior, of glycans on the free and bound state. Additionally, key concepts for the expression and purification of proteins will be also provided.

## 2.1 Extraction and purification of bacterial complex glycans

The isolation of bacterial cell wall glycoconjugates initiates with their extraction from microbial cells. This thesis focuses on the extraction methods for LOS, LPS, and CPS, as these glycoconjugates constitute the primary subjects of study. Within this framework, two distinct extraction procedures are employed, each tailored to the varying amphiphilic character of the compounds. Specifically, the phenol-chloroform-petroleum ether (PCP) extraction method is employed for obtaining more hydrophobic compounds, such as the **LOS from *E. coli*** utilized in this thesis, which possesses a shorter saccharide chain with respect to the LPS, due to the absence of the O-antigen portion. The PCP extraction involves treating dried microbial cells with a mixture of petroleum ether, chloroform, and 96% aqueous phenol in an 8:5:2 proportion (v/v/v). The second extraction method, designed for more hydrophilic compounds, like CPS and LPS, is known as hot phenol-water extraction. In this method, cells are treated with a 1:1 phenol-water mixture at 70°C, resulting in a water phase and a phenol phase. For instance, **CPS from *A. baumannii*** and **LPS from *B. vulgatus***, objects of the present study, are recovered from the water phase due to their highly soluble polysaccharide nature. However, in cases where the core oligosaccharides of LOS exhibit variable sizes, possibly influenced by bacterial growth conditions or LPS modifications, it is advisable to combine the two described methods to obtain the desired LOS molecules. Consequently, a single extraction method is rarely employed for obtaining polysaccharides, making a combination of methods more appropriate. Following dialysis, the extracted phases undergo enzymatic hydrolysis with DNAse, RNAse, and protease to remove nucleic acids and protein contaminants. After another step of dialysis, the obtained samples are examined on a sodium dodecyl sulfate-polyacrylamide electrophoresis gel (SDS-PAGE) with silver nitrate staining. CPS (over 100kDa), LPS (50-100 kDa) and LOS (few kDa) appear on the gel in different ways due to the variation in sugar content. Usually, CPS appears as a band on the top of the gel because, owing to its high molecular weight, it does not migrate. The higher molecular weight of LPS, compared to LOS, results in the observation of multiple dispersed bands throughout the gel profile. These bands arise from the heterogeneity of LPS, which can be attributed, for instance, to variations in the lengths of the O-chain moiety. In contrast, LOS bands are identifiable at the lower region of the gel, consistent with their lower molecular weight. After gel electrophoresis, additional purification steps may be required, such as size exclusion chromatography (SEC), ultracentrifugation or mild acidic hydrolysis, to eliminate the lipidic component, essential for cell membrane attachment (Figure 2.1).

The chemical structure of the purified samples can be determined by using different techniques, mainly including GC-MS and NMR, as detailed in the following paragraphs.[129]

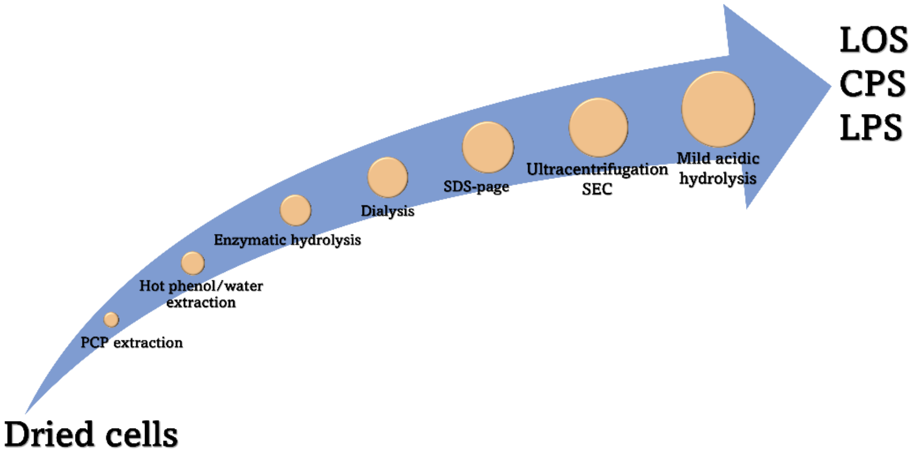


**Figure 2.1:** General scheme of the experimental procedures used to isolate bacterial glycoconjugates.

## 2.2 Structural characterization of bacterial glycans

The determination of glycan chemical structure consists of establishing the following characteristics:[130]

- Monosaccharides' composition and their absolute configurations;
- Linkage patterns as identification of glycosidic linkage positions and potential branching sites;
- Furanosidic or pyranosidic rings;
- α- or β- anomeric configurations;
- Sequences of monosaccharide residues and, in the case of glycans with a regular structure, determination of the repeating units;
- Identification and localization of substituents on hydroxyl groups;
- Molecular size.

To reach all this information, it is necessary to combine various biophysical techniques, such as Mass Spectrometry and NMR spectroscopy.

## 2.2.1 Mass Spectrometry (GC-MS analysis)

The coupling of Gas Chromatography (GC) with Mass Spectrometry (MS) is a potent analytical technique to perform a preliminary compositional analysis of glycoconjugates. However, since this method allows to analyze volatile compounds, derivatization of glycan samples is needed.

Generally, glycans are firstly separated into their constituent monosaccharides (the depolymerization procedure consists of acid catalyzed solvolysis, using either water or other solvents) and subsequently transformed into derivatives suitable for the GC-MS analysis. Analysis of derivatized monosaccharides allows for the identification of sugar residues based on their fragmentation pattern and retention time (Figure 2.2).

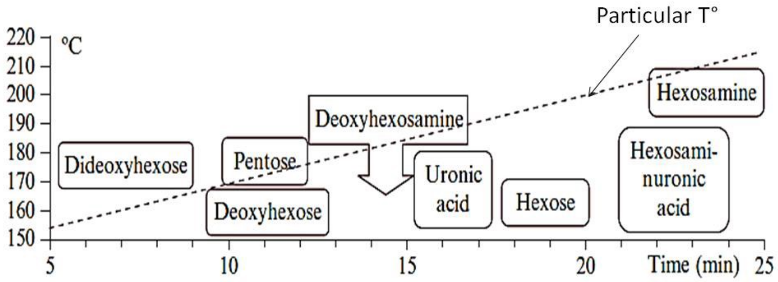


**Figure 2.2:** Elution order corresponding to each type of derivatized sugar in the GC-MS chromatogram.

Due to the heterogeneity of sugar molecules, the choice of the derivatization method is crucial and allows to achieve different information. Thus, as examples, four derivatization methods will be detailed: (i) *Acetylated Alditols* (AA) and (ii) *Acetylated Methyl Glycosides* (AMG), both used for the determination of monosaccharide composition, (iii) *Partially Methylated and Acetylated Alditols* (PMAA) useful for the linkage analysis, (iv). *acetylated octyl- (or butyl-) glycosides*, used for the determination of the absolute configuration.

The AA method is applicable for identifying neutral sugars, including aldoses, ketoses, and aminosugars. The procedure typically involves an initial hydrolysis, commonly with trifluoroacetic acid (TFA) to break the glycosidic linkage.

Subsequently, there is a reduction step targeting the free carbonyl group, followed by a final acetylation of the hydroxyl groups (Figure 2.3).

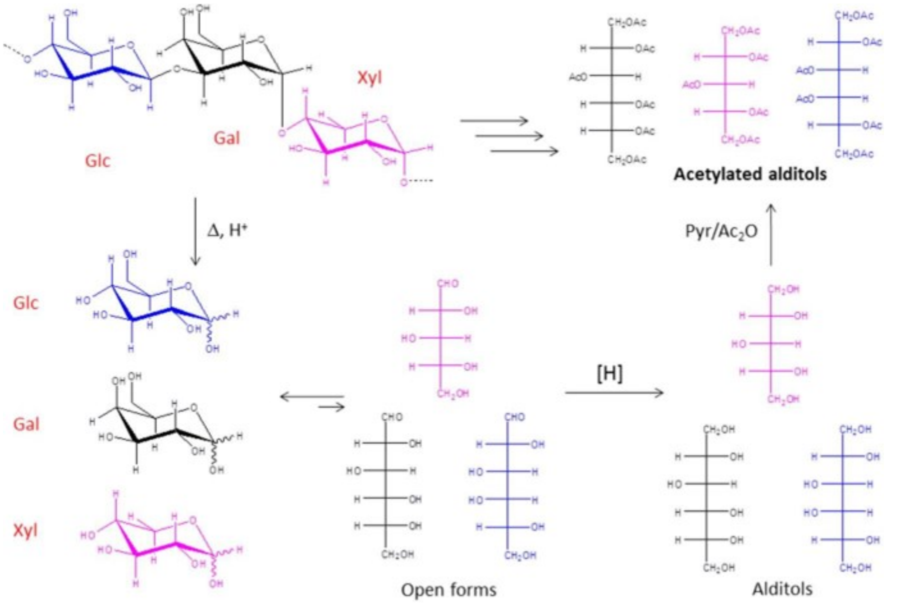


**Figure 2.3:** Scheme of the reactions involved in converting a polysaccharide into its corresponding acetylated alditols (adapted from ref. [130]).

AMG, in contrast to AA, is suitable for detecting acidic sugars and serves as a general procedure for identifying hexoses, aminosugars, uronic acids, and ulosonic acids. This method requires particular attention in the initial step, involving methanolysis under anhydrous conditions, which can pose a limitation.

In detail, the saccharide sample undergoes methanolysis with MeOH/HCl, followed by acetylation of the free monosaccharides. The resulting acetylated O-methyl glycosides are then analyzed using GC-MS, where identification occurs through fragmentation pattern and retention time analysis, and also by comparison with standards. This approach employs mild degradation conditions (MeOH/HCl 0.5-2M, 80°C) and is especially recommended for acid-labile sugars, such as Kdo and deoxy-sugars (Figure 2.4).

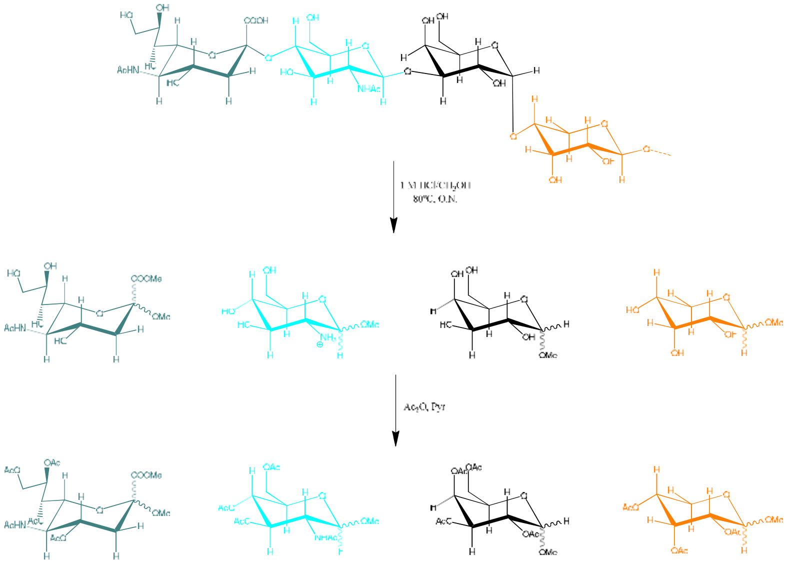


**Figure 2.4:** Scheme of the reactions involved in converting a polysaccharide into its corresponding acetylated methyl glycosides (adapted from ref.[130]).

The derivatization method for the obtainment of PMAA is employed for the determination of monosaccharide ring size and attachment points. In this process, the sample undergoes methylation at its free hydroxy groups, followed by hydrolysis, reduction with $NaBD_4$, and acetylation. The resulting PMAA are then analyzed using GC-MS. The presence of acetyl groups in the fragments indicates the attachment point or the position of cyclization in the pyranose or furanose cycle, while methyl groups signify free positions not involved in linkages. While this methodology is effective for detecting neutral and basic sugars, it necessitates an additional reduction step for the carboxymethyl group of fully methylated polysaccharides if acidic sugars need to be detected (Figure 2.5).

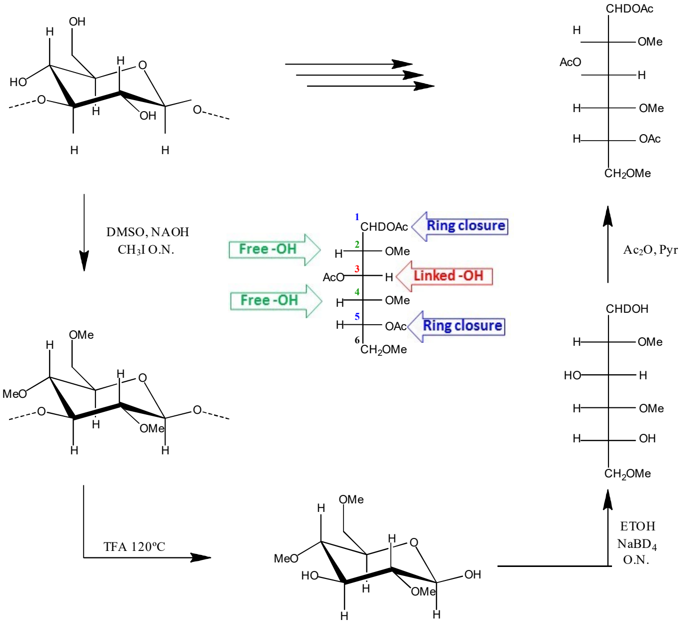


**Figure 2.5:** Scheme of the reactions involved in converting a polysaccharide into its corresponding partially methylated acetylated alditols (adapted from ref.[130]).

Finally, acetylated octyl- (or butyl-) glycosides allow to determine the absolute configuration, using an enantiomerically pure alcohol, such as 2-(+)-octanol or 2-(+)-butanol. The use of this enantiomerically pure alcohol facilitates the formation of a diastereoisomeric mixture, which can be separated in the chromatographic column of the GC-MS.

The retention time of the acetyl 2-(+)-octyl glycosides is compared with that of a standard mixture of O-2-(±)-octyl glycosides representing standard monosaccharides in D or L configuration. By comparing the retention time of the octyl glycosides in the sample with those of the standard, the D- or L-configuration of the monosaccharides can be determined (Figure 2.6).

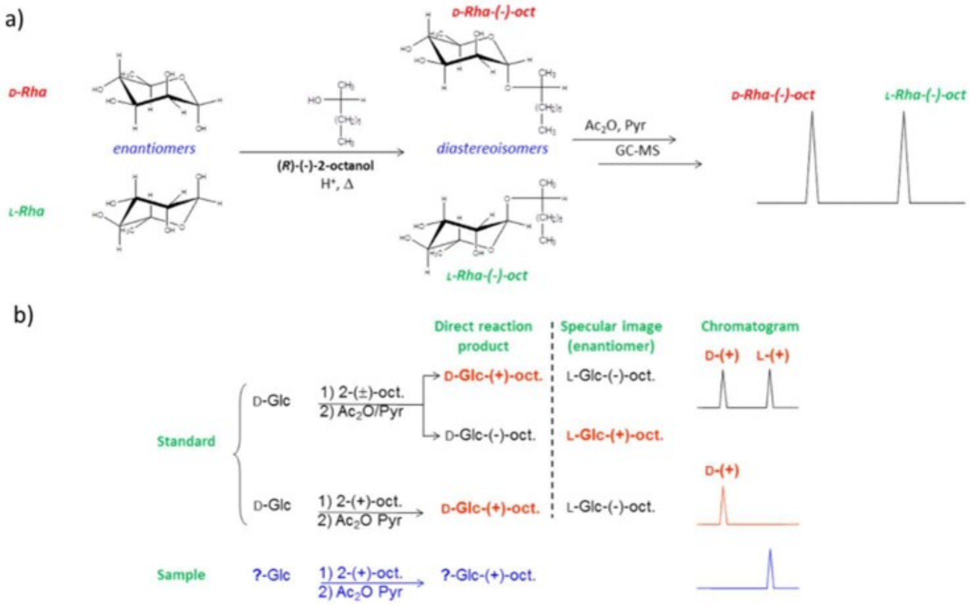


**Figure 2.6:** Acetylated octyl glycosides. a) Scheme of the enantiomeric mixture of rhamnose undergoing solvolysis using enantiomerically pure (R)-(-)-2-octanol. b) Approach employed for the synthesis and characterization of 2-octyl glycosides (adapted from ref. [130])

### 2.2.2 NMR analysis

NMR spectroscopy stands as a powerful and *non*-destructive technique widely used for investigating both synthetic and natural compounds in solution. It provides insights at atomic and molecular level by observing the behavior of atomic nuclei in a magnetic field. This technique is of paramount importance in the structural characterization of carbohydrates, offering the advantage of studying molecules in their native state.[131]

NMR spectroscopy observes the transitions between different spin states of nuclei when subjected to a static magnetic field ($B_0$). Elements with a *non*-zero quantum nuclear spin moment (I) induce a nuclear magnetic moment ($\mu$), categorizing them as NMR active nuclei. When subjected to an external magnetic field ($B_0$) the nuclear spins of the atoms align the magnetic field with S orientations ($S = 2I + 1$) known as spin states. In the case of glycans, commonly observed nuclei include $^{1}H$, $^{13}C$ and $^{31}P$, all possessing $I = ½$, indicating that the nuclear spins will orient with or against $B_0$.

The NMR spectrum is composed of (i) signals indicating the number of nuclei object of the study, (ii) signal multiplicity revealing scalar interactions between vicinal and geminal nuclei, propagated through the bond electrons and measured by the coupling constant (J), a value used to gain structural information like bond distances in the molecule, (iii) chemical shift (δ) indicating frequency of resonance, and (iv) integral namely the number of equivalent nuclei for each signal.

Structural analysis by NMR is accomplished by employing a combination of homo- and heteronuclear mono/bi-dimensional NMR experiments. Specifically, $^1H$ and $^{13}C$ NMR experiments are pivotal in determining the number and nature of monosaccharides, α or β anomeric configuration, ring size and the nature of *non*-carbohydrate substituents. Distinctive chemical shift regions, characteristic of sugar units, can be obtained from these experiments (Table 3). Moreover, typical values of coupling constants $^3J_{H1,H2}$ and $^1J_{C1,H1}$ discriminate between the anomeric configuration of the sugar units. Indeed, for pyranose rings with *gluco* or *galacto* configuration (axial H2), a $^3J_{H1,H2}$ above 8 Hz is indicative of a β-configuration while below 3 Hz is indicative of an α-configured anomer. Instead, *manno* configured sugars (equatorial H2) are characterized by $^3J_{H1,H2}$ below 3 Hz. Furthermore, a value of $^1J_{C1,H1}$ below 165 Hz corresponds to a β-anomer whereas for an α-anomeric configuration the value is above 170 Hz.

**Table 3:** 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 sugar |
| 2.0-1.0 | Methyl protons of the 6-deoxy sugars and the acetyl groups |
| **δ (ppm)** | **$^{13}C$** |
| 180-160 | Carbonyl carbons |

| | |
|---|---|
| 105-95 | Anomeric carbons |
| 80-60 | Sugar ring carbons |
| 60-45 | Nitrogen-bearing carbon signals |
| ~30 | Aliphatic methylene carbons of deoxy sugars |
| 20-17 | Methyl carbons of deoxy sugars, acetyl group |

To determine a sugar sequence, a set of complementary bidimensional homo- and hetero-nuclear NMR experiments is required. At this stage, it is fundamental to underline that NMR experiments employ two mechanisms for magnetization transfer: scalar coupling or dipolar interaction (NOE effect). Scalar coupling provides structural insights into two protons separated by two (geminal) or three (vicinal) bonds, while the dipolar interaction arises from two protons in close spatial proximity, within less than 5 Å. In detail, homonuclear COSY (COrrelation SpectroscopY) and TOCSY (TOtal Correlation SpectroscopY) measure the correlation of $^{1}$H nuclei with geminal and vicinal couplings and those belonging to the entire spin network, respectively. NOESY (Nuclear Overhauser Effect SpectroscopY) measures the cross-relaxation rates of spins, disclosing the proximity of specific protons in space. The NOESY variant under spin-locked conditions, known as ROESY (Rotating-frame Overhauser Effect SpectroscopY), always shows a positive signal and is commonly employed when the NOE is close to zero, such as in small oligosaccharides, including many trisaccharides.

On the other hand, HSQC (Heteronuclear Single Quantum Correlation) directly correlates $^{13}$C and $^{1}$H, while HMBC (Heteronuclear Multiple Bond Correlation) employs multiple-bond couplings over two or three bonds (J=2-15 Hz) to ascertain long-range $^{1}$H-$^{13}$C connectivity. All together these techniques allow the depiction of the ligand structure (Figure 2.7).

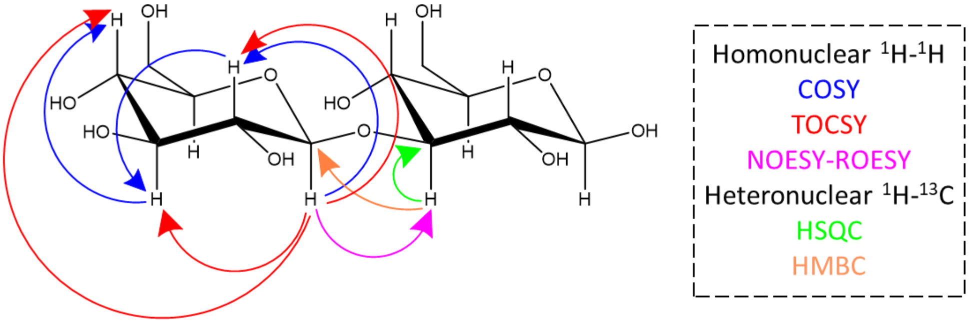


**Figure 2.7:** Representation of the correlations observable by different 2D NMR experiments.

## 2.3 Methodologies for the expression of C-type lectins

C-type lectins are usually expressed in *E. coli.* The sequence of the protein of interest is firstly cloned into a plasmid, such as pET-30(a)(+) (Figure 2.8).

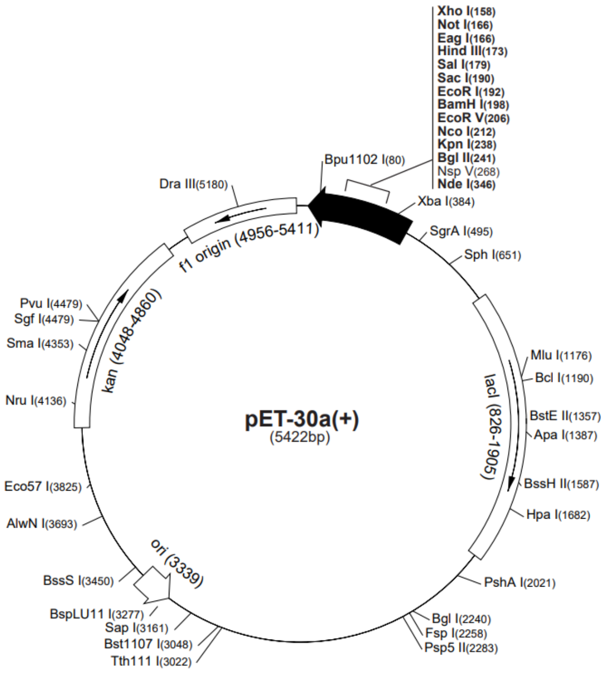


**Figure 2.8:** Restriction map of the plasmid pET-30(a)(+).

Competent cells, for example, *E.* BL21(DE3), are then transformed with the plasmid, using the protocol suggested by the provider. *E. coli* BL21(DE3) serve for the expression of *non*-toxic heterologous genes and contain the lambda DE3 prophage that carries the gene for T7 RNA polymerase under the control of a lacUV5 promoter, allowing expression of the T7 RNA polymerase to be induced with Isopropyl β-d-1-thiogalactopyranoside (IPTG). BL21(DE3), although it is an *E. coli* strain, does not contain the Lon protease and the outer membrane protease OmpT, thus reducing the degradation of heterologous proteins expressed in these cells (Figure 2.9).

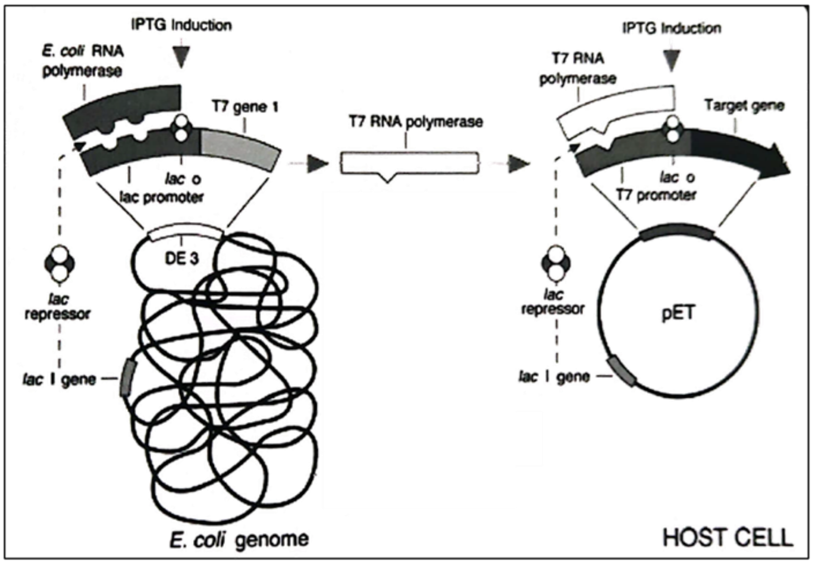


**Figure 2.9:** Scheme of *E. coli* BL21(DE3) competent cells.

Transformed cells grow in a certain medium, like Luria-Bertani broth (LB), seed in Agar-LB plate with the corresponding antibiotic, in this case, kanamycin and incubated. The presence of the antibiotic allows the selection of the cells with the plasmid integrated. At this point, one single colony of transformed cells is selected from the plate and inoculated for growth. When the optical density at 600 nm ($OD_{600}$) is 0.6, the culture is induced with 1 mM of IPTG, and the growth is prolonged. When the final $OD_{600}$ is around 4-5, cells are harvested by centrifugation and once collected, resuspended in buffer with EDTA-free protease inhibitor.

Note that for the expression and purification of proteins, different methods can be used, for example, they can be expressed as fusion proteins or in inclusion bodies. In the first case, once the obtained transformed cells are lysed, the protein of interest can be recovered after centrifugation in the supernatant which contains soluble proteins, and, then, purified; otherwise, if the expression is in inclusion bodies, some additional steps are required. Inclusion bodies are cytoplasmic granules, composed of insoluble aggregates of the expressed protein. In this case, cells collected by centrifugation, are lysed by mechanical methods, such as sonication or lysozyme and detergents. The inclusion bodies, collected by centrifugation, are washed (Triton X-100 + EDTA or urea) to separate as much as possible soluble proteins from aggregated recombinant proteins. In most cases, isolated inclusion bodies contain more than 90% pure recombinant protein. The washed inclusion bodies are solubilized, and the recombinant protein is subjected to a refolding procedure. Various conditions can be used to solubilize the inclusion bodies: guanidine-HCl, urea, basic pH, SDS, acetonitrile/propanol.[132] Renaturation procedures vary by protein, but, in general, consist in the gradual removal of the denaturing agent, like for the drop-by-drop dilution. After dialysis, centrifugation and filtration, the protein can be purified.

The purification of the recombinant protein can consist of different steps of chromatography. For example, an affinity chromatography can be employed to selectively bind the protein of interest and then elute it in the appropriate buffer; or another step can be a SEC to further purify the protein according to its MW. Once purified, the protein is controlled on an SDS-PAGE using Coomassie staining, filtered and quantified. In Figure 2.10 a summary of all the steps necessary for the expression and purification of a recombinant protein produced as inclusion bodies is reported.

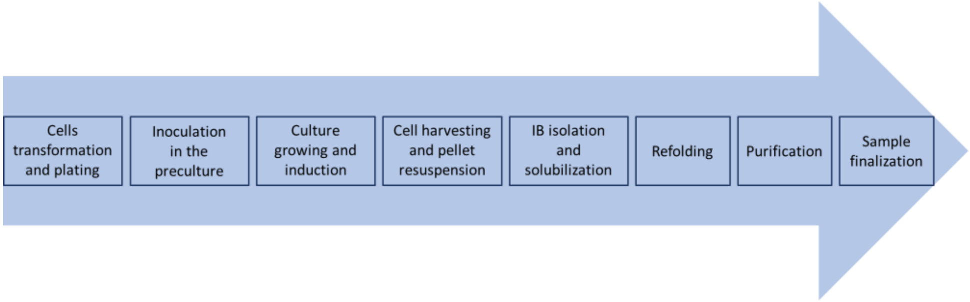


**Figure 2.10:** Scheme of all the steps necessary to produce a recombinant protein as inclusion bodies.

Once obtained the chosen protein and characterized the glycoconjugate of interest, the study of their molecular binding can be investigated using different techniques, as explained in the next paragraphs.

## 2.4 Biophysical techniques to unveil protein-glycan molecular interactions

Investigating the molecular basis of protein-ligand interactions is a prerequisite for structure-based drug design, facilitating the development of innovative therapeutic approaches and strategies. A myriad of techniques, including biophysical, spectroscopic, spectrometric, and computational methods, is available for a full comprehension of protein-ligand interactions. Among them, NMR techniques, fluorescence spectroscopy and surface plasmon resonance (SPR) allow to achieve crucial molecular details.

### 2.4.1 NMR techniques for the analysis of protein-glycan complexes

NMR spectroscopy is a very useful tool for gaining information regarding protein-ligand interactions at atomic resolution. Several approaches are available for investigating protein-ligand interactions using solution NMR spectroscopy. These methods can be broadly categorized into *protein-based* and *ligand-based* approaches. In a protein-based method, a spectrum of the protein is acquired, and the ligand is subsequently titrated. This approach provides insights into the specific residues of the protein that directly interact with the ligand. Conversely, in ligand-based methods, a spectrum of the ligand is acquired, followed by the addition of the protein. The ligand can range from small molecules, like chemical compounds or peptides, to larger macromolecules, like DNA or other interacting proteins. In this thesis, carbohydrate molecules are used as ligands.

The equilibrium associated with the formation of a protein (A)–ligand (B) complex is given by the following reaction $A + B \rightleftharpoons AB$, and it is characterized by the equilibrium dissociation constant ($K_D$, measured in $M^{-1}s^{-1}$) that, in case of 1:1 stoichiometry, is:[133]

$$K_D = \frac{[A][B]}{[AB]} = \frac{k_{off}}{k_{on}} \qquad \textbf{Equation 2.1}$$

where [A], [B], and [AB] are equilibration concentrations of the protein A and the ligand B and the complex protein-ligand AB, respectively; $k_{off}$ is the dissociation rate constant and $k_{on}$ is the association rate constant which determines the magnitude of $K_D$.

The lifetime of the protein–ligand complex, determined by $1/k_{off}$, is responsible for the overall appearance of the NMR spectra. Two limiting cases can be delineated: (i) a *slow chemical exchange* between free and protein-bound states of the ligand, where the lifetime of the protein-ligand complex is much longer than the difference in chemical shifts between two signals observed for free and bound states, resulting in the presence of two NMR signals, typically found in strong complexes; (ii) a *fast chemical exchange* between free and bound state, generating the collapse of the two signals into a single peak, due to the short lifetime of the complex (Figure 2.11).[134] Some protein-ligand interaction NMR techniques are useful only for strongly interacting molecules in slow exchange, while others operate exclusively in the fast exchange regime.

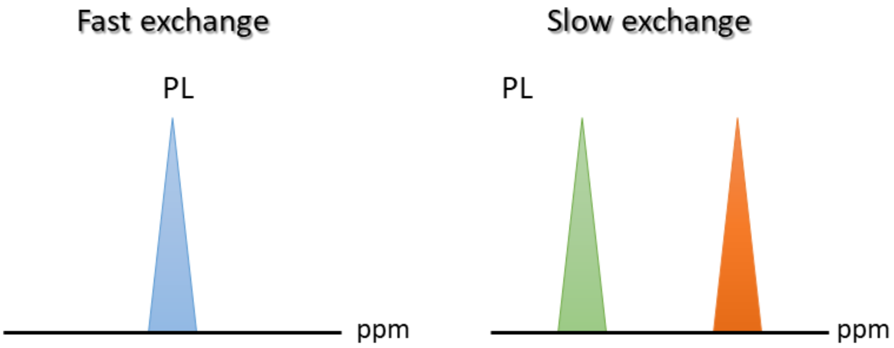


**Figure 2.11:** Illustration of fast and slow exchange.

In the next subparagraphs, two ligand-based NMR methods will be discussed more in detail: transferred NOESY (tr-NOESY) and saturation transfer difference (STD NMR). Both are based on the transfer of NMR parameters from the protein-bound state to the ligand-free state and are particularly useful in the medium–low affinity range, with $K_D \geq 100$ μM. Furthermore, these NMR approaches work in a "fast-exchange" regime, with the dissociation rate constant $k_{off}$ in the range of $1000<k_{off}<100000$ $s^{-1}$.[135, 136] Additionally, some details regarding chemical shift perturbation, a protein-based NMR method, will be also provided.

### 2.4.1.1 tr-NOESY

tr-NOESY experiments are based on the different behavior of a ligand in free and bound states, giving information about conformational changes, upon binding. These experiments are based on the Nuclear Overhauser Effect (NOE), used for measuring homonuclear dipolar interactions that can be detected up to 5 Å.[137] It arises from direct dipolar cross-relaxation between neighboring nuclear spins, exhibiting a sharp decline with increasing distance. Sign and magnitude of the NOE are influenced by the rotational correlation time ($\tau_c$), which, in turn, depends on the size of the ligand. For small molecules (like ligands up to 2 kDa MW), $^1$H–$^1$H NOEs are small and positive, whereas for larger molecules (like proteins), they become larger and negative. The detection of transferred-NOEs (tr-NOEs) is contingent on the distinct tumbling times ($\tau_c$) of the ligand in the free and bound states. Indeed, it reflects the binding to the receptor and gives information regarding the conformational behavior of the ligand. In detail, when binding to a macromolecular receptor, the ligand assimilates into the macromolecule, exhibiting the corresponding NOE behavior. The ligand, undergoing free exchange between the bound and free states, preserves the NMR characteristics of the protein receptor. Due to the slower relaxation in the free state, the spectrum provides information about the ligand in the bound state.[138] Thus, the analysis of NOE-derived *intra*-molecular $^1$H–$^1$H distances in both free and bound states enables the monitoring of conformational alterations in the ligand upon binding, allowing the determination of the bioactive conformation, namely, the conformation adopted by the ligand in the bound state (Figure 2.12).[139]

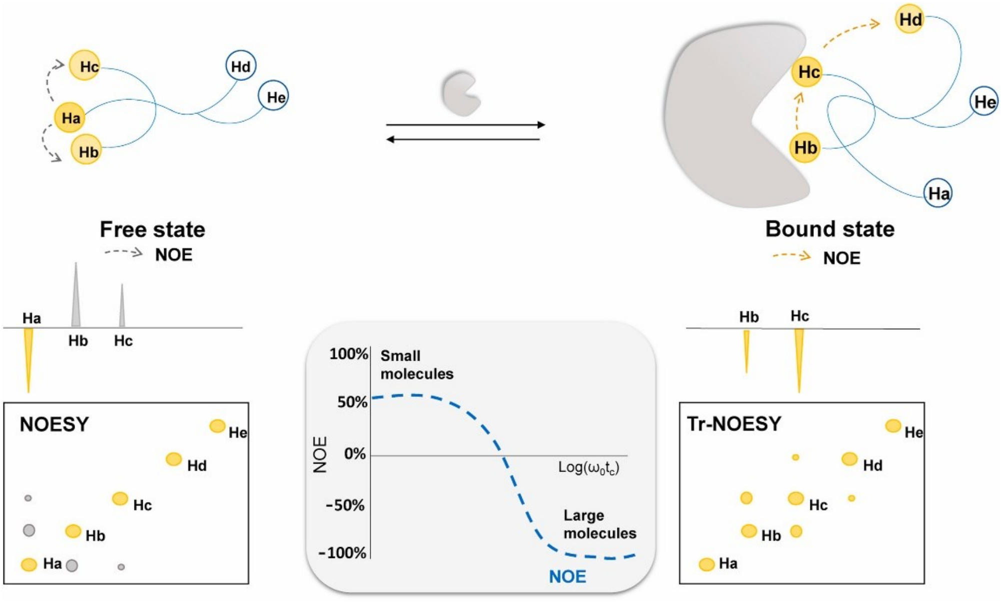


**Figure 2.12:** NOE effect schematic representation. In the free state, the ligand exhibits positive NOE with cross-peaks with opposite signs to the diagonal peaks of the NOESY. In the bound state, the ligand adopts a negative NOE behaving as the macromolecule as shown in the tr-NOESY spectrum (adapted from ref.[138]**)**.

The construction of NOE build-up curves allows to specifically determine *intra*-molecular $^1$H–$^1$H distances. The highest enhancement in tr-NOE experiments occurs at shorter mixing times, typically ranging from 50 to 100 ms, whereas it is four to ten times longer for the free ligand.[136] Consequently, when integrating the NOEs at various mixing times, the build-up curves are fitted to a double exponential function:

$$\boldsymbol{f = a\,(e^{-ct})(1 - e^{-bt})} \qquad \textbf{Equation 2.2}$$

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:

$$\boldsymbol{f'(0) = a * b} \qquad \textbf{Equation 2.3}$$

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

$$\boldsymbol{r_{ij} = r_{ref}\sqrt[6]{\frac{\sigma_{ref}}{\sigma_{ij}}}} \qquad \textbf{Equation 2.4}$$

where $r_{ij}$ is the distance to calculate, $r_{ref}$ is the reference distance, $\sigma_{ref}$ is the cross-relaxation rate and $\sigma_{ij}$ is the cross-relaxation time that gives the desired distances.[136]

In conclusion, tr-NOESY experiments serve as a valuable tool for elucidating dynamic processes and bioactive conformation, with possible applications in rational structure-based drug design.

#### 2.4.1.2 STD NMR

STD NMR is a ligand-based method able to detect molecular interactions with $K_d$ values in a range of $10^{-8}$ –$10^{-3}$ mol $L^{-1}$. This technique possesses an easy implementation and allows to (i) determine the interaction between small ligands and their receptors, (ii) analyze the binding processes at a molecular level and (iii) derive the epitope maps namely the regions of the ligands directly involved in the binding processes. STD NMR experiments are based on intermolecular NOE,[140] by transferring the magnetization from the protein to the ligand protons. By selectively saturating protein resonances using a specific radio frequency-pulse train (saturation time in the range of seconds), the magnetization spreads rapidly throughout the entire protein via spin diffusion and intramolecular NOEs. This saturation is then transferred from the protein to the interacting ligand during its residence time in the protein binding site, through intermolecular saturation transfer and chemical exchange. Ligand protons in close proximity to the protein binding site exhibit higher magnetization, while those farther away receive little or no saturation (Figure 2.13).

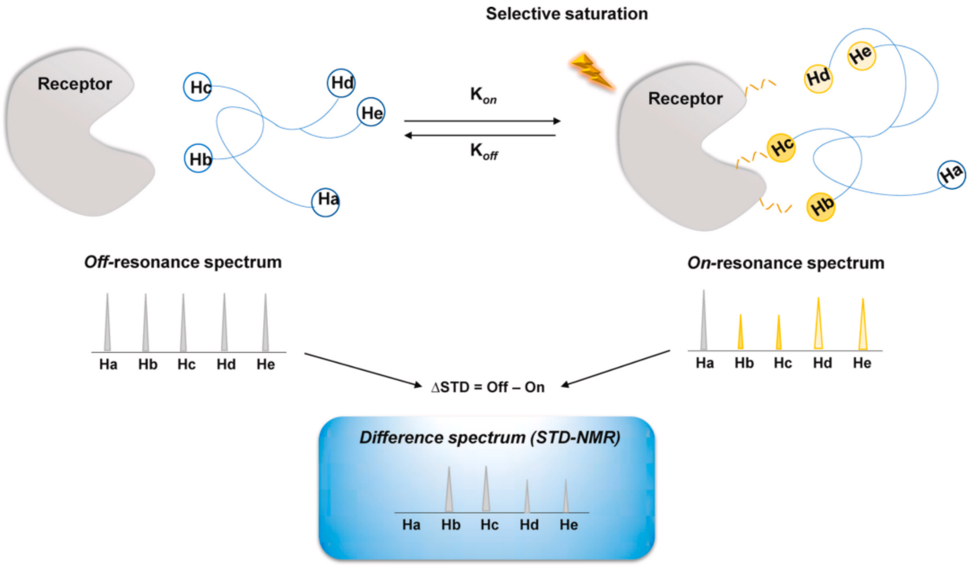


**Figure 2.13:** STD NMR method schematic representation. In the *off*-resonance spectrum, ligand protons do not show an intensity decrease, whereas in the *on*-resonance spectrum, the saturation is transferred to the ligand after irradiating the protein with RF pulses. The final STD spectrum is the result of the subtraction of the off- and on-resonance spectra, showing only the ligand signals involved in the binding process (adapted from ref.[138]).

From the experimental point of view, the STD experiment is based on the subtraction of two 1D NMR spectra: (i) the *off-resonance* spectrum, representing the reference, in which the irradiation occurs usually at 40ppm, which is a region far from ligand and protein signals and (ii) the *on-resonance* spectrum, in which only protein signals are selectively saturated, usually in the aliphatic or aromatic regions (0 to -2 ppm and 6.5-8.5 ppm, respectively). The signal is subsequently transferred from the saturated protein to the bound ligand via intermolecular $^1H$–$^1H$ cross-relaxation pathways. Once in solution, the ligand maintains the saturated state due to its enthalpic relaxation ($R_1$), which is slower than the $k_{off}$. Consequently, an accumulation of the saturated free ligand in solution leads to a reduction of the signal intensity, because of the acquired magnetization. The resulting STD spectrum comes out from the subtraction of the *off-* and *on-resonance* experiments, containing only signals of bound ligand protons. The STD intensities ($I_{STD}$) are calculated as follows:

$$\boldsymbol{I_{STD} = \frac{I_0 - I_{sat}}{I_0}} \qquad \textbf{Equation 2.5}$$

where $I_0$ is the intensity of a signal in the *off*-resonance experiment, $I_{sat}$ is the intensity of a signal in the *on*-resonance experiment. Since the degree of saturation varies among ligand protons, based on their proximity to the binding pocket,[141] it is possible to delineate the ligand's interacting epitope. Additionally, STD% values can be obtained by normalizing the STD intensities, against the most intense signal: a higher STD% value indicates greater proximity of the ligand proton to the receptor binding pocket. However, STD intensities can vary depending on the length of the saturation or due to the different relaxation times of the protons of the ligand. To solve these issues, a commonly used approach is the initial growth rates method,[142] in which these rates are derived from STD intensities at the point of zero saturation time, where neither ligand re-binding nor relaxation occurs. STD NMR spectra can be acquired at different saturation times in the range of 0.5-5 s, to obtain the STD build-up curves. Particularly, STD build-up curves can be calculated through the STD amplification factor ($A_{STD}$), which is obtained by multiplying the relative STD effect ($I_{STD}/I_0$) at a given ligand concentration ($[L]_0$) by the molar ratio between in-excess ligand and the protein ($[L]_0/[P]_0$), where $[L]_0$ and $[P]_0$ refer to the total concentration of ligand and protein, respectively:

$$\boldsymbol{A_{STD} = \frac{I_0 - I_{sat}}{I_0} * \frac{[L]_0}{[P]_0} = \frac{I_{STD}}{I_0} * \frac{[L]_0}{[P]_0}} \qquad \textbf{Equation 2.6}$$

$A_{STD}$ is calculated for each proton involved in the interaction, for each saturation time and then data are fitted in the following monoexponential function:

$$\boldsymbol{STD\ (t_{sat}) = STD_{max} * \left(1 - e^{(-k^{sat} * t^{sat}}\right)} \qquad \textbf{Equation 2.7}$$

where STD ($t_{sat}$) 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 $STD_{fit}$ indicates the slope of the STD build-up curves at a saturation time of 0 and it is influenced by the proximity of the ligand to the protein. To generate the ligand's epitope map, the $STD_{fit}$ values are normalized relatively to the highest value, which is set to 100%.[138]

In conclusion, STD NMR provides a deep characterization of the interactions occurring between ligands and proteins, allowing the determination of the epitope map of the ligand. Further applications also implicate the use of 2D STD experiments, like 2D-STD-TOCSY and 2D-STD-HSQC, that can solve cases in which signals overlapping are present, by extracting the corresponding rows through the 2D spectra.[143]

### 2.4.1.3 Chemical shift perturbation

Protein-based NMR methods offer a direct means of observing receptor signals, enabling the characterization of interactions with specific ligands and the identification of unknown receptor binding sites.[144] To employ these NMR experiments, the expression of an isotope-labeled protein, such as with $^{15}N$, is necessary. Subsequently, a preliminary assignment of the protein NMR spectra through specific 3D NMR experiments is essential to decipher the protein backbone and its later chains.[145]

Upon the formation of protein–ligand interactions, various physical parameters of both the protein and the ligand undergo changes. Differences in hydrophobicity at the interaction surface, for example, lead to alterations in local electron density. These changes influence the chemical shift, one of the most easily observable NMR parameters. Large changes in chemical shifts are also induced by the spatial proximity of groups with magnetic susceptibility anisotropies, such as aromatic rings. Importantly, the chemical shift is affected not only by the covalent molecular structure of a protein but also by *non*-covalent interactions with ligands and solvent molecules. A crucial protein-based NMR method is the analysis of chemical shift perturbation (CSP) or chemical shift mapping (CSM), employed to map the chemical shift of a

$^{15}$N-labeled protein when titrated with a ligand.[146, 147] CSP analysis of a protein/ligand complex requires the acquisition of the $^{1}$H-$^{15}$N-HSQC spectra for the apoprotein as a reference. Sequential $^{1}$H-$^{15}$N-HSQC spectra are then performed with increasing ligand concentrations, ideally until the protein binding site is saturated. Binding is visualized by overlaying all HSQCs recorded during the titration. In the presence of an interaction, the chemical shifts of the residues involved in complex formation with the ligand, observed as peaks in a $^{15}$N-HSQC, shift from their original positions. Two limiting cases are encountered: (i) in *fast exchange*, the chemical shift represents the population-averaged value between the free and bound forms, with peaks moving linearly by adding amounts of ligand; (ii) in *slow exchange*, both signals of the bound and free states are observed, with signal integrals representing their relative amounts (Figure 2.14).

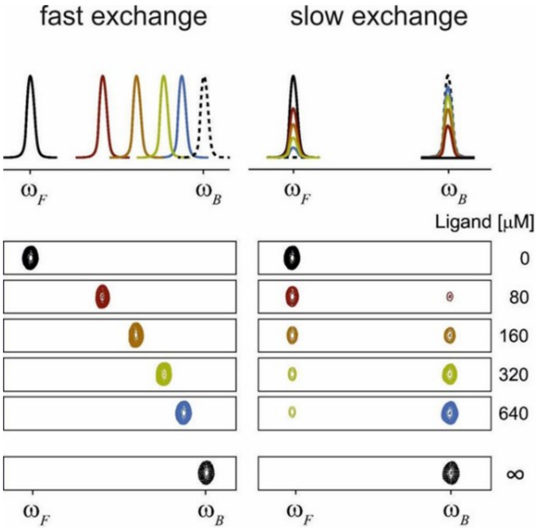


**Figure 2.14:** Changes in protein peaks upon titration with a ligand are shown schematically for fast and slow exchange. For binding in fast exchange, the signal of the free-protein peak at $\omega_F$ is moving towards the fully ligand-saturated protein peak $\omega_B$. In slow exchange, only the relative signal intensities of free and bound protein peaks change (adapted from ref.[134]).

Chemical shift perturbation analysis requires that the protein and the ligand be dissolved in the same buffer and the measurements during the titration must be acquired under the same conditions. This is crucial as chemical shifts are highly sensitive to differences in temperature, pH value and, buffer composition, especially those of amide protons. Additionally, a limitation of protein-based NMR is the size of the macromolecule which should not exceed 50 kDa due to the negative impact on relaxation time, resulting in low-quality $^1H$-$^{15}N$-HSQC spectra. TROSY (Transverse Relaxation Optimized SpectroscopY) can be a solution, using triple resonance experiments to select a single component of different relaxation $T_2$ mechanisms, leading to a single and sharp peak in the spectrum.[148]

While most peaks affected by a change in chemical shift determine the binding site of the protein accommodating the ligand, conformational changes in amino acids also result in differences in resonance frequencies. Thus, the shifting of a signal is not always indicative of proximity to the binding interface but can provide information about allosteric changes in the protein structure when a ligand is bound.

### 2.4.2 Fluorescence spectroscopy

Fluorescence spectroscopy is a quantitative analytical technique, widely employed to offer insights into the structural and dynamic aspects of glycan-protein complexes. Its benefits encompass high sensitivity, ease of application, and a diverse selection of fluorescence labels and probes.

Fluorescence originates from electronic transitions within molecules, triggered by the absorption of a photon that induces the formation of an excited singlet state. This excited state emits light as fluorescence, allowing the molecule to revert to the ground state. Nevertheless, *non*-radiative processes like vibrational loss (heat), the creation of a triplet state, and collisions with a quencher (such as iodide or oxygen) may interfere with fluorescence, deactivating the excited state and reducing its effectiveness. The emission rate of most fluorophores typically stands at 108 s−1, resulting in a typical fluorescence lifetime of ~10 ns. This timeframe aligns with conformational transitions, making fluorescence spectroscopy well-suited for detecting conformational changes occurring during ligand binding. Fluorescence spectroscopy is useful as long as ligand binding induces a change in the fluorophore, which may be an intrinsic part of a protein (e.g., aromatic amino acids (Figure 2.15)) or may be introduced extrinsically (e.g., a *non*-covalent probe).

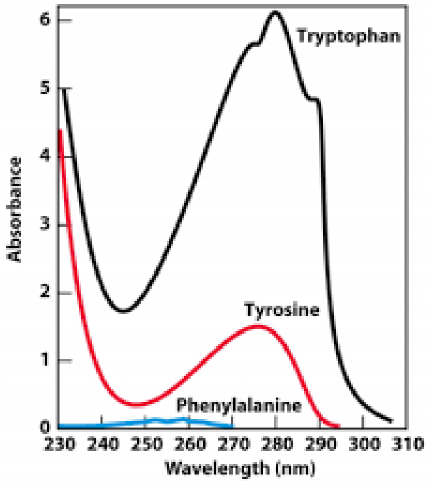


**Figure 2.15:**UV-vis absorption spectra of aromatic amino acids.

In most of the studies related to glycan-protein interactions, changes in fluorescence intensity of aromatic amino acids are monitored, using a fixed concentration of the protein and following the ligand binding through its titration, until the protein is saturated, obtaining the $K_D$ (**Equation 2.1**).

The binding curve is obtained by plotting $\Delta F/F_0$ values versus ligand concentration. Data analysis is performed using the following equation:

$$\frac{\Delta I_f}{I_0} = \frac{\Delta I_{max}}{I_0} X_{FY} \qquad \textbf{Equation 2.8}$$

where $\Delta I_f$ is the fluorescence intensity change upon addition of the ligand, $\Delta I_{max}$ is the maximal fluorescence intensity change, F is the fluorophore, X is the interferent species, and $X_{FY} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, with a = $[F]_t K_b$, b = $1+[Y]_t K_b$ and c = $[Y]_t K_b$.

For the evaluation of the binding interactions, in particular, the $K_D$, data can be analyzed by non-linear regression equation considering one site-specific binding model:

$$Y = \frac{B_{max} * X}{K_D + X} \qquad \textbf{Equation 2.9}$$

where Y is the fluorescence intensity change at the maximum wavelength, X is the ligand concentration, and $B_{max}$ is the maximum specific binding.[149]

### 2.4.3 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) is a valuable technique for studying protein-glycan interactions. This versatile method can be applied in various contexts, allowing for the measurement of binding kinetics, affinity, specificity, and the concentration of biologically active compounds.

The SPR *detection system* operates based on the principle of total internal reflection (TIR) and the generation of surface plasmon. Essentially, when light passes through a surface with a different refractive index, TIR occurs if the refractive index is lower on the other side of the boundary and the incident angle is greater than the critical angle. This leads to the complete reflection of the wave. Thus, SPR is observed under TIR conditions of light at the interface of two media with different refractive indexes, coated with a metal film (gold layer interface). Plasmons can be described as oscillations of free electron densities with respect to the fixed positive ions in a metal. Surface plasmon waves are excited at the gold interface and SPR is observed as a dip in the intensity of reflected light at a specific angle of incidence. Binding events impact the refractive index (RI) close to the metal surface and are monitored in real-time as changes in the specific angle where SPR induces the extinction of light ($\Delta\theta$, resonance angle shift). There is a direct relationship to quantify the binding which can slightly vary according to the used compounds; in general, it is 1pg/mm$^2$ surface concentration change – $10^{-6}$ RI change – $10^{-4}$ deg deviation – 1 Resonance Unit (RU).

In the studies carried out in this thesis, the used *sensor chips* (containing the nanoscale film to functionalize to study the interactions) are composed of dextran polymers, with the advantages of hydrophilicity, easy use for covalent coupling, high binding capacity, flexibility, and high chemical resistance. To immobilize the ligand of interest, there are different types of dextran-activated surfaces, such as CM5, a dextran matrix covered with many carboxyl groups, and CM3 in which the interaction takes place closer to the surface to improve sensitivity (Figure 2.16).

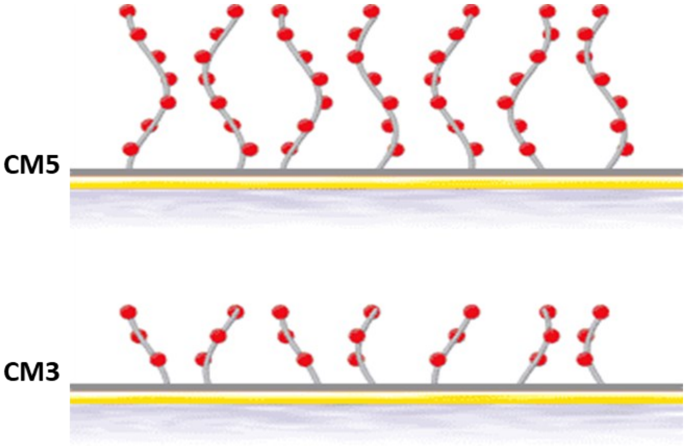


**Figure 2.16:** Examples of diverse dextran activated surfaces.

There are two possible ways for the immobilization of the *ligand* on the surface: (i) in the *direct* immobilization the ligand is directly functionalized on the surface by covalent coupling, (ii) in the *indirect* immobilization, a capturing molecule (es. streptavidine), attached on the surface by covalent coupling, catches the ligand (es. linked to the biotin); instead, the *analyte* is injected free in solution. Among the different coupling chemistries, the activation of the carboxylic groups of the dextran matrix mostly occur by amine coupling, using the mixture of two reactants 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC) and N-hydroxy-succinimide (NHS), forming amine-reactive intermediates that spontaneously reacts with primary amines to form amide bonds. On the sensorchip, there is a microfluidic system with four channels called flow cells (FC), for individual or serial use, meaning that with the same amount of analyte, it could be possible to analyze four different ligands consecutively or, also, to use one of the FC as reference.

The result of an SPR analysis is the sensorgram, reporting RU on the Y-axis and time (s) on the X-axis, giving information regarding association and dissociation of the complex of interest. In particular, from the sensorgram, it is possible to calculate (i) for the *affinity* analysis, how strong is the binding at equilibrium, by quantifying the $K_D$, (ii) for the *kinetic* analysis, how fast is the binding, through $k_{on}$ (measuring the recognition) and $k_{off}$ (measuring the stability) (**Equation 2.1**).

A scheme of SPR method is shown in Figure 2.17. In conclusion, SPR is a versatile and powerful analytical tool with a wide range of applications. It excels in determining small molecules with exceptional accuracy, reaching concentrations as low as $10^{-13}$ – $10^{-16}$ M. Its diverse applications include not only the study of different molecular interactions such as glycan-protein, antigen-antibody, DNA-protein, RNA-protein, DNA-DNA, RNA-RNA, along with the kinetics of these reactions, but also the quantitative and qualitative assessment of toxins, monitoring gene expression. The method has numerous advantages, including high sensitivity and selectivity, real-time kinetics analysis, simplicity in sample processing, short analysis time, and the ability to reuse the sensor after proper regeneration. Overall, SPR stands out as a valuable and efficient technique in the realm of molecular analysis.[150]

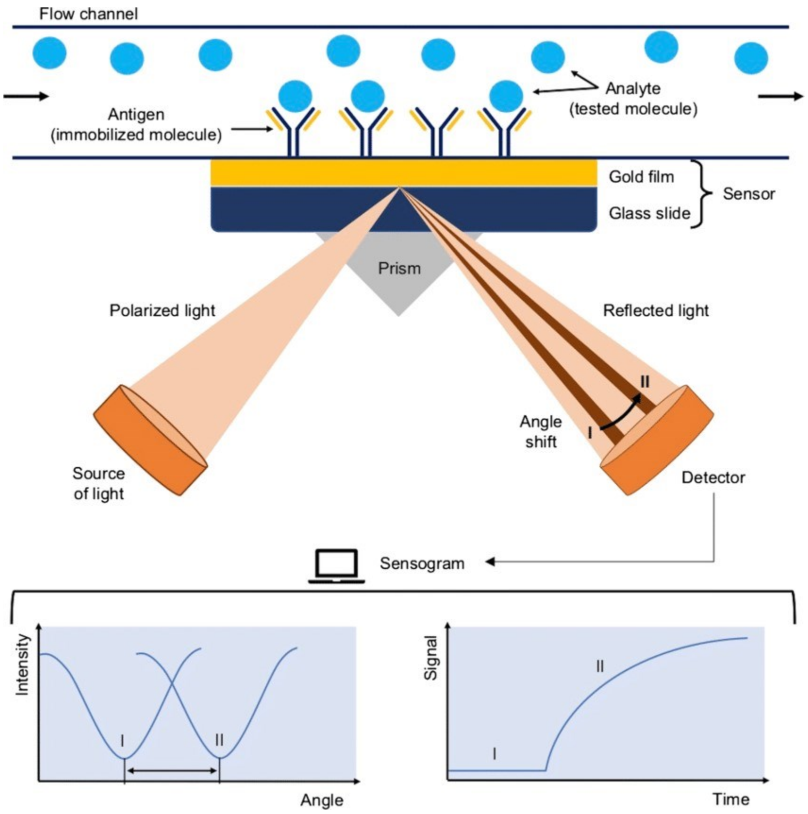


**Figure 2.17:** Scheme of operation of SPR (extracted from [150]).

## 2.5 Computational studies

Computational studies in glycoscience have witnessed significant advancements, playing a crucial role in unraveling the structural and conformational intricacies of carbohydrate molecules, in their free state or bound to proteins. Molecular modeling techniques combined with analytical and biophysical techniques, allow to provide structural and dynamical features of 3D complexes [151, 152]

Computational techniques range from ab initio to coarse-grained (CG) methods and from deterministic to heuristic approaches. In detail, they include (i) *DFT-based ab initio simulations*, (ii) *quantum mechanics/molecular mechanics (QM/MM)* and *QM/QM hybrid methods*, (iii) *semi-empirical methods*, (iv) *molecular mechanics*

*(MM) and molecular dynamics (MD) simulations*, (v) *heuristic methods* (Monte Carlo and genetic algorithms), (vi) *coarse-grained methods*, (vii) *docking calculations*.[153]

Importantly, carbohydrates are subject to stereo electronic effects, which are integrated into force field-based protocols. Significant advancements in force field development have led to high-quality force fields that can faithfully reproduce the dynamic and electronic properties of carbohydrates.[154] Additionally, when dealing with bacterial glycans, another challenge arises due to the absence of a wide library of parametrized structures. Consequently, it becomes necessary to utilize QM to parameterize the sugar units before being able to work with them.

Subsequent paragraphs will provide an overview of docking, MM and MD simulations, as these computational tools play a central role in this thesis.

### 2.5.1 Docking

In the realm of molecular modeling, docking stands out as a widely employed method for structure-based drug design. Its utility lies in predicting the binding conformation of small molecules within specific protein binding sites, offering a model of the interaction. Docking protocols generally start with conformational sampling of the ligand, consisting in finding potential binding poses for a small molecule within a given receptor, by adjusting their conformation to achieve the best fit. The subsequent step is the clustering, entailing energy scoring of the resulting binding poses and evaluating the energy of the ligand-target complex. Although these steps are common in various docking programs,[155] differences arise in the algorithm for computational search and the nature of the scoring function applied to rank the docked poses.[156, 157]

The conformational ligand search can be carried out using three main types of algorithms: (i) *shape matching*, which assesses geometrical overlap, identifying potential binding sites through macromolecular surface searches,[158] (ii) *systematic search*, ideal for flexible ligand docking, exploring all degrees of freedom and (iii) *stochastic or non-deterministic algorithms*, like Monte Carlo (MC) and evolutionary programming (EP), which are the most time efficient algorithms, involving random changes in the ligand.[159]

Regarding energy evaluation, three scoring functions are recognized: (i) *force-field based*, which computes individual interaction terms such as van der Waals and electrostatic energies, stretching, bending and torsional energies,[160] (ii) *knowledge-based*, that relies on statistical analysis of interacting contacts between the ligand and the target[161, 162] and (iii) *empirical-based*, computing the collective binding free energy by aggregating a series of weighted empirical energy terms, encompassing interactions such as hydrogen bonds (H-bonds) and hydrophobic interactions.[163, 164]

For a docking calculation on a protein-ligand complex, the 3D structures of the interaction partners are required. These structures can be obtained through experimental techniques, such as NMR spectroscopy and /or X-ray crystallography, or, if the protein 3D structure is unknown, homology modeling methods can be employed to predict it. Information about the protein binding site is not mandatory but, using AutoDock,[165] as in this thesis, it can help in building the AutoGrid grid box, allowing ligand freely rotation, even in its fully extended conformation. In cases where the binding pocket is unknown, "blind docking" firstly explores the entire protein surface, in order to find regions preferred by the ligand, and then, a second round of docking calculations permits the use of a smaller grid in which ligand can move around the protein. Of course, when the protein binding site is unknown, a higher computational cost is required. Autodock program employs EP algorithms, specifically the Lamarckian Genetic Algorithm,[166] based on the genetic algorithm (GA) and local search (LS) method, which tries to find the closest conformation of the global energy minimum.[167] This hybrid method treats ligand variables, like translation, orientation and conformation, as a "gene" characterizing a "genotype" (state of the ligand), while atomic coordinates correspond to the "phenotype." Mutation and crossover define the genotype, whereas the phenotype is established by the energy function to be optimized. Individual conformations search their local minima, inheriting the information to later generations (Lamarckian aspect). Thus, the Lamarckian Genetic Algorithm combined with a semi-empirical free energy force field to score docked binding poses, generates a set of docked conformations with intermolecular and internal energy components:

$$\boldsymbol{\Delta G} = \left(\boldsymbol{V_{bound}^{L-L}} - \boldsymbol{V_{unbound}^{L-L}}\right) + \left(\boldsymbol{V_{bound}^{P-P}} - \boldsymbol{V_{unbound}^{P-P}}\right) + (\boldsymbol{V_{bound}^{P-L}} - \boldsymbol{V_{unbound}^{P-L}} + \boldsymbol{\Delta S_{conf}}) \qquad \textbf{Equation 2.10}$$

where P is the protein, L is the ligand, V represents each potential energy contribution, given by the sum of van der Walls, H-bonds, electrostatic and solvation terms, and $\Delta S_{conf}$ is the loss of conformational entropy upon binding. The conformation with the lowest docked energy is considered the best docking result.

### 2.5.2 Molecular Mechanics and Molecular Dynamics simulations

MM and MD simulations play a crucial role in predicting the structure and energy of molecules based on nuclear motions. At atomic level, particles behavior is accurately described by the principles of quantum chemistry; however, this level of theory requires high computational costs especially for big size macromolecules. Molecular Mechanics (MM) or classical mechanics describes molecules as an assembly of hard, impenetrable spheres known as atoms, each possessing a characteristic mass and size (proportional to the van der Waals radius). The positions of nuclei are fixed within a defined electron distribution, interconnected by elastic bonds (stretch, bend, torsion) following Hooke's law, and characterized by *non*-bonded interactions (van der Waals and electrostatic forces). In this context, the atomic interactions can be modelled with simple parameterized functions measuring the energetics of the system known as force fields. These force fields are typically derived from either *ab initio* or semi-empirical QM calculations or by fitting experimental data. Hence, the careful selection of an appropriate force field is essential for ensuring the validity of any MM and MD simulation.

One of the prominent force fields utilized for modeling carbohydrate structures is included in AMBER,[168] encompassing a wide array of common monosaccharides and incorporating diverse glycosidic linkages. In detail, within AMBER, there exist several distinct force fields designed for specific molecular components, such as lipids (lipids21),[169] sugars (GLYCAM_06j),[170] proteins (ff14SB),[171] nucleic acids (OL15 for DNA,[172] OL3 for RNA[173]), water molecules (TIP3P)[174] and general organic molecules (gaff2).[175] Notably, GLYCAM_06j is tailored for use with both D and L enantiomers, mono- and oligosaccharides, and all glycosidic linkages. This force field stands out as the only one where the same atom type (Cg) is assigned to the anomeric carbon (C1) in both α and β anomers. This unique feature facilitates simulations involving ring-flipping and ensures equilibrium between conformers with axial and equatorial substituents at the anomeric center.

### 2.5.2.1 Molecular Mechanics calculations

The 3D arrangement of glycans is determined by factors like sugar composition and the type of the glycosidic linkage, which define the shape and conformation. Ring shapes are classified based on reference conformations, including chair (C), twist (T), boat (B), envelope (E), skew (S). The sugar conformation is predominantly characterized by glycosidic torsion angles Φ (H1-C1-O1-CX) and Ψ (C1-O1-CX-HX). In addition, for glycosidic bonds not involving an endocyclic carbon, such as in 6-linked sugars, the ω (O5-C5-C6-O6) angle has also to be considered. (Figure 2.18).

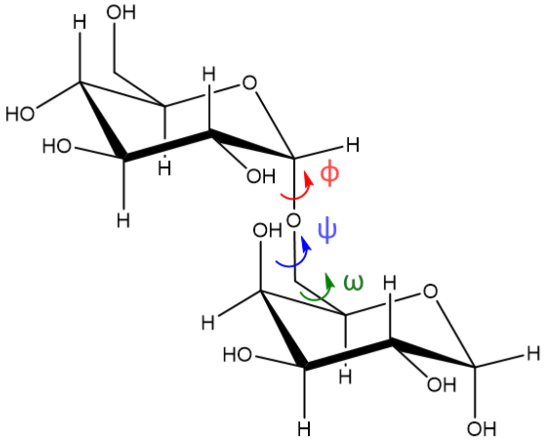


**Figure 2.18:** Representation of Φ, Ψ and ω torsion angles in a disaccharide.

The sampling of the ω torsion angle involves the populations of gauche−gauche (*gg*), gauche−trans (*gt*), and trans−gauche (*tg*) rotamers, represented in Figure 2.19.[176]

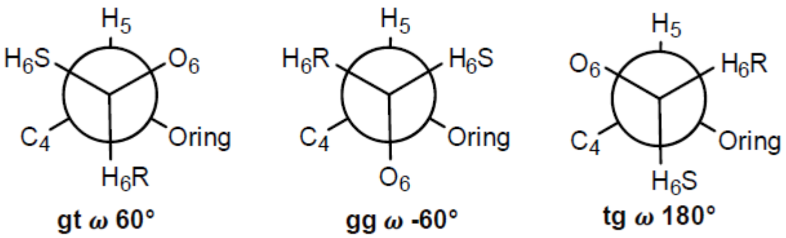


**Figure 2.19:** Newman projections of staggered rotameric states of the ω torsion angle along the C6-C5 bond: *gg* (gauche-gauche), *gt* (gauche-trans) and *tg* (trans-gauche).

Various methods can be employed to calculate Φ and Ψ, defining the energy map typically applied to each disaccharide unit in a glycan chain.[177] MM3 serves as the primary force field for determining local minima and flexibility of glycosidic torsions,[178, 179] but also free carbohydrate databases (like http://csdb.glycoscience.ru/database/) can be employed. Consequently, potential energy surfaces depicting conformational energy for the Φ and Ψ dihedral angles can be represented on "adiabatic" maps. These maps resemble energy graphs analogous to Ramachandran plots used for proteins, showing the energetically favourable conformations of a carbohydrate dimer.[180]

### 2.5.2.2 Molecular Dynamic simulations

MD simulations stand as primary tool in the computational investigation of biological molecules, offering insights into the dynamics and conformational features of molecules and their complexes. First of all, initial velocities and positions of all atoms in the biomolecular system are established. The forces acting on each atom are computed, and, in accordance with Newton's second law of motion, the spatial position of each atom is predicted as a function of time.[181] For a given atom "i" with mass "$m_i$" and Cartesian position "$x_i$", $F_{xi}$ denotes the force acting on the atom during the time "t":

$$\frac{d_{x_i}^2}{d_t 2} = \frac{F_{x_i}}{m_i} \qquad \textbf{Equation 2.11}$$

This iteration is repeated at defined time intervals and the position and velocity of each atom in each step are assessed, generating a trajectory that delineates the atomic-level configuration of the system at every time point.[182]

Classical MD simulations consists of different key steps. The initial phase is energy minimization, to find a minimum in the system's energy landscape. Two methods are commonly applied for this purpose: (i) the *steepest gradient*, in which system's geometry is optimized to attain a local energy minimum and (ii) the *conjugate method*, in which subsequent minimization steps are performed, taking information by the previous ones.[183] The optimal strategy involves incorporating the environment into the simulation, through the addition of explicit water molecules or other surrounding molecules. This approach allows the simulation of the complex within a solvent box, implementing periodic boundary conditions (PBC) to mitigate surface artifacts. This is particularly crucial for carbohydrates, given their hydrophilic nature resulting from the abundant presence of hydroxylic groups in their structures, leading to a strong affinity for water molecules. Subsequently, the system undergoes a heating phase to eliminate unfavourable contacts between solvent and solute, addressing issues like steric clashes. During this phase, the atoms' velocities are heightened and calculated using the standard temperature-dependent Maxwell-

Boltzmann distribution. The final step involves the equilibration of the system under controlled conditions of energy, temperature, pressure, and volume before initiating the production of the MD trajectory.

MD simulations, spanning various timescales and conducted in solution, have become indispensable for studying molecular properties and dynamic phenomena, providing a comprehensive understanding of interactions between ligands and proteins.

### 2.5.3 CORCEMA-ST

MD simulations can support the experimental data, such as results obtained by NMR studies, offering detailed insights into the conformational behavior of ligands and the construction of validated 3D models for complexes. In this context, CORCEMA-ST (COmplete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) is a tool designed for the quantitative analysis of STD NMR data.[184] This algorithm facilitates the prediction of STD intensities based on the Cartesian atomic coordinates of the ligand-receptor complex and allow their comparison with the experimental data.[185] Indeed, known some properties of the complex of interest, including for example the $k_{off}$ and the rotational correlation times of both the receptor and ligand, CORCEMA-ST can predict theoretical ligand STD intensities from a provided molecular model of the protein-ligand complex.

The comparison of experimental STD values with theoretical predictions enables the validation of a 3D model for the complex. This method relies on matrix calculations, where the Cartesian coordinates of all ligand and protein protons within a specified cut-off distance are taken into account. The accuracy of the molecular model in reproducing the experimental NMR data can be quantified through the R-NOE factor, calculated as follows:

$$\boldsymbol{R - NOE = \sqrt{\frac{\sigma W_k (STD_{exp,k} {*}\, STD_{cal,k})^2}{\sigma W_k (STD_{exp,k})^2}}} \qquad \textbf{Equation 2.12}$$

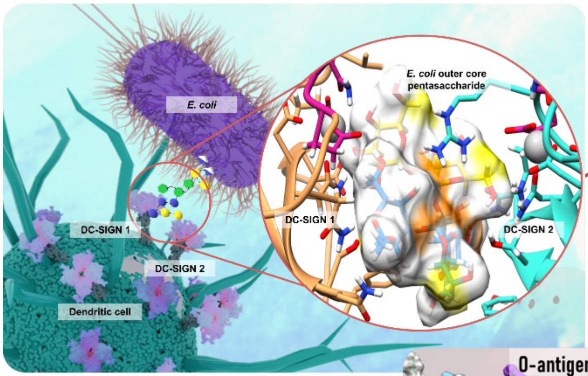
where $STD_{exp,k}$ are the experimental STD intensities and $STD_{cal,k}$ are the calculated STD intensities.

A good agreement between experimental and theoretical data is indicated by low R-NOE values, serving as a reliable validation for the 3D complex. Consequently, various structural models, derived from different sources like distinct docking runs or clusters of molecular dynamics (MD), can be ranked based on their ability to elucidate the experimental STD NMR data, aiming to identify the optimal model. The R-NOE factor serves as a scoring function, driving the conformational search for the ligand within the protein binding site. Following the identification of a promising starting pose, R-NOE values can be minimized by optimizing key torsion angles through simulated annealing, facilitating the discovery of the global energy minimum of the ligand bound to the receptor. Additionally, ligand geometry refinement within the binding site can be performed using experimental STD data.

To conclude, the integration of experimental techniques with computational approaches, like those described in this chapter, are essential for determining the 3D structures of ligand-receptor complexes.

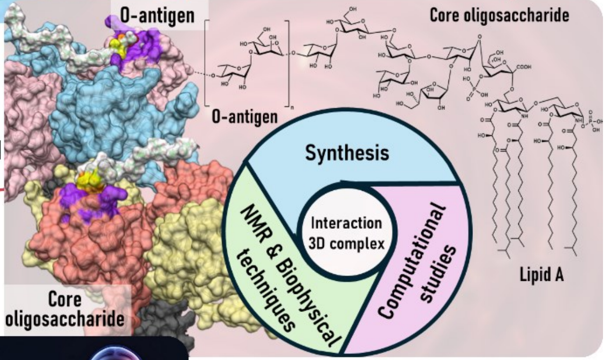
# SECTION II – RESULTS AND DISCUSSION

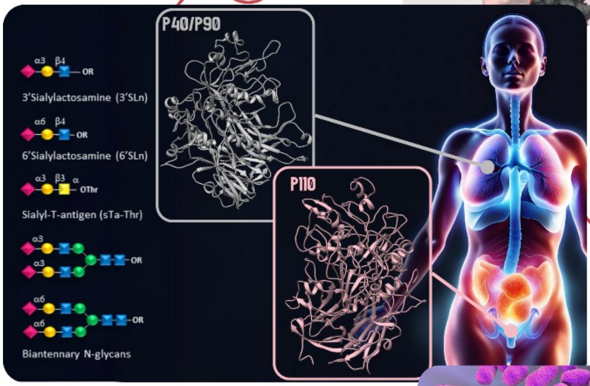
**The following chapters are linked throughout this manuscript by the "*fil rouge*" of the analysis of host-guest interactions.**



**Chapter III:**
Molecular recognition of *Escherichia coli* glycoconjugates by DC-SIGN

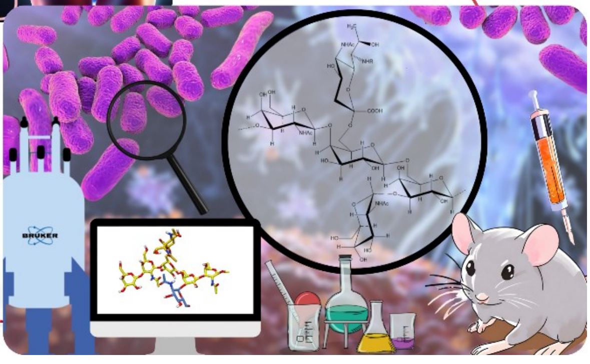
**Chapter IV:**
Molecular recognition of *Bacteroides vulgatus* glycoconjugates by DC-SIGN





**Chapter V:**
Molecular recognition of eukaryotic glycans by *Mycoplasma pneumoniae* and *Mycoplasma genitalium* cytoadhesins

**Chapter VI:**
Isolation and characterization of CPS from an *Acinetobacter baumannii* clinical isolate



# Chapter 3 – Molecular recognition of *Escherichia coli* glycoconjugates by DC-SIGN

## 3.1 Introduction

Lipopolysaccharides (LPS) are distinctive glycolipids located in the OM of Gram-negative bacteria (cf. §1.3).

Briefly, in their smooth form, LPS consist of three structurally and genetically distinct portions: the lipid A, the core oligosaccharide and the O-antigen. Conversely, rough form LPS (also known as LOS) lacks the O-antigen moiety. These bacterial cell envelope glycoconjugates, possessing the ability to trigger the innate immune response in eukaryotes and in humans, can lead to a cytokine storm, causing, in some cases, sepsis-induced death. Due to their capacity to initiate the host immune response, LPS are recognized as potent MAMPs, playing a crucial role in the pathogenesis of Gram-negative infections.[186-188]

The lipid A is considered the primary immunostimulatory component of LPS, modulating the immune response upon recognition by the MD-2/TLR4 receptor complex.[54] Despite numerous studies on the correlation between lipid A and immune system stimulation, depending on, as examples, from the number of acyl chains or phosphate groups it possesses,[189] limited information is available regarding the immune recognition of the saccharide portions (O-antigen and core oligosaccharide) of LPS. Actually, intense research in the field of innate immunity has unveiled the ability of the host immune system to respond to Gram-negative bacteria thanks to carbohydrate sensing macromolecules, *inter alia*.[190] However, both molecular insights and effects of these PRR-LPS interactions is far to complete. Moreover, it is easy to assume that there are unknown or uncharacterized receptor(s) with the ability to recognize O-PS and/or OS structures in bacterial LPS.

In innate immunity, lectins, a ubiquitous group of *non*-immunoglobulin proteins, often serve as PRRs. The C-type lectin (CTL) family, the largest and most diverse class of human lectins, includes both transmembrane and soluble receptors that recognize specific glycan structures through their Carbohydrate Recognition Domain (CRD), typically in a $Ca^{2+}$-dependent manner (cf. §1.5.1.1). One notable transmembrane CTL is DC-SIGN (cf. §1.5.1.2), expressed predominantly in dendritic cells (DCs) and present in macrophages and monocytes, acting as phagocytic cells. Through its adherence and phagocytosis functions, DC-SIGN plays a pivotal role in defending the host against various bacterial strains, underscoring its significance in the host's immune response.

DC-SIGN, belonging to the mannose receptor family, is expressed as a tetramer on the cell surface, with each monomer composed of a single CRD, a neck region, and an intracellular domain.[191] Thanks to the presence of the EPN (Glu-Pro-Asn) motif, DC-SIGN can bind bacterial glycans containing Fuc and Man residues, as well as GlcNAc-containing structures.[98, 104, 192] Some studies reported the phagocytic activity of DC-SIGN against different Gram negative bacteria including *E.coli,*[103, 193-195] especially in the absence of O-antigen polysaccharides, and in the presence of a complete core oligosaccharide.[196] *E.coli* is a commensal bacterium of the gastrointestinal tracts of humans and warm-blooded animals, but also can be pathogenic, causing a wide array of infections, such as enteritis, urinary tract infection, septicemia and neonatal meningitis.[197] As reported in literature, the LPS structure of *E. coli* comprises a highly conserved lipid A moiety and an extremely variable O-antigen portion, with over 170 known structures.[198] Regarding the core oligosaccharide, the inner region is highly conserved, featuring Kdo and Hep residues, often phosphorylated and/or decorated with nonstoichiometric glycosyl substituents. In contrast, the outer core exhibits higher variability and can be classified into five different families: R1, R2, R3, R4, and K-12 (Figure 3.1).[199]

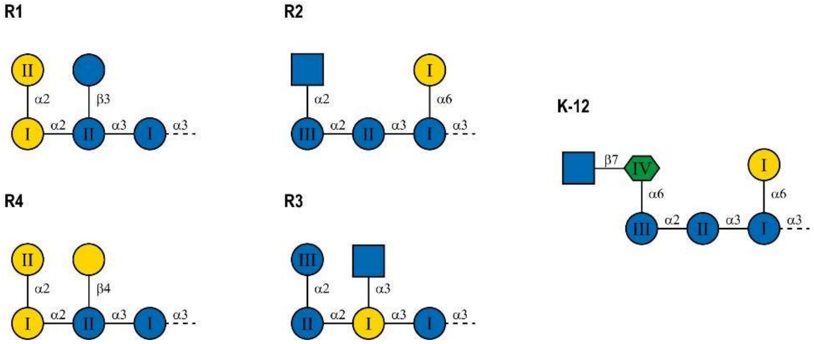


**Figure 3.1:** Schematic representation of the five types of *E. coli* outer core using SNFG nomenclature.

To better elucidate the molecular mechanisms underlying bacterial recognition by human lectins, the molecular binding between *E. coli* core OS type R1, the most abundant in clinical isolates, and DC-SIGN has been investigated through a combination of techniques, including fluorescence microscopy, Surface Plasmon Resonance (SPR), Nuclear Magnetic Resonance (NMR) spectroscopy, and computational studies, in order to build a 3D model of the interaction, as depicted in the workflow represented in Figure 3.2.

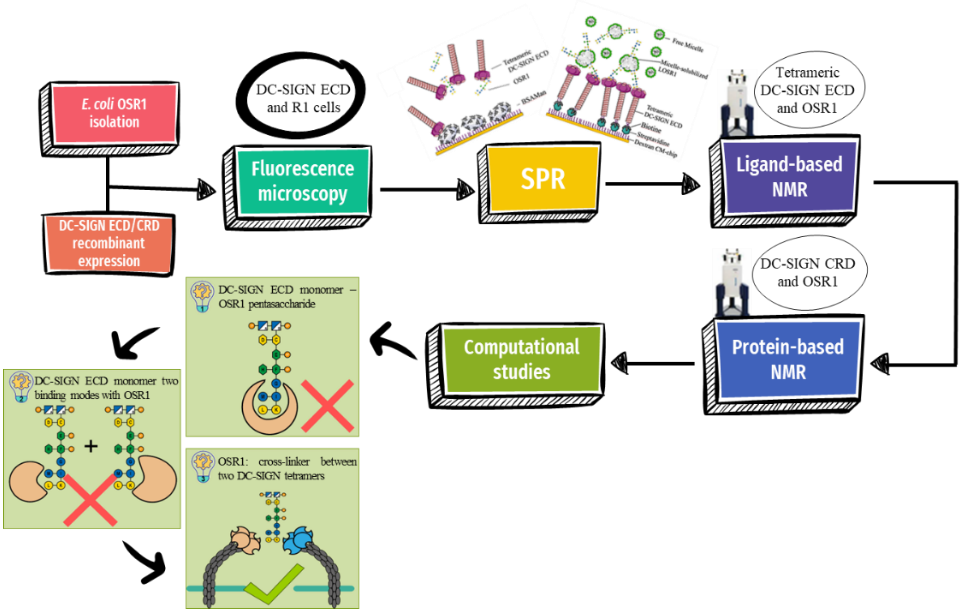


**Figure 3.2:** Workflow of the experimental activities.

## 3.2 Fluorescence microscopy

Preliminary fluorescence microscopy studies, conducted in collaboration with Dr. C. Laguri (Institut de Biologie Structurale, Grenoble, France), marked the initial exploration of the molecular recognition of *E. coli* core OS type R1 by DC-SIGN. In these *in vitro* experiments, DC-SIGN ECD labelled with Alexafluor 647 (AF647) was employed, and the interaction on the surface of *E. coli* R1 cells was observed through both Differential Interference Contrast (DIC) and epifluorescence imaging. A strong fluorescence signal was evident on the bacteria, confirming the binding of DC-SIGN to R1 cells (Figure 3.3A). To enhance the visualization of DC-SIGN binding on *E. coli*, Stochastic Optical Reconstruction Microscopy (STORM) was applied to the DC-SIGN/R1-cells interaction. This microscopy technique surpasses the resolution limits of traditional microscopy by exploiting the blinking characteristics of the fluorophore, which is maintained by depleting free oxygen in the buffer during the experiment. Serial images of bacteria labelled with DC-SIGN-AF647 were acquired over minutes, enabling the visualization of individual fluorophores' fluorescence emission (Figure 3.3B). STORM imaging of DC-SIGN labelled with AF647 on R1-cells clearly demonstrated the strong binding of the C-type lectin to the core oligosaccharide of R1-cells on the surface of *E. coli*. The specificity of this binding was further assessed through flow cytometry experiments conducted in the presence of increasing concentrations of OSR1. (Figure 3.3D).

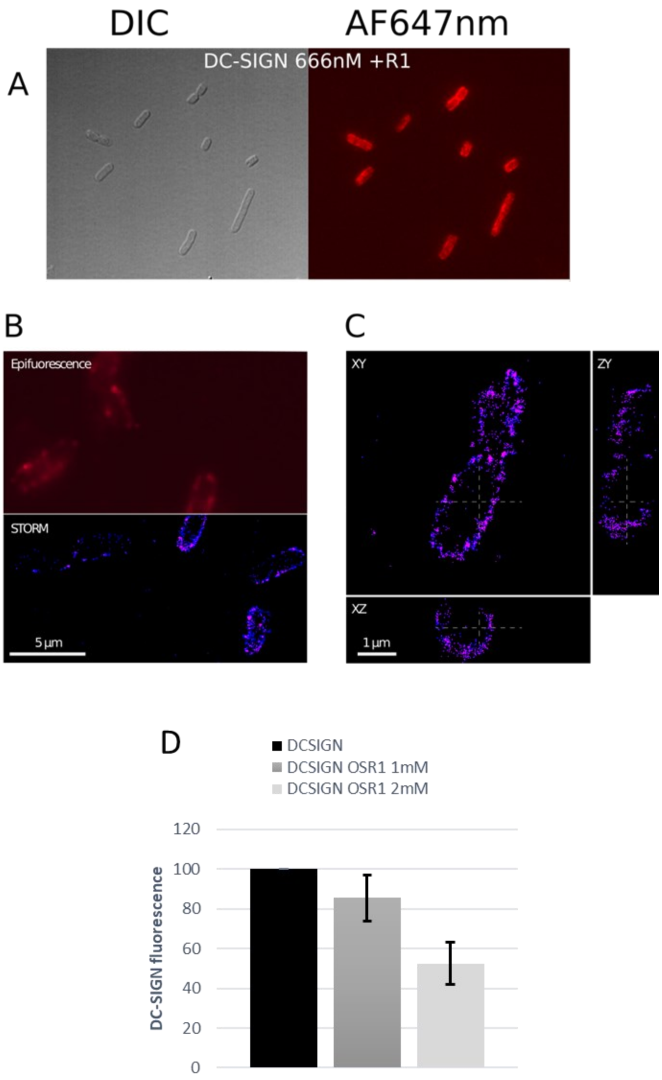


**Figure 3.3:** DC-SIGN ECD binds strongly to R1 core oligosaccharide at the surface of *E. coli.* A) DIC and epifluorescence imaging of R1-cells incubated with DC-SIGN ECD. STORM imaging of DC-SIGN ECD labelled with AF647 on E. coli R1 cells. B) Split panel showing wide field epifluorescence and STORM reconstructed imaging of the labelled bacteria. C) Orthogonal plane projections (according to the dashed lines) of a single bacterium from the field shown in the left panel. A cylindrical lens is used to provide the 3D localization precision. D) Flow cytometry quantification of DC-SIGN ECD labelled with AF647 bound to R1-cells in absence or presence of 1 and 2 mM OSR1. 50% binding inhibition is achieved with 2 mM OSR1. Experiments were done in triplicates and standard deviation is shown**.**

## 3.3 *E. coli* R1 LOS isolation and DC-SIGN recombinant expression

To comprehensively investigate the interaction between *E. coli* R1 LOS and DC-SIGN, the key components involved in the interaction have been isolated. Specifically, *E. coli* cells carrying R1 type core oligosaccharide were grown in LB at 37°C under agitation up to 0.9 $OD_{600nm}$. Cells were collected by centrifugation, a R1 type LOS from *E. coli* was extracted with PCP protocol (Figure 3.4A), as described in Chapter II. The LOS was then de-lipidated to remove ester- and amide-linked fatty acids. The structure of the obtained oligosaccharidic core, denoted as OSR1, (Figure 3.4B), was firstly controlled by NMR spectroscopy. In detail, OSR1 is a dodecasaccharide consisting of two Gal residues and three Glc units in the outer core region, along with three L-glycero-D-manno-heptoses (two phosphorylated at position 4) and two 3-deoxy-D-manno-oct-2-ulosonic acids (Kdo) in the inner core portion. The two GlcN residues at the reducing end (one phosphorylated at position 1 and the other at position 4) belong to the Lipid A moiety.[187]

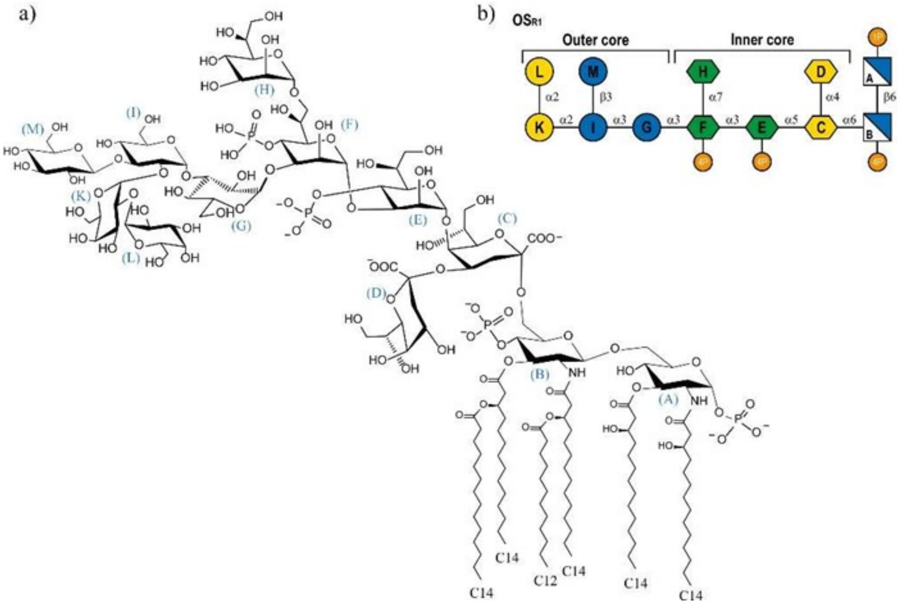


**Figure 3.4:** a) structure of the full extracted R1 type LOS from *E. coli*. b) Schematic representation of OSR1 after LOS de-lipidation, using SNFG nomenclature.

DC-SIGN ECD and DC-SIGN CRD were expressed and purified with a similar protocol, as described in Chapter II, during my secondment at Institut de Biologie Structurale (Grenoble, France), in the group of Prof. F. Fieschi. Proteins were expressed in *E.coli* BL21(DE3) competent cells, using the plasmids pET30b, containing cDNA encoding the ECD or CRD of DC-SIGN.[200] Proteins produced in inclusion bodies have been refolded by drop-by-drop dilution and dialyzed.[97] Purification of functional DC-SIGN proteins were achieved by an affinity chromatography on a mannan-agarose column, equilibrated in 25 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $CaCl_2$ (Buffer A), and eluted in the same buffer, lacking $CaCl_2$ but supplemented with 10 mM EDTA. This step was followed by a Superose 6 SEC equilibrated in buffer A (Figure 3.5).

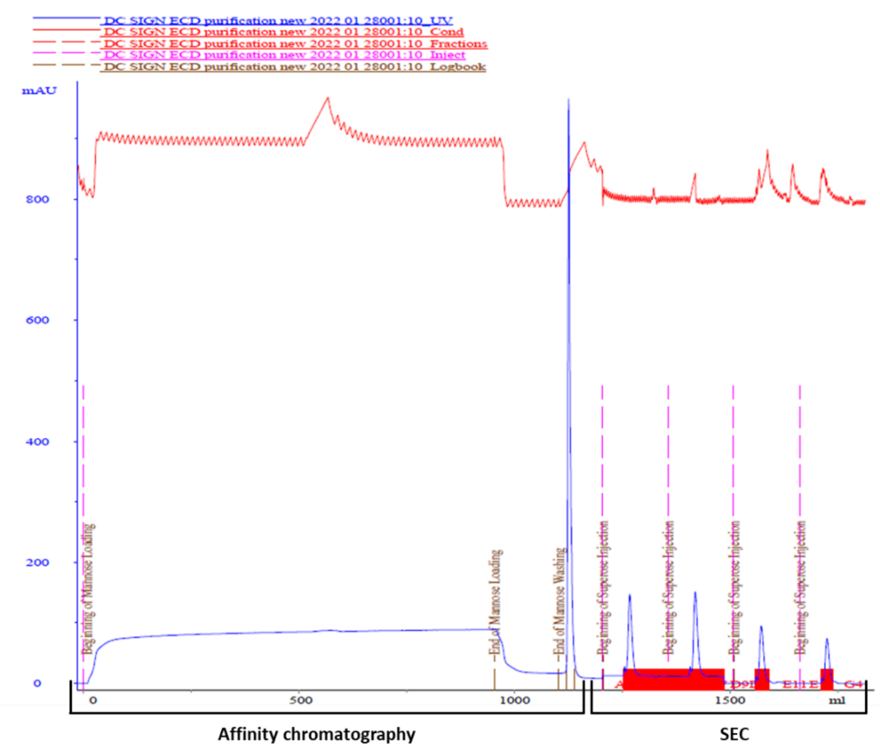
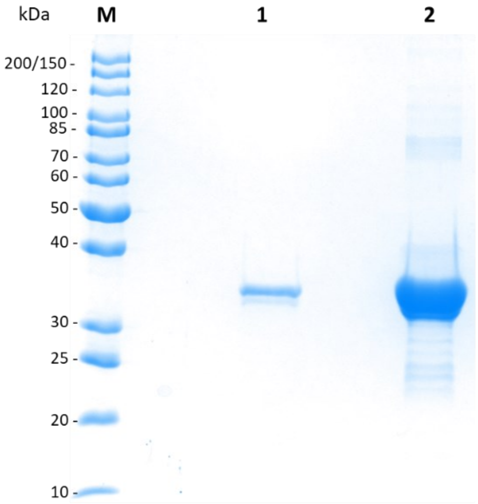


**Figure 3.5:** Purification of functional DC-SIGN proteins. First step: affinity chromatography on mannan-agarose column equilibrated in 25 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $CaCl_2$ (Buffer A), and eluted in the same buffer lacking $CaCl_2$ but supplemented with 10 mM EDTA. Second step: Superose 6 SEC equilibrated in buffer A.

The final pure sample of DC-SIGN ECD in two different quantities is reported as example in Figure 3.6.



M= MW standard (Page ruler unstained protein ladder)
1= 1 μg sample of pure DC-SIGN ECD
2= 10 μg sample of pure DC-SIGN ECD

**Figure 3.6:** 12% SDS-PAGE of two different quantities of pure DC-SIGN ECD, coloured with Coomassie staining.

Subsequently, SPR, NMR spectroscopy, and computational techniques like docking and molecular dynamic simulations have been employed to elucidate the molecular recognition between DC-SIGN and *E. coli* OSR1, constructing and validating a 3D model of the ligand-protein complex.

## 3.4 SPR analysis

The ability of DC-SIGN ECD to recognize OSR1 has been evaluated by SPR analysis, performed in collaboration with Prof. F. Fieschi (Institut de Biologie Structurale, Grenoble, France). The affinity of the OSR1 for DC-SIGN has been estimated through a classical competition assay. Tetrameric DC-SIGN ECD has been injected over a sensorchip functionalized with BSA-Mannotriose (Figure 3.7A). This interaction has been challenged with increasing concentration of OSR1 leading to inhibition of DC-SIGN interaction with the surface in a concentration dependant manner (Figure 3.7B). The inhibitory curve resulting from this competition experiment (Figure 3.7C) allowed to determine an $IC_{50}$ of 1.059 ± 0.003 mM. Thus, SPR experiments demonstrated the interaction between OSR1 and DC-SIGN. The observed affinity was significantly higher than those would have been expected, due to the presence of terminal Gal or Glc, as shown in position **L** and **M** (Figure 3.4b), which are not very strong ligands of DC-SIGN. As a reference, in this sort of competition tests, Man monosaccharide or tMan oligosaccharides classically give an $IC_{50}$ around 3 mM. Glc and Gal, instead, would provide an $IC_{50}$ just below 10 mM or even 30 mM respectively.[97] Thus, the observed $IC_{50}$ of 1 mM, strongly suggests a larger epitope of interaction. In this SPR competition test, DC-SIGN ECD was free in solution, exposed at the cell surface like in a physiological situation. Thus, DC-SIGN ECD has much less constraint. It is also the case for OSR1 extracted from the LPS and now as a soluble diffusible ligand.

To better mimic the physiological conditions of such interaction, a direct interaction test using SPR, has been also performed, in which C-type lectin receptor is functionalized in an oriented way.[201] Moreover, purified R1-cells LOS has been solubilized in detergent allowing the formation of mixed micelles of LPS/DDM, used for direct interaction study and injected onto DC SIGN oriented surface (Figure 3.7D) at different concentration of LOS. From the obtained sensorgrams (Figure 3.7E) a titration curve has been traced and allowed to evaluate a $K_{Dapp}$ of 15.6 μM (Figure 3.7F). From the first competition test of OSR1 to the second R1-cells LOS tested in direct interaction on an oriented surface, an increase of affinity by several order of magnitude has been observed (from $IC_{50}$ of 1mM to $K_{Dapp}$ of 15 μM). This suggested that, statistically, micelles containing several LOS can generate avidity through multivalency on the DC-SIGN oriented surface. In addition, it is not possible to exclude the possibility that the lipid A moiety of the LOS might contribute also to this improved binding strength.

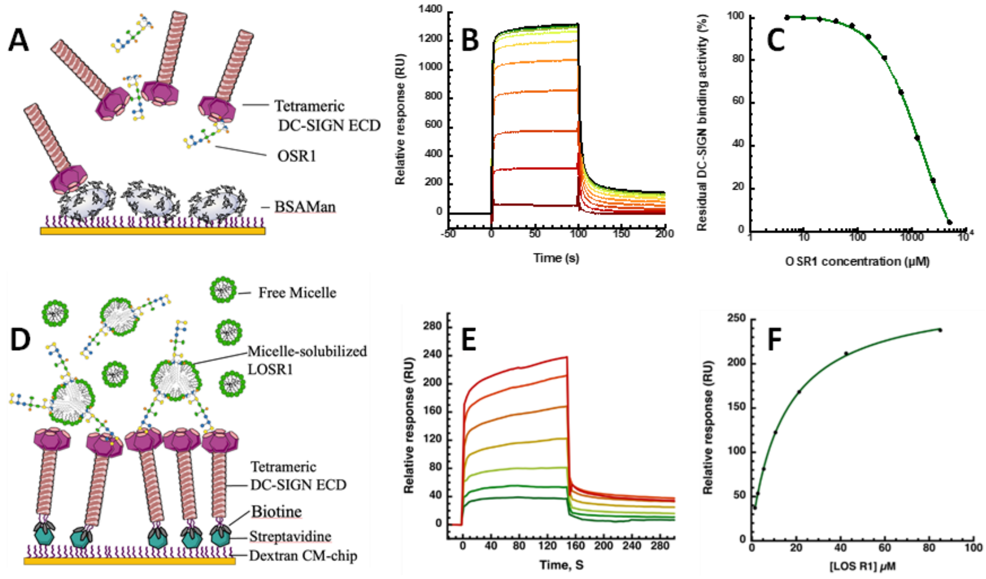


**Figure 3.7:** Analysis of DC-SIGN ECD interaction with OSR1 by SPR. (A) Principle of competition experiment. DC-SIGN ECD (20 µM) was incubated without or with OSR1 at increasing concentrations from 5 µM to 5 mM (from dark green to dark red)). The samples were co-injected over BSA-Man surface. The steady state responses were extracted from the sensorgrams (B), converted to DC-SIGN ECD residual activity, and plotted against OSR1 concentration (C). The inhibition curves were then fitted using four-parameter logistic model. The experiment was done twice using two distinct surfaces. (D) Principle of direct interaction of R1-cells LOS/DDM micelles with DC-SIGN ECD oriented surface. R1-cells LOS incorporated in DDM micelle have been injected onto DC-SIGN ECD at increased concentrations from 85 to 1.3 µM (serial dilution within running buffer by a factor of 2 from dark red to dark green). The steady state binding responses from the sensorgram (E) were plotted against LOSR1 concentration (F).

## 3.5 NMR analysis

The molecular interaction between tetrameric DC-SIGN ECD and OSR1 was further investigated using STD NMR.[135, 136] Firstly, a blank STD spectrum of OSR1 alone was acquired to confirm the absence of artifact signals (data not shown). STD NMR results supported the ability of DC-SIGN ECD to recognize OSR1, as evidenced by multiple enhancements observed in the STD NMR spectrum acquired upon the addition of OSR1 to a solution of the recombinant form of the protein, previously used also for the SPR experiments. The analysis of the less crowded regions of the spectrum, namely the one containing the anomeric signals, and the comparison of both multiplicity and intensity of the STD NMR signals with respect to the reference spectrum, revealed an extended binding epitope involving the outer core region, accordingly with previous SPR analysis results (Figure 3.8A). Specifically, the anomeric protons of **I**, **K**, **L** and **G** sugars (Figure 3.8B) gave rise to remarkable STD effects; on the contrary, the resonances of **E** and **F** residues were not observed in the STD NMR spectrum, thus suggesting the outer core moiety as the one mainly involved in the binding with DC-SIGN. Accordingly, the most intense signals corresponded to the H4 of Glc residues **M** and **I**. Also, H2, H3 and H5 protons of **M** strongly contributed to the interaction. An intermediate contribution was observed for H1, H2 and H3 protons of Glc **I** and H3, H4 and H5 of Gal **L**. Lower STD enhancements were observed for signals belonging to some protons of the Gal residues **K** and **L** and Glc **G**. For **K**, H1, H3 and H4 exhibited a higher contribution than H6. On the other hand, H3 of **G** showed a higher contribution than H1 and H4 of the same residue. The other protons of these two saccharides (**K** and **G**) did not seem to have an interaction with the protein. Moreover, signals belonging to the inner core residues were not observed as indicated, for example, by the absence of STD NMR signals either of the diastereotopic methylene protons of Kdo residues (**C** and **D**) and of the protons at position 2 of the GlcN residues **A** and **B** of the lipid A, further suggesting the main contribution of the outer core to the interaction of OSR1 with DC-SIGN.

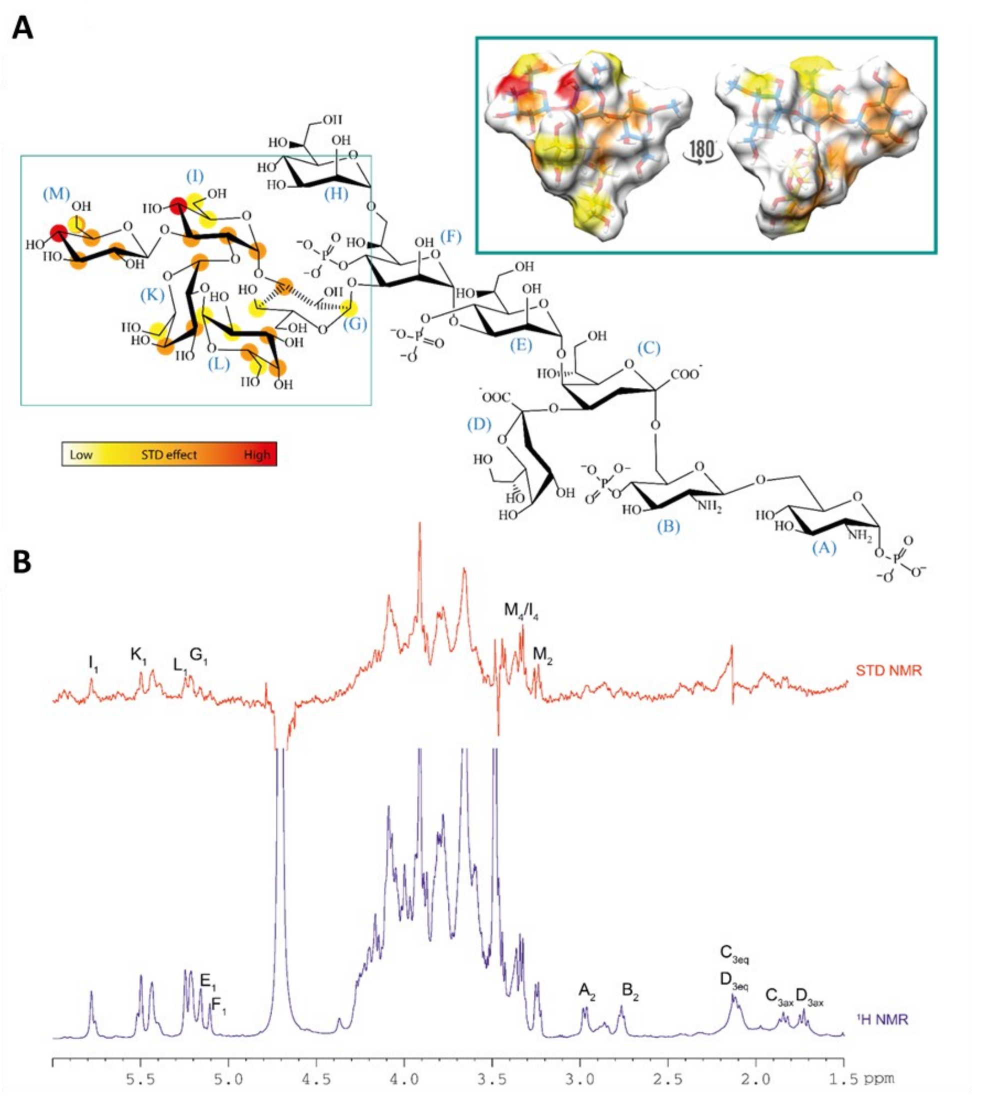


**Figure 3.8:** A) STD-derived epitope mapping of the interacting oligosaccharide with colour coding from the highest (red) to the lowest (yellow) observed STD effect and the 3D representation of the STD-derived epitope mapping in the square. B) $^{1}$H NMR reference spectrum at the bottom (blue) with the 1D STD NMR spectrum on top (red) of the 1:90 mixture for DC SIGN:OSR1 complex with some of the key proton resonance signals labelled.

To pinpoint the LOS binding site and gain insights into the involved protein residues, protein-based NMR experiments were also conducted. Once expressed and purified the $^{15}$N-labelled DC-SIGN CRD, its interaction with OSR1 was examined using $^{1}$H-$^{15}$N NMR spectroscopy, in collaboration with Dr. Laguri. Unfortunately, any significant chemical shift perturbation came out in DC-SIGN CRD upon the addition of 20 molar equivalents of the oligosaccharide in solution (Figure 3.9). This could be attributed to a very low affinity between the protein CRD and the core oligosaccharide, probably due to distinct conformational differences in the recombinant CRD compared to the ECD in solution.

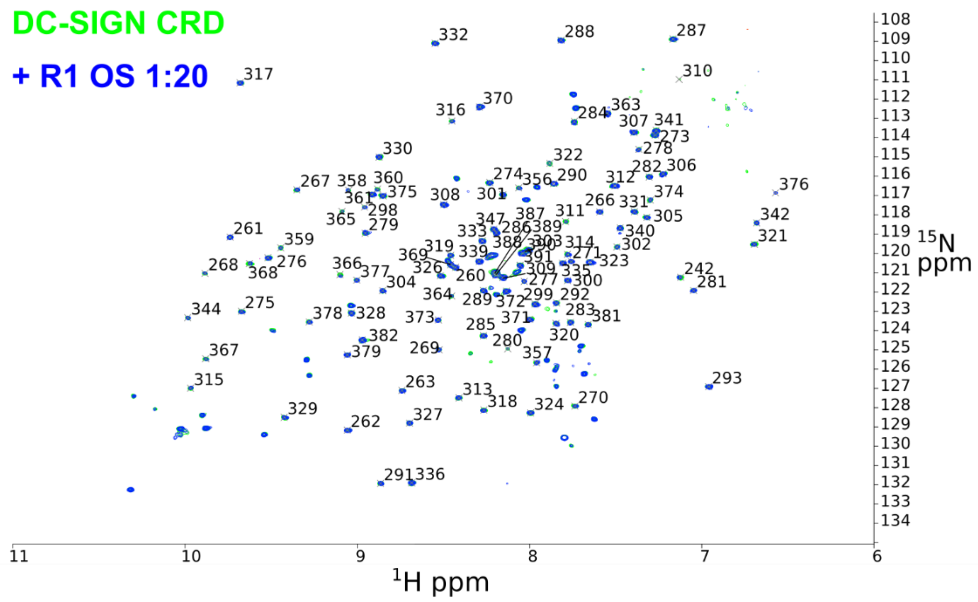


**Figure 3.9:** $^{1}$H-$^{15}$N Best-Trosy spectra of DC-SIGN CRD without (green) and with (blue) 20 molar equivalents of OSR1.

## 3.6 Molecular modeling analysis

Due to the presence of isochronous NMR signals, not allowing a precise assignment of all the STD NMR effects and hindering the quantitative determination of the ligand epitope mapping, computational studies were employed to support and validate the experimental findings permitting to construct a 3D model of protein-ligand complex. Applying a reductionist approach, the pentasaccharide composed of the five outer core residues (**G**, **I**, **K**, **L** and **M**) from OSR1 was used as ligand and built with the GLYCAM[202] carbohydrate builder. Then, its interaction with a single monomeric DC-SIGN subunit (PDB:1K9I), has been investigated.

By means of Autodock 4.2[167], docking calculations were performed to initially predict the interactions between the protein and ligand in the modelled complex. This enabled the selection of representative poses from the most populated clusters to serve as the starting point for subsequent MD simulations in explicit solvent using AMBER.

Considering the apparent recognition of the entire outer core through NMR, it was initially assumed that the protein had a sufficiently large binding pocket to accommodate the complete pentasaccharide. However, computational studies revealed that the five sugars of the outer core could not interact simultaneously with the protein binding pocket, as half of them were found to be solvent exposed. Taking into account that STD NMR experiments can be influenced by various ligand orientations within the protein binding pocket,[167, 203] a second hypothesis was postulated based on the possibility of having multiple binding modes. The docking results, in fact, already presented two interesting poses for OSR1 pentasaccharide recognition by DC-SIGN ECD. The less energetic cluster (cluster A) displayed residues **L** and **G** in the protein binding site (Figure 3.10A) while the most populated cluster (cluster B) presents **K**, **I** and **M** accommodated inside the pocket (Figure 3.11A). The best representative poses from the different orientations of the ligand, taking always into account that the chain extends through residue **G**, were used for running MD simulations. Regarding cluster A, although the **L** residue could be found coordinating with the $Ca^{2+}$ ion, after a few nanoseconds the ligand left the binding pocket, resulting in a *non*-stable MD (Figure 3.10B).

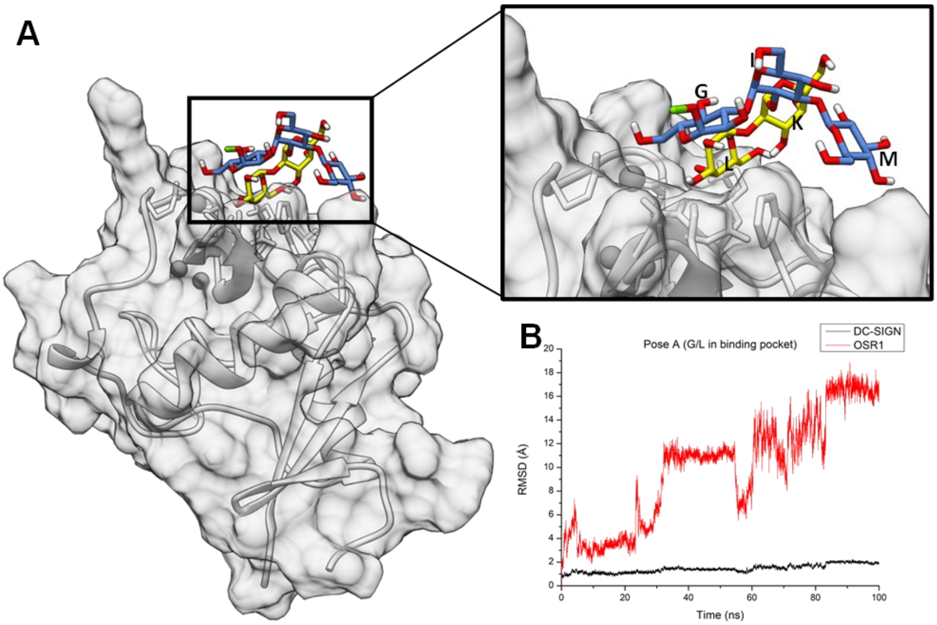
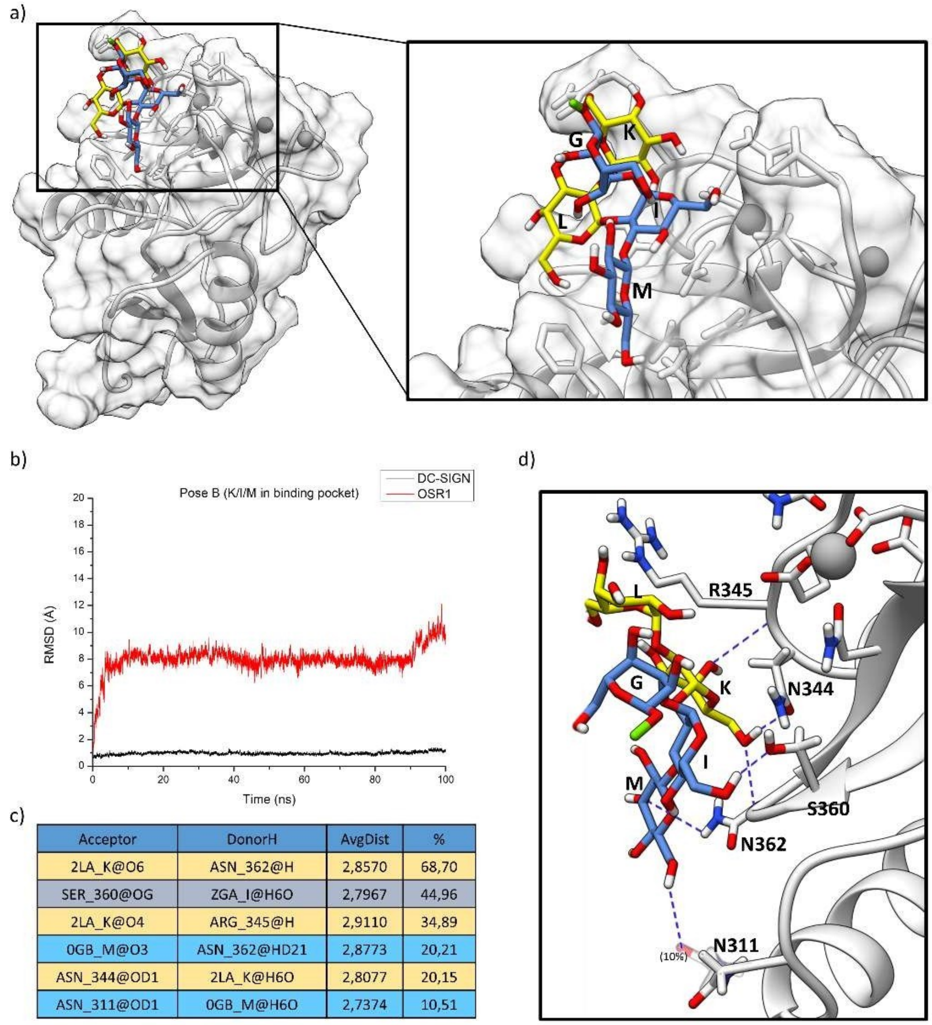


**Figure 3.10:** Less energetic cluster (cluster A) from the docking studies between DC-SIGN monomer and OSR1 outer core pentasaccharide. It is possible to see residues L and G in the binding pocket while the rest are solvent exposed. b) RMSD of the DC-SIGN monomer (black) and OSR1 pentasaccharide (red) after MD simulation calculated having the protein as reference The protein is stable while it is possible to observe how after a few nanoseconds the ligand leaves the binding pocket.

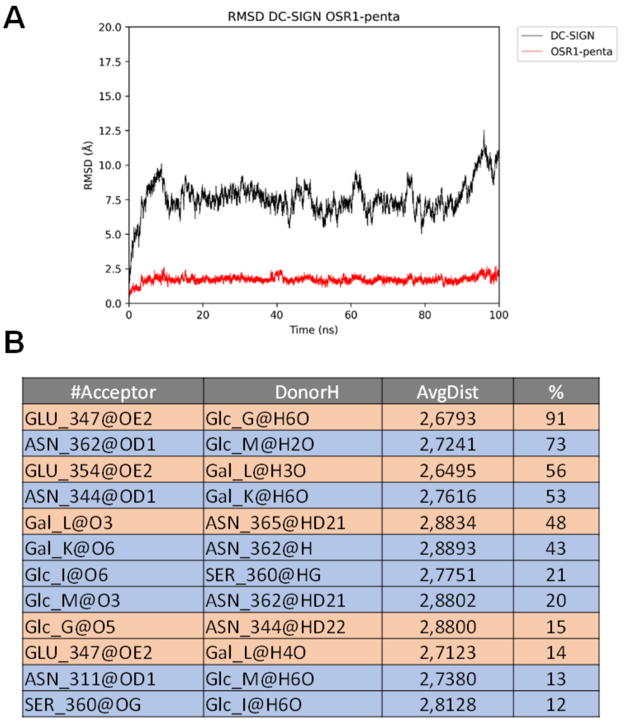


| Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| 2LA_K@O6 | ASN_362@H | 2,8570 | 68,70 |
| SER_360@OG | ZGA_I@H6O | 2,7967 | 44,96 |
| 2LA_K@O4 | ARG_345@H | 2,9110 | 34,89 |
| 0GB_M@O3 | ASN_362@HD21 | 2,8773 | 20,21 |
| ASN_344@OD1 | 2LA_K@H6O | 2,8077 | 20,15 |
| ASN_311@OD1 | 0GB_M@H6O | 2,7374 | 10,51 |

**Figure 3.11:** a) Representative pose of the most populated cluster (cluster B) from the docking studies between DC-SIGN monomer and OSR1 outer core pentasaccharide. Residues K, I, and M are in the binding pocket while L and G are solvent exposed. B) RMSD of the DC-SIGN monomer (black) and OSR1 pentasaccharide (red) calculated having the protein as reference. The protein is stable while the ligand after few nanoseconds shifts from the binding pocket to interact with a polar region. C) Protein-ligand H bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD coloured according to the ligand residue involved in the interaction. D) Representative pose of the interaction between DC-SIGN monomer protein and OSR1 pentasaccharide. The ligand is coloured according to the SNFG nomenclature. $Ca^{2+}$ ion is coloured in grey. In green the linking point between residue G and the rest of the LOS.

For the cluster B, something less expected happened; after a few nanoseconds the ligand shifted towards a polar region instead of interacting with the expected residues in the protein binding pocket, establishing contacts with Asn362, Arg345 and Asn344 through residue **K**, with Ser360 through unit **I**, and with Asn362 and Asn311 via **M** (Figure 3.11D). Thus, the obtained results indicated that the ligand was not stable in the principal $Ca^{2+}$-dependent binding pocket, suggesting that the different binding modes hypothesis were not feasible.

Structurally, DC-SIGN CRD consists of two α-helices and five β-strands. The loop extending beyond the protein surface is responsible for forming two cavities that accommodate $Ca^{2+}$ ions and plays a crucial role in carbohydrate binding. In particular, the principal binding site, constituted by the EPN motif (Glu347, Pro348 and Asn349) together with Glu354 and Asn365 residues, is the one essential for carbohydrate coordination and manages the specificity for the recognized ligands. The considered hypothesis was that OSR1 could be recognized and linked at the interface of two DC-SIGN monomers, serving as a linker to cluster two distinct tetrameric DC-SIGN units.[98, 203-205] To test this hypothesis, a new complex was modelled, consisting of two DC-SIGN subunits with the pentasaccharide positioned at the interface between them. Interestingly, the complex was stable along the 100 ns trajectory (Figure 3.12).



| #Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| GLU_347@OE2 | Glc_G@H6O | 2,6793 | 91 |
| ASN_362@OD1 | Glc_M@H2O | 2,7241 | 73 |
| GLU_354@OE2 | Gal_L@H3O | 2,6495 | 56 |
| ASN_344@OD1 | Gal_K@H6O | 2,7616 | 53 |
| Gal_L@O3 | ASN_365@HD21 | 2,8834 | 48 |
| Gal_K@O6 | ASN_362@H | 2,8893 | 43 |
| Glc_I@O6 | SER_360@HG | 2,7751 | 21 |
| Glc_M@O3 | ASN_362@HD21 | 2,8802 | 20 |
| Glc_G@O5 | ASN_344@HD22 | 2,8800 | 15 |
| GLU_347@OE2 | Gal_L@H4O | 2,7123 | 14 |
| ASN_311@OD1 | Glc_M@H6O | 2,7380 | 13 |
| SER_360@OG | Glc_I@H6O | 2,8128 | 12 |

**Figure 3.12:** Analysis of the interaction and stability of OSR1 recognition by two DC-SIGN units. A) DC-SIGN (black) and OSR1-pentasaccharide (red) RMSD. The ligand RMSD was calculated in reference to the protein. B) Protein-ligand H-bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD.

Particularly, as shown in Figure 3.13, where the most representative complex is reported, the binding site of one of the subunits still encloses the ligand, while the other half, solvent-exposed in the monomer, can interact with the second DC-SIGN subunit. In detail, **G** and **L** residues were interacting with DC SIGN in the principal $Ca^{2+}$-dependent binding pocket. The most stable H-bond involved the hydroxyl moiety at position 6 of residue **G** and Glu347. Other H-bonds were instead formed with residue **L**, which was interacting with Glu354 and Asn365 through 3-OH and 4-OH, respectively. Indeed, Gal **L** hydroxyl moieties at positions 3 and 4 were also coordinating with the $Ca^{2+}$ ion (Figure 3.14).

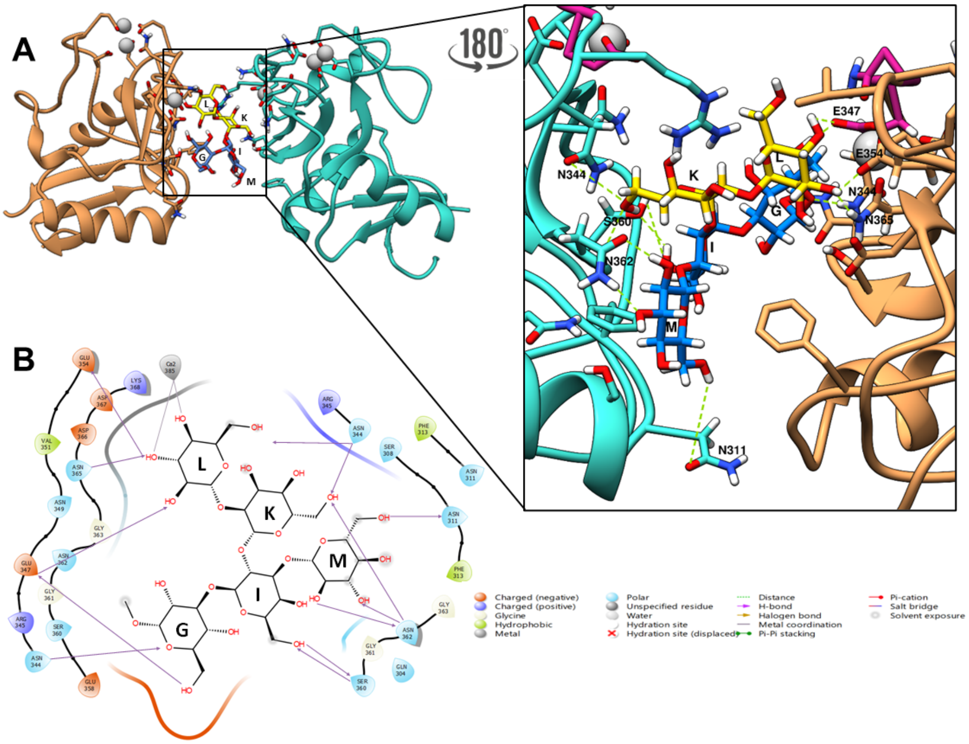


**Figure 3.13:** A three-dimensional representative model coming from the 100 ns MD for the complex DC-SIGN dimer and OSR1 outer core pentasaccharide. A) The ligand is coloured with the SNFG colour assignment. The outer-core pentasaccharide acts as linker between two different DC-SIGN tetramers. Residues G and L interact with one subunit of a tetrameric DC-SIGN (orange) while K, I and M interact with the other subunit of a different tetrameric DC-SIGN (cyan). The interactions between the pentasaccharide and the binding pocket residues are depicted. The most relevant AA involved in the interaction are labelled. The polar contacts are highlighted in green. The EPN motif is highlighted in purple. B) Two-dimensional schematic plot of the interactions between the DC-SIGN dimer and the outer core pentasaccharide: solid arrows represent the hydrogen bonds with functional groups of the backbone; solid lines represent the $Ca^{2+}$- L coordination; the other depicted residues participate with polar and hydrophobic interactions.

Those interactions were consistent with the literature as recent studies demonstrated the ability of DC-SIGN to recognize α-Gal,[206] through the coordination of the hydroxyl group at positions 3 and 4 by $Ca^{2+}$. In addition, residues **M**, **I** and **K** were involved in the interaction with the second subunit. Interestingly, the interaction did not take place in the principal $Ca^{2+}$-dependent binding site of this second subunit; those residues were indeed interacting with polar amino acids Asn311, Asn344, Ser360 and Asn362, which formed stable hydrogen bonds. In detail, 2-OH and 3-OH of residue **M** were stabilized by hydrogen bonds with Asn362, observed for the 73% of the simulation time. This was in agreement with the NMR results, given the significant contribution of **M** to the interaction. Moreover, the residue Asn362 also formed a hydrogen bond with 6-OH of **K**, which remained stable during 43% of the simulation time. The same 6-OH was acting as a H donor with Asn344 for another 53% of the simulation time. Similarly to **K**, 6 OH from residue **I** alternated between hydrogen bond donor and acceptor when interacting with Ser360. Noteworthy, these interactions were comparable to those observed in the previously mentioned MD of cluster B. This suggested that the coordination and interaction between residues **L** and **G** alone might not be sufficiently stable, thus the stability was achieved through the engagement of a polar region in another tetrameric DC-SIGN in the interaction.

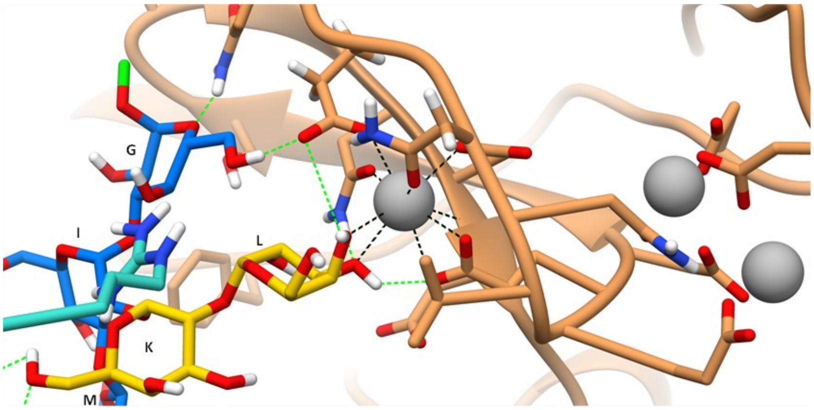
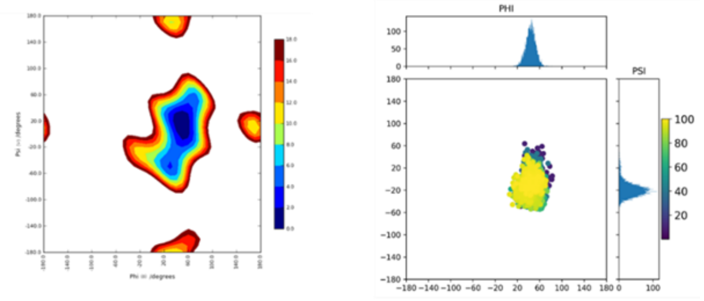


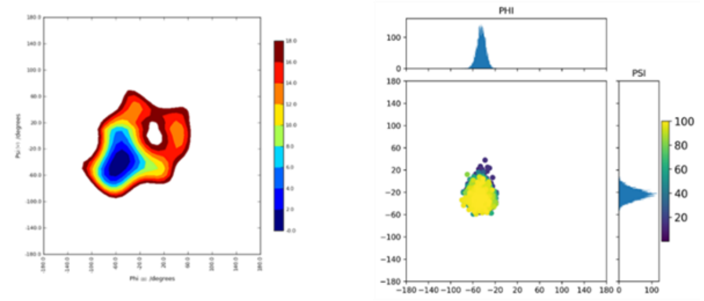
**Figure 3.14:** Close up view of the interaction between residues G and L and one of the DC-SIGN subunits. The ligand is coloured with the SNFG colour assignment. Highlighted in green in G the linking point to the rest of OSR1. $Ca^{2+}$ ion is coloured in grey. The coordination interactions are coloured in black while the H-bonds are coloured in green.

The conformational behavior of the ligand was also evaluated; the dihedral angles $\phi$ and $\Psi$ were monitored along the trajectories of both free state (data not shown) and bound state. All dihedral values around the glycosidic linkages aligned with the *exo*-anomeric effect (Figure 3.15), and there were no notable differences between the free and bound states. This observation implies that the ligand maintained the same conformation both before and after binding.

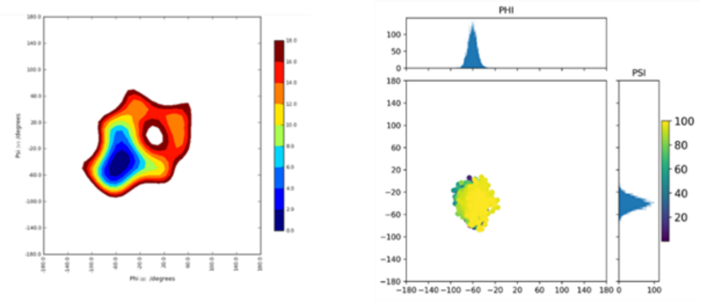
β-Glc-(1-3)-α-Glc
**(M-I)**



α -Glc-(1-3)-α-Glc
**(I-G)**



α -Gal-(1-2)-α-Gal
**(L-K)**



α -Gal-(1-2)-α-Glc
**(K-I)**

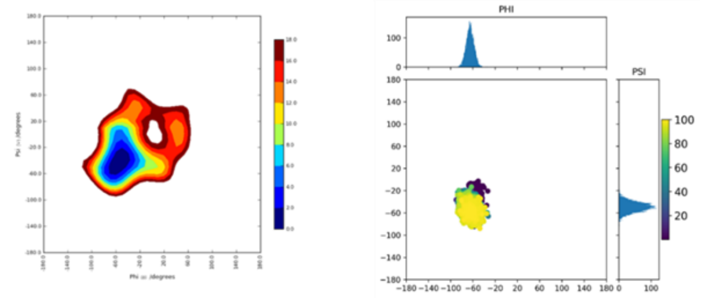


**Figure 3.15:** Adiabatic energy maps, obtained by MM calculations, illustrating the energetic minima of the glycosidic torsion angles ϕ/ψ of the outer core LOS R1 constituent disaccharides. Outer core pentasaccharide dihedral angles analysis in the bound state represented as scatter plots of the ϕ torsion against ψ, during the MD simulation with the relative histograms to represent the most populated energies**.**

To assess potential steric hindrance or clashes during the interaction between the protein and the entire core oligosaccharide, the complete OSR1 was modelled into one of the representative poses obtained from the MD simulations with the two DC-SIGN units. The entire saccharide can fit into the binding site, with the inner core and lipid A sugars solvent exposed, with no clashes with the protein (Figure 3.16).

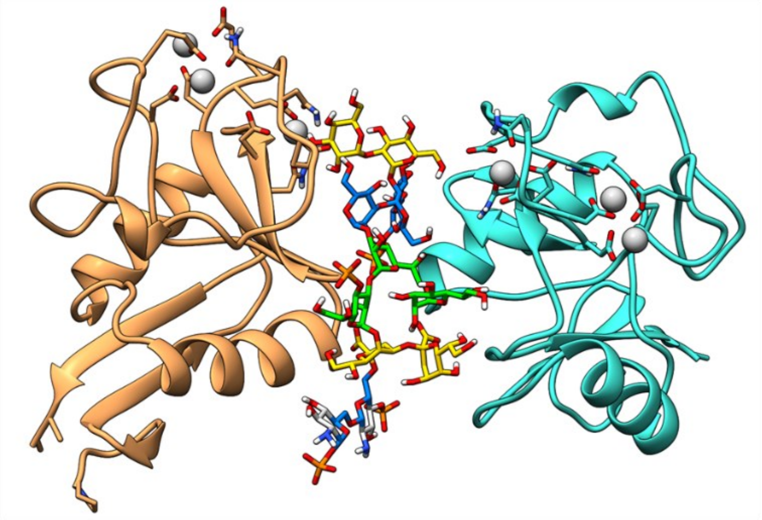


**Figure 3.16:** Manual superimposition of the full OSR1 dodecasaccharide with the dimeric DC-SIGN. The monosaccharides are coloured using the SNFG colour code with the phosphate groups in orange. The subunits coming from each tetrameric DC-SIGN unit are coloured in sandy brown and turquoise with the calcium ions in grey.

Additionally, after the superimposition of the two DC-SIGN tetramers to the system, any clashes between protein chains were also excluded, validating the proposed complex (Figure 3.17).

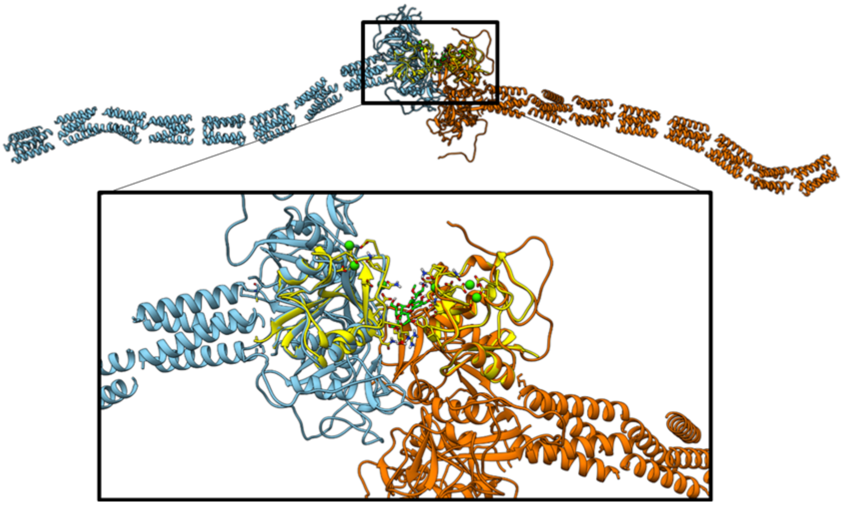


**Figure 3.17:** Proposal of the 3D model for the recognition of OSR1 by DC-SIGN with the glycoconjugate acting as crosslinker between two different DC-SIGN tetramers. Two full DC-SIGN tetramers with neck (blue and orange) superimposed to the MD pose (yellow).

These findings align with NMR studies, supporting the interaction between the second subunit and saccharides **M**, **I**, and **K**, which could explain the STD NMR enhancements across the entire pentasaccharide (Figure 3.18).

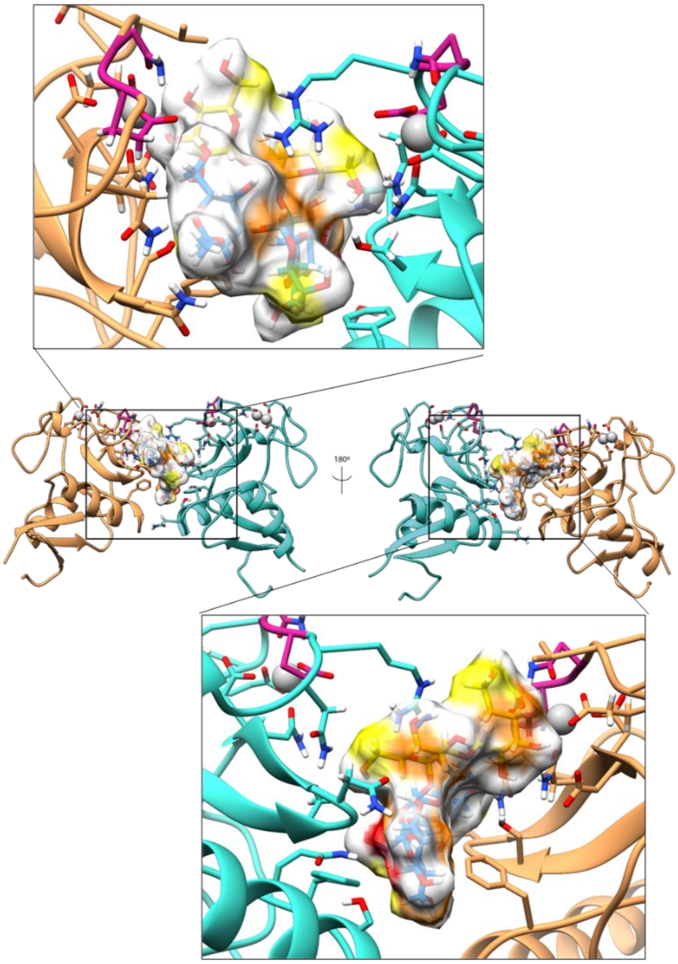


**Figure 3.18:** A 3D representative model coming from the 100 ns MD for the complex DC-SIGN dimer – Outer core pentasaccharide. The ligand is coloured with the SNFG colour assignment while its surface is coloured according to the STD data. In sandy brown and turquoise there are the subunits coming from each DC-SIGN tetrameric unit. The calcium ions can be found coloured in grey.

## 3.7 Discussion

Lectins, crucial in cellular trafficking and cell–cell communication, are potential therapeutic targets. Among these, the C-type lectin DC-SIGN plays a pivotal role in immune response regulation by recognizing carbohydrate structures on pathogens or self-glycoproteins. Despite growing understanding of DC-SIGN biological functions, a comprehensive molecular understanding of its interaction with bacterial glycan ligands remains elusive.

Based on these premises, in this thesis, once isolated the DC-SIGN and the oligosaccharide core OSR1, the protein ability to recognize *E. coli* LOS has been investigated by using an integrated approach including fluorescence microscopy, SPR, NMR and computational studies. Fluorescence microscopy assessed the strong binding of DC-SIGN ECD to *E. coli* surfaces with R1 type core oligosaccharides, while SPR estimated affinity values. Ligand conformation and epitope mapping were characterized through a combination of results from STD NMR and computational studies, enabling the proposal of a 3D interaction model.

In summary, DC-SIGN ECD selectively binds the outer core carbohydrate residues of OSR1, with an $IC_{50}$ of 1 mM, indicating a large epitope of interaction. The proposed 3D model highlights the interaction of **G** and **L** OS units (Figure 3.4B) with DC SIGN in the principal $Ca^{2+}$-dependent binding pocket, with the **L** residue coordinating with the $Ca^{2+}$ ion through the hydroxyl groups at positions 3 and 4. Additional contribution to the binding comes from ligand interaction with a second protein subunit, mediated by residues **M**, **I** and **K**, probably accommodated in a $Ca^{2+}$-independent secondary binding-site. The proposed 3D model suggested a potential multivalent interaction mode, recognizing OSR1 at the interface of two tetrameric units of DC-SIGN. The main engagement of the outer core region in the interaction involved both the primary $Ca^{2+}$-dependent binding pocket and a secondary binding site on a second subunit. Different techniques have been also employed to mimic physiological conditions in which DC-SIGN ECD is embedded in the cell membrane and the core OS is presented on LOS inserted in the *E. coli* OM. These experiments allowed to exclude the possibility of a different interaction mode. Thus, epifluorescence, STORM microscopies and SPR confirmed the relevance of the DC-SIGN/OSR1 interaction.

The presented results provide detailed atomic structural insights into the interaction between DC-SIGN and microbial glycans. Notably, the high homology in gene sequences within the *Enterobacteriaceae* family suggests that DC-SIGN may serve as a receptor for genetically related bacterial strains, paving the way for further molecular insights. Furthermore, the discovery of the interaction between DC-SIGN and OSR1 from the pathogenic bacterium *E. coli* has several potential applications and implications. CTLs and, more specifically, DC-SIGN, can be regarded as mediators between innate and adaptive immunity. Due to its capability to enhance immune responses, DC-SIGN can be a highly attractive target for developing CD8 cytotoxic T lymphocyte-agonistic vaccine adjuvants. Additionally, this interaction can be explored for its potential in immunotherapeutic strategies. Knowing the protein residues that interact with OSR1, thanks to the results obtained from MD simulations and protein-based NMR, by using protein engineering and the manipulation of the encoding DNA, it may be possible to modify the DC-SIGN sequence in order to incorporate specific properties, like increased substrate specificity, to boost the immune response against *E. coli* and related pathogens. Yet, this interaction also provides a potential target for drug development. Small molecules or biologics capable of modulating DC-SIGN's interaction with *E. coli* could be developed as potential therapeutics. It has been reported that DC-SIGN recognizes the SARS-CoV-2 Spike Protein, facilitating viral trans-infection and enhancing viral transmission to susceptible cells. This process is significant in the context of viral pathogenesis and the development of severe COVID-19.[207] Understanding the molecular mechanisms by which DC-SIGN facilitates viral entry and dissemination not only have provided valuable insights into the pathogenesis of COVID-19 but, also, have enabled the development of glycomimetic antagonists to inhibit the interaction with the SARS-CoV-2 spike glycoprotein, potentially reducing viral transmission.[208] The results described in this chapter expand the scope by suggesting that DC-SIGN interaction, not only involves viral pathogens like SARS-CoV-2 but, also, encompasses bacteria such as *E. coli.* Thus, the same approach employed in the design of glycomimetic antagonists against the SARS-CoV-2 spike glycoprotein[208] could be applied to target *E. coli,* ubiquitous pathogenic bacterium crucial for public health, food safety and antimicrobial resistance.

# Chapter 4 - Molecular recognition of *Bacteroides vulgatus* glycoconjugates by DC-SIGN

## 4.1 Introduction

The human gastrointestinal tract hosts the largest population of microorganisms, known as the gut microbiota (GM). This complex and ever-changing assembly consists of commensal bacteria, fungi and viruses, playing a crucial role in activating, developing and functionally maturing both the mucosal and systemic immune systems.[209] The GM modulate host immunity and maintain homeostasis, providing several benefits, including detoxification and processing of nutrients from dietary intake. Additionally, the GM contributes to protect against external pathogens by strengthening the intestinal barrier through direct mechanisms such as competition for common nutrients and niches, as well as indirect mechanisms, such as enhancing host defense.[52, 210]

As previously mentioned, the immune system possesses a large variety of PRRs, able to recognize specific MAMPs (cf. §1.4). Among them, LPS, the major component of Gram-negative bacteria OM, is a crucial MAMP that has traditionally been believed to stimulate only harmful interactions with the host. However, recent findings suggest that LPS can also beneficially stimulate the host's immune system. Processes such as colonization, virulence, adhesion, symbiosis, or tolerance are attributed to LPS and depend on its chemical structure; any minor structural change can impact its function and recognition by the immune system receptors.[211, 212]

*Bacteroides* genus comprises obligate anaerobic Gram-negative bacteria, widely distributed as commensals and essential beneficial components of the human colon.[213, 214] For instance, oral administration of *B. vulgatus* can restore colonic gap junctions and prevent colitis in a strain-dependent manner;[215] similarly, *B. vulgatus* mpk strain

(BVMPK)[211, 216] can protect against *E. coli*-induced colitis or *Yersinia enterocolitica*-induced inflammation, thereby reducing intestinal inflammation and promoting tissue repair.[211, 214] BVMPK LPS has been identified as a weak agonistic LPS, lacking a pro-inflammatory response in both human *in vitro* and murine *in vivo* experiments, stimulating the release of anti-inflammatory cytokines.[217]

Moreover, BVMPK LPS exhibits selective affinity for human CTL DC-SIGN, associated with gut lymphoid tissue and homeostasis, playing a pivotal role in the host immune system, as previously mentioned in Chapter III. Given DC-SIGN's capability to recognize bacterial cell envelope glycoconjugates[218-221] and the significance of specific glycan motifs in LPS that modulate immune functions, this chapter presents a structural evaluation of BVMPK LPS recognized by DC-SIGN. The interaction with DC-SIGN will be depicted using different glycoforms of BVMPK O-antigen and core oligosaccharide to assess the contribution of both regions, employing synthetic, spectroscopic, and biophysical approaches, as depicted in the workflow represented in Figure 4.1 .

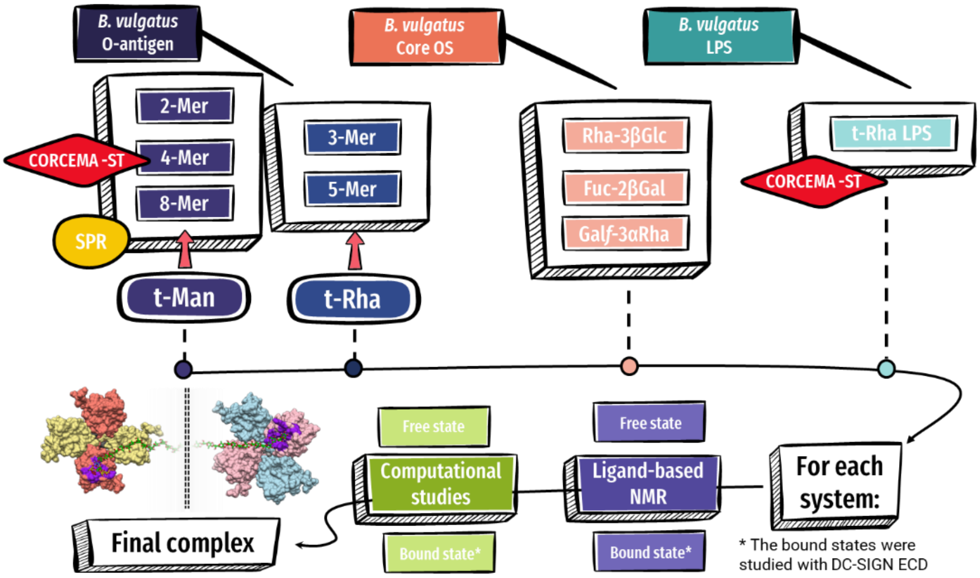


**Figure 4.1:** Workflow of the experimental activities.

This exploration highlights the importance of glycan complexity and variability in *B. vulgatus* LPS, identifying specific regions that interact with DC-SIGN in an anti-inflammatory manner. This contrasts with the previous chapter, where the interaction of LOS from *E. coli* with DC-SIGN was associated with a pro-inflammatory response.

## 4.2 DC-SIGN recognition of BVMPK LPS

Motivated by the previously investigated interaction between synthetic O-antigen oligomers and DC-SIGN,[221] a deeper exploration into the binding of DC-SIGN to constituents of BVMPK LPS (Figure 4.2) was initiated through competition Enzyme-linked immunosorbent assay (ELISA) experiments. For better comparison, two representative synthetic oligomers, namely the 4-Mer and 8-Mer (Figure 4.3), were also introduced to the plate.

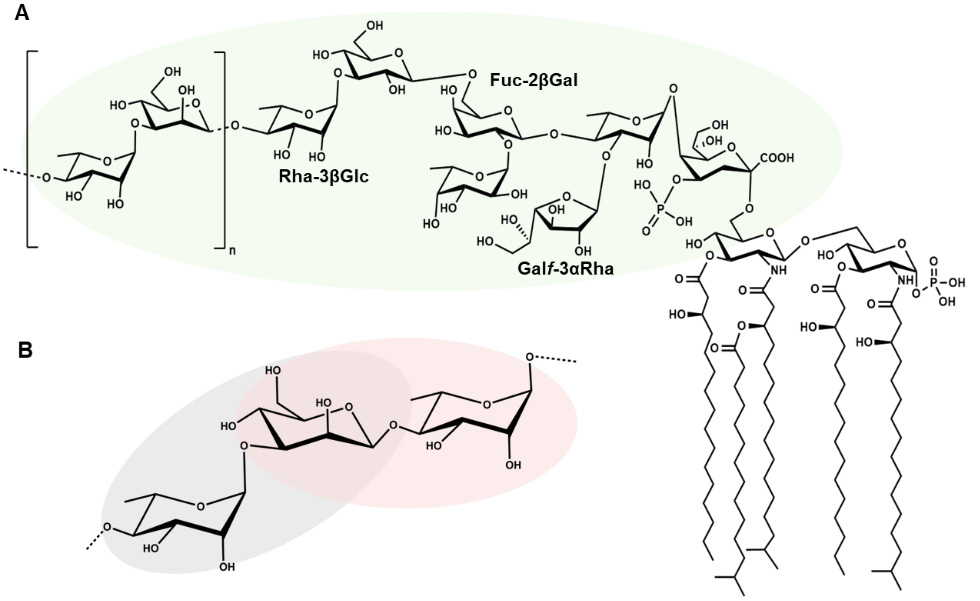


**Figure 4.2:** *Bacteroides vulgatus* mpk structure. A) BVMPK complete LPS structure. B) O-antigen disaccharide repeating unit constituted by an α-L-rhamnopyranose and a β-D-mannopyranose. Circled in green and grey/pink are the core and O-antigen LPS portions here evaluated in their interaction with DC-SIGN.

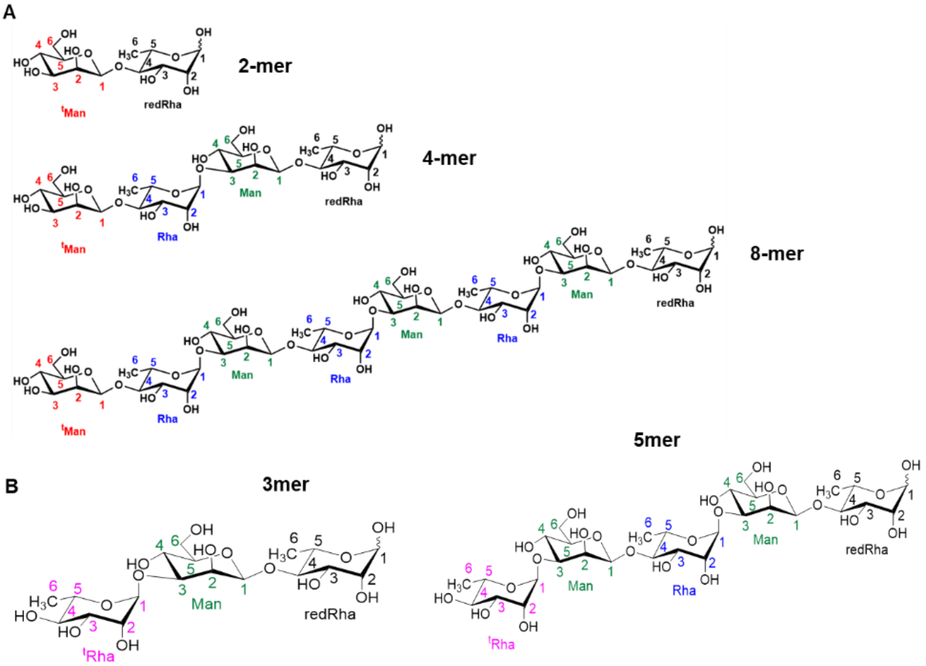


**Figure 4.3:**Synthetized BVMPK O-antigen studied ligand. A) 2-Mer, 4-Mer and 8-Mer possess a terminal Man residue. B) 3-Mer and 5-Mer possess terminal Rha residues.

Remarkably, ELISA assays revealed binding in all cases and both LPS components (the expected O-antigen and, more crucially, the core portion) exhibited the ability to inhibit LPS binding to DC-SIGN, following a typical dose-response curve, with the O-antigen showing a slightly better affinity if compared to the core oligosaccharide (Figure 4.4).

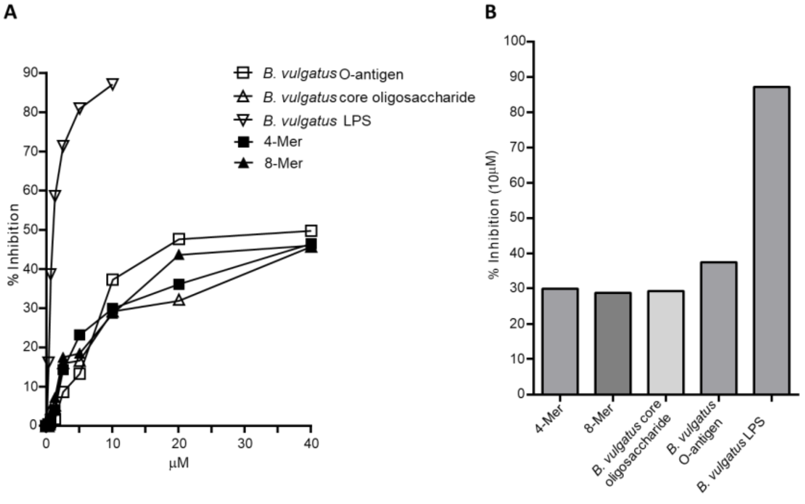


**Figure 4.4:** A) ELISA analysis of the competition binding of *B. vulgatus* core, O-antigen and LPS and the two synthetic oligomers, 4-Mer and 8-Mer to human lectin DC-SIGN. Competition experiments have been performed three times showing similar results; the graph shows data of one of these experiments. B) A focus on the results obtained at 10uM.

Considering these findings, the recognition of BVMPK LPS by the recombinant DC-SIGN ECD was assessed using a combination of synthetic approaches, NMR spectroscopy, computational, and biophysical techniques. Specifically, both the O-antigen and the core oligosaccharide were investigated in their interaction with DC-SIGN. In particular, different length O-antigen repeating units, containing either a terminal Man or a terminal Rha, were considered. Furthermore, three disaccharides composing the core oligosaccharide structure (excluding the Kdo), as well as the full LPS, containing a terminal Rha residue, three OPS repeating units and the core heptasaccharide, Kdo included, were studied (Figure 4.2).

### 4.2.1 BVMPK LPS O-antigen – DC-SIGN interaction

As BVMPK O-antigen was composed of [→3)-β-D-Man-(1→4)-α-L-Rha-(1→] disaccharide repeating units (RU), oligomers with different numbers of RU were sinthetized by the group of Prof. B. Yu (Shanghai Institute of Organic Chemistry, China), terminating both with Man[221] (Figure 4.3A) or Rha (Figure 4.3B).

These oligomers, containing one to four RU, a reducing L-Rha residue ($^{red}$Rha) and alternatively terminal Man (tMan, 2-Mer, 4-Mer and 8-Mer, Figure 4.3A) or terminal Rha residues (tRha, 3-Mer, 5-Mer, Figure 4.3B), were thoroughly evaluated in order to assess the role of Man and Rha in DC-SIGN recognition and binding process, and, also, the role, if any, played by the length of the saccharide chain.

#### 4.2.1.1 O-antigen oligomers with tMan – binding studies

Binding studies were performed on various *B. vulgatus* O-antigen oligomers, with the same outcomes. Binding epitopes were obtained through STD NMR[135, 138] experiments, revealing that predominant DC-SIGN-oligomers interactions occur at the terminal portion. Specifically, the STD NMR analysis of the mixture composed of DC-SIGN and 4-Mer (Figure 4.5) showed a binding epitope where the primary interaction was around the tMan residue, exhibiting the most significant STD enhancement. Additionally, a minor contribution to the binding from Rha was observed. Consequently, the interacting epitope was predominantly constituted from the terminal portion of the oligo, with the key region of tMan in contact with DC-SIGN. Indeed, the H3 proton of tMan unit exhibited the most pronounced STD effect and significant signals were also detected for protons H4 and H5, with STD effects of 98% and 73%, respectively. Additional contributions from H2 and H6 were also noted. In contrast, only H4 from the directly linked L-Rha unit was observable, displaying a low STD response. This observation indicated that the interaction with DC-SIGN primarily involved the tMan sugar, while no STD effect was observed for either the inner Man or the $^{red}$Rha unit, suggesting their solvent exposure. NMR binding studies using the longer ligand 8-Mer yielded a comparable STD NMR-derived epitope mapping, suggesting the crucial role of the tMan unit in the recognition process.

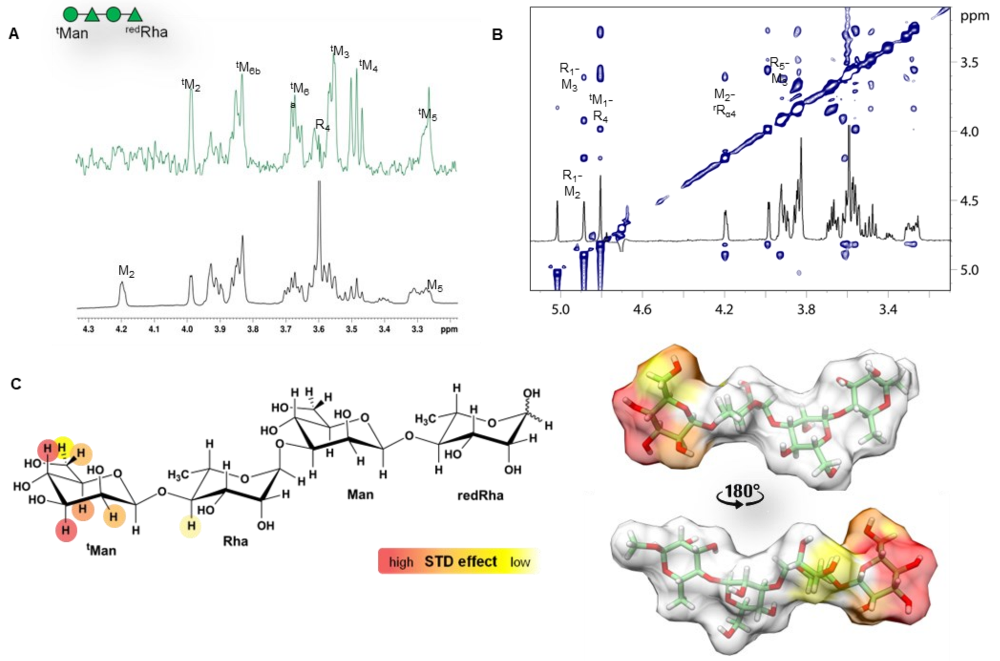


**Figure 4.5:** NMR interaction studies between 4-Mer and DC-SIGN. A) 4-Mer ligand schematic structure sketched according to SNFG nomenclature. STD NMR zoomed spectrum. $^1$H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:30 mixture of DC-SIGN:4-Mer. B) tr-NOESY NMR on the 1:20 mixture of DC-SIGN:4-Mer. C) 2D and 3D representation of the interacting epitope map of the interaction between 4-Mer and DC-SIGN derived from STD NMR data and computational studies.

The bioactive conformation assumed by the ligand upon binding was investigated through tr-NOESY experiments (Figure 4.5B). A comparison of spectra obtained in the free state[221] and bound state revealed no significant changes, confirming that there was no conformer selection of the glycan chain when complexed with DC-SIGN (Figure 4.5C and Table 4).

**Table 4:** DC-SIGN: 4-Mer experimental and calculated key *inter*-residue distances

| | **Exp. (tr-NOESY)** | **Calc. (MD derived)** |
|---|---|---|
| **$R_1$-$M_3$** | 2.97 | 3.43± 0.34 |
| **$R_1$-$M_2$** | 2.62 | 2.68 ± 0.38 |
| **$^tM_1$-$R_4$** | 2.54 | 2.48 ± 0.24 |
| **$M_2$-$R_{α4}$** | 2.73 | 3.13 ± 0.37 |
| **$R_5$-$M_3$** | 2.62 | 3.53 ± 0.49 |

Further insights into the molecular recognition features were unveiled by combining NMR-derived binding data with computational studies. Energetically accessible conformational regions of glycosidic torsion angles (α-L-Rha-(1→3)-β-D-Man and β-D-Man-(1→4)-α-L-Rha), constituting BVMPK LPS O-antigen, were determined (Figure 4.6).

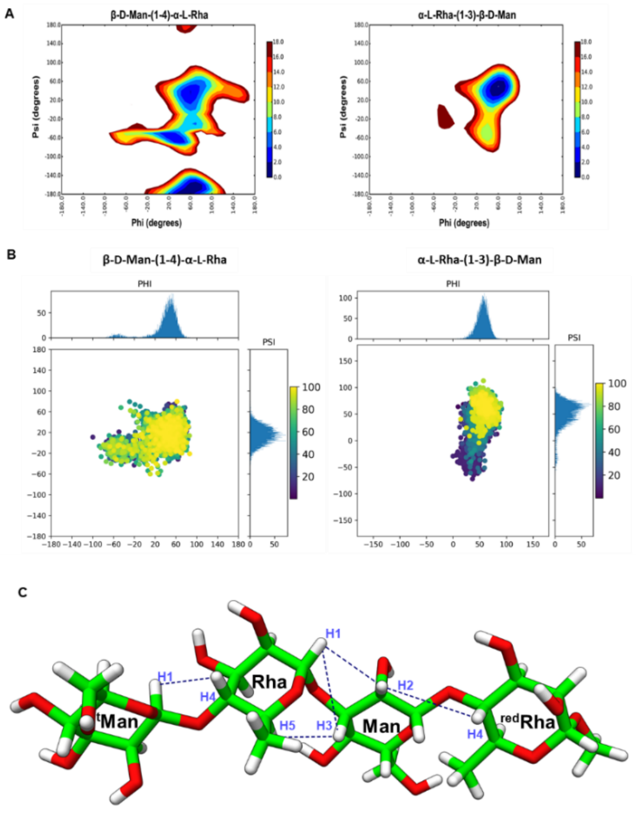


**Figure 4.6:** A) Adiabatic energy maps of the two different glycosidic linkages present in the O-chain of BVMPK. B) Analysis of the bound state dihedral angles represented as a scatter plot of the ϕ torsion against ψ during the time with the relative histograms to represent the most populated energy. C) Calculated distances from the DC-SIGN:4Mer MD simulation.

Then, starting complexes were generated by manually docking the oligomers into the DC-SIGN monomeric subunit (PDB:1SL4).[98] The tMan guided the superimposition of the oligomers in the binding site, using the $GlcNAc_2Man_3$ unit as reference

(PDB:1SL4). The complex generation and subsequent optimization were conducted using the software Maestro.[222, 223] Similar results were obtained studying all the complexes; here, the behavior of DC-SIGN/4Mer complex is described as example.

Specifically, Man could adopt two different orientations in the binding pocket:[205, 224, 225] a binding mode A (**BmA**) with the 6-OH pointing towards the adjacent shallow groove (Figure 4.7), and, after rotating the ligand through the Man residue by 180°, a second binding mode B (**BmB**) obtained with the axial 2-OH pointing towards the shallow groove (Figure 4.8).

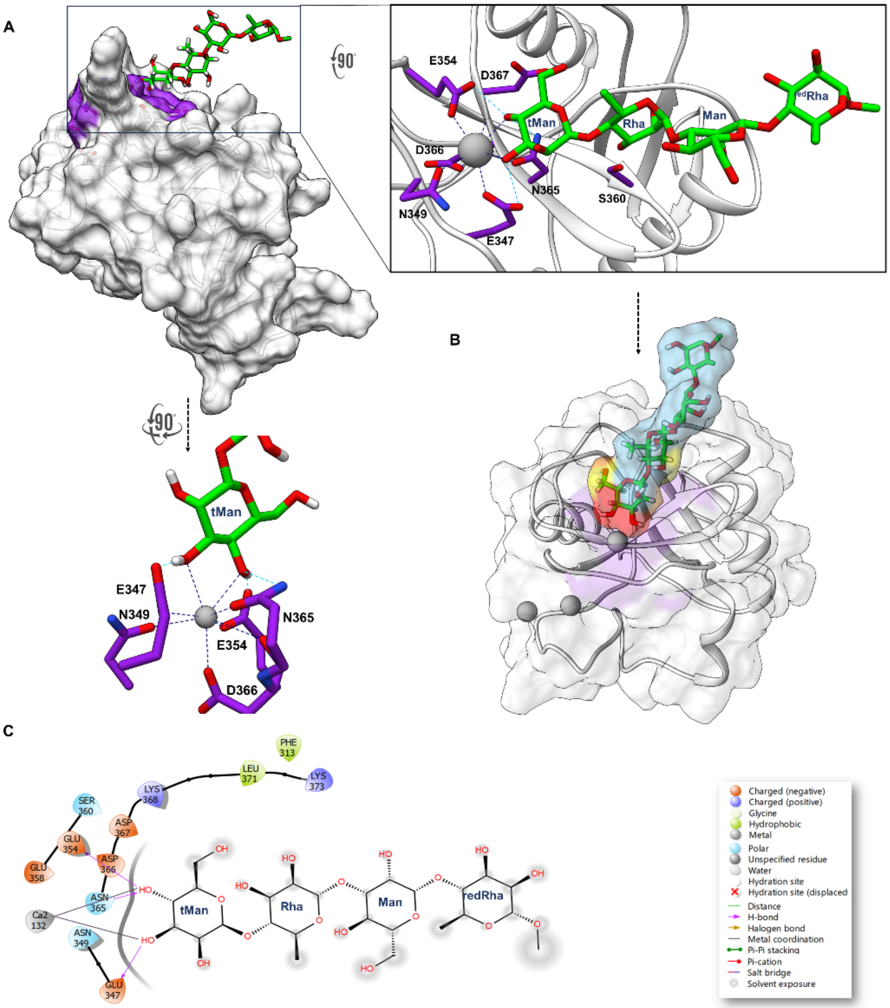


**Figure 4.7:** MD simulation analysis of **BmA** for the DC-SIGN:4-Mer complex A) 3D model of DC-SIGN – 4-Mer complex with the protein amino acids involved in the interaction highlighted in purple. Close-up view of a representative pose from the most populated MD family (top right) and the tMan bound to the principal $Ca^{2+}$. The main amino acids involved in the binding are coloured in purple and the $Ca^{2+}$ ion in grey. The ligand is colored according to SNFG. $Ca^{2+}$ coordination bonds are shown as dashed blue lines while the observed H-bond interactions are indicated by dashed cyan lines. B) 3D representation of the interaction with the ligand surface colored according to the STD NMR results. C) BmA two-dimensional plot representing the interactions between 4-Mer and DC-SIGN binding pocket residues.

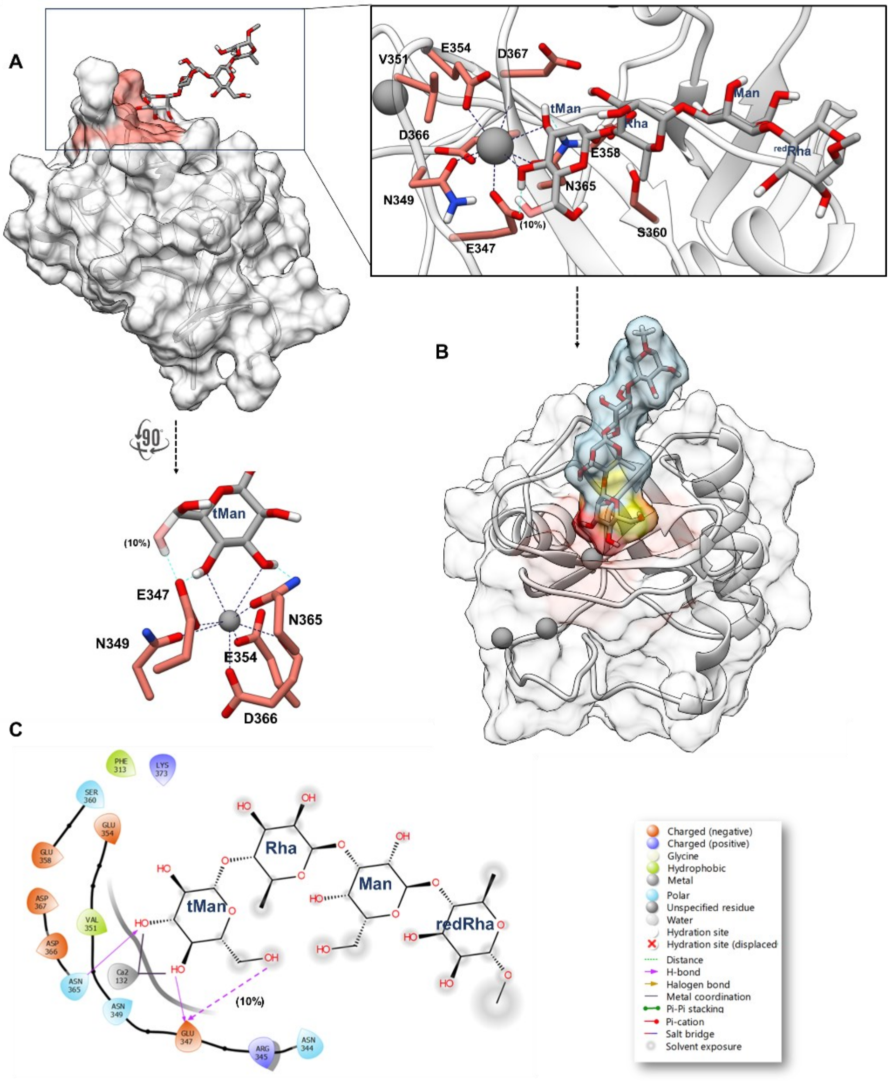


**Figure 4.8:** MD simulation analysis of the **BmB**, interaction between 4-Mer and DC-SIGN. A) 3D model of DC-SIGN – 4-Mer complex with the protein amino acids involved in the interaction highlighted in salmon. Close-up view of a representative pose from the most populated MD family (top right) and the tMan bound to the principal $Ca^{2+}$. The main amino acids involved in the binding are coloured in purple and the $Ca^{2+}$ ion in grey. The ligand is colored according to SNFG. $Ca^{2+}$ coordination bonds are shown as dashed blue lines while the observed H-bond interactions are indicated by dashed cyan lines. B) 3D representation of the interaction with the ligand surface colored according to the STD NMR results. C) BmB two-dimensional plot representing the interactions between 4-Mer and DC-SIGN binding pocket residues.

Both accommodations, **BmA** and **BmB**, were subjected to extensive MD simulations, to validate experimental results. Glycosidic torsion angles $\phi$ and $\psi$ were sampled along the MD simulations in both free[221] and bound states to further evaluate the conformational behavior of BVMPK LPS O antigen oligomers in complex with DC SIGN. Ensemble average interproton distances were extracted and translated into NOE contacts using a full-matrix relaxation approach. Experimental and calculated key *inter*-residue distances have revealed an excellent agreement between the experimental and calculated data (Table 4). Therefore, no significant differences in the ligands' extended conformation were found upon binding (Figure 4.6). The most representative poses for the two binding modes, **BmA** and **BmB**, in the DC-SIGN/4mer complex were chosen through cluster analysis of the MD simulations and subjected to analysis using the CORCEMA-ST program.[184] The results suggested that in both **BmA** and **BmB** complexes, tMan played a central role in the interaction, with a substantial contribution from H3 and H4, along with evident involvement of the other ring protons (Figure 4.9). Specifically, in both cases, the highest predicted STD value was for tMan H3, consistent with the experimental findings. Conversely, no saturation was anticipated for the other sugar units, except for the internal Rha, which contributed to the binding with significant involvement only for the H4 proton. Moreover, in accordance with the STD NMR results, no contribution was predicted for Man and $^{red}$Rha residues. However, overall, the calculated STD values for each individual binding mode did not entirely align with the experimental STD NMR results, confirming the coexistence of two binding modes.

Considering a theoretical bimodal binding equilibrium, involving both **BmA** and **BmB**, the contribution of each binding mode to the interaction was discerned, as detailed in a prior publication.[204] The observed NMR spectrum is essentially considered as the ensemble spectrum, resulting from the weighted combination of the STD signals of both binding modes contributing to the interaction. By identifying the ensemble of ligand-bound modes best aligned with the experimental data, the quality of CORCEMA-ST predictions was assessed using the R-NOE factor. The results showed a satisfactory agreement between theoretical and experimental outcomes. The equilibrium of the two bound conformations in solution was characterized by **BmA** as the major binding mode (70% of contribution), and **BmB** as the minor binding mode (30% of contribution) (NOE R-factor = 0.2; Table 5 and Figure 4.9).

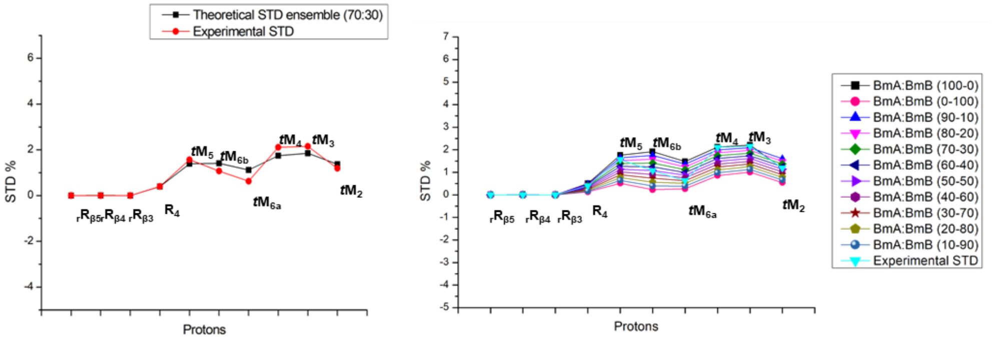
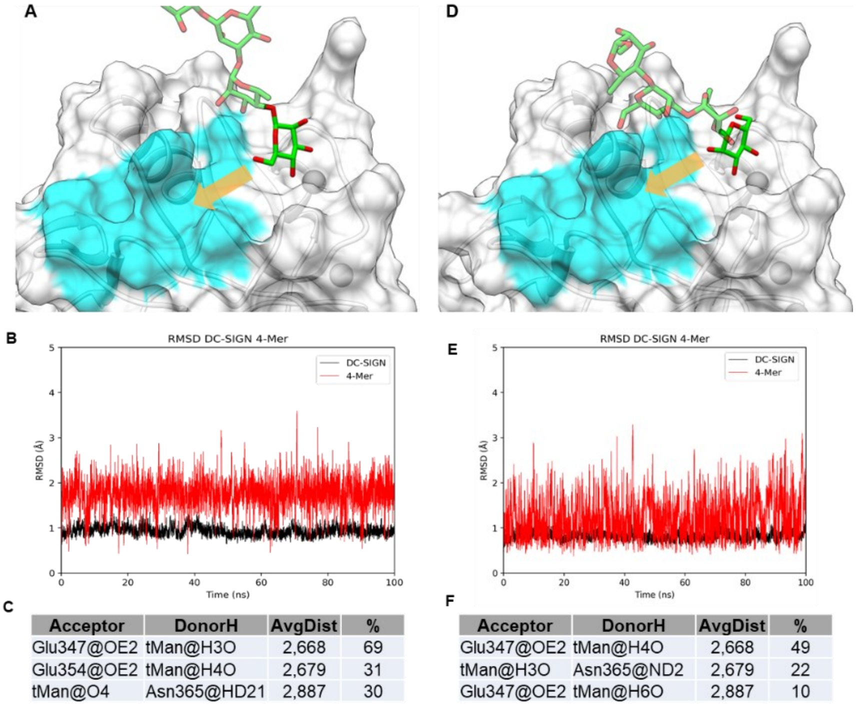


**Figure 4.9:** CORCEMA-ST analysis derived from the most representative poses of the DC-SIGN/4-Mer complex. The analysis of the contribution of **BmA** and **BmB** to the ensemble was performed by analysing the quality of the prediction by NOE R-factor, being determined that the contribution to the ensemble was 70% for **BmA** and 30% for **BmB**, same results as obtained from the analysis of the MD simulation. the most, demonstrating the best fit between the ensemble theoretical (blue) and experimental (red) STD data.

**Table 5:** NOE R-factor as a function of the proportional weight of **BmA** and **BmB.**

| BmA | BmB | NOE R-factor |
|---|---|---|
| **100** | 0 | 0.3199 |
| **90** | 10 | 0.2646 |
| **80** | 20 | 0.2252 |
| **70** | 30 | 0.2006 |
| **60** | 40 | 0.2257 |
| **50** | 50 | 0.2655 |
| **40** | 60 | 0.3209 |
| **30** | 70 | 0.3853 |
| **20** | 80 | 0.4548 |
| **10** | 90 | 0.5275 |
| **0** | 100 | 0.6021 |

Representative poses from the energetically favorable and highly populated clusters of the 4-Mer in the **BmA** and **BmB** binding modes (Figure 4.7 and Figure 4.8) illustrate the involvement of the tMan residue in coordinating with $Ca^{2+}$ ion through OH at positions 3 and 4 in the canonical binding mode.[98, 205] RMSD calculations on protein backbone and ligand residues, with the initial frame as a reference, indicated the stability of the complexes, with a similar accommodation of the oligomers in the binding pocket (Figure 4.10). During the entire MD simulations, the tMan unit maintained canonical coordination with the $Ca^{2+}$ ion through positions 3 and 4, while the remaining part of the ligand was solvent exposed.



C

| Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| Glu347@OE2 | tMan@H3O | 2,668 | 69 |
| Glu354@OE2 | tMan@H4O | 2,679 | 31 |
| tMan@O4 | Asn365@HD21 | 2,887 | 30 |

F

| Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| Glu347@OE2 | tMan@H4O | 2,668 | 49 |
| tMan@H3O | Asn365@ND2 | 2,679 | 22 |
| Glu347@OE2 | tMan@H6O | 2,887 | 10 |

**Figure 4.10:** Schematic representation of the two possible binding mode and MD simulation analysis of the two binding modes for the DC-SIGN:4-Mer complex. A) **BmA** with the 6-OH pointing towards the adjacent shallow groove (orange arrow). B) Protein (black) and ligand (red) RMSD of the **BmA**. The ligand RMSD was calculated in reference to the protein. C) **BmA** protein-ligand H-bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD. D) **BmB** with the 2-OH pointing towards the adjacent shallow groove (orange arrow). E) Protein (black) and ligand (red) RMSD of the **BmB**. The ligand RMSD was calculated in reference to the protein. F) **BmB** protein-ligand H-bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD.

Molecular interactions within the complexes were monitored and compared to NMR-derived experimental data, particularly focusing on the 4-Mer in **BmA** and **BmB**, with similar results for other oligomers (data not shown). The $Ca^{2+}$ ion, positioned in the main DC-SIGN binding site and accommodated by Glu347, Pro348, and Asn349, played a pivotal role in carbohydrate recognition and binding, coordinating with tMan through the oxygen atoms of 3OH and 4OH. Distinctions between **BmA** and **BmB** binding modes were observed. In **BmA**, 3OH tMan acted as a hydrogen donor in interaction with Glu347 carboxylic moiety, while tMan 4OH served as a hydrogen donor for the interaction with Glu354 carboxylic moiety and an acceptor for Asn365 (Figure 4.7 and Figure 4.10B-C). Conversely, in **BmB**, due to a 180° rotation, significant shifts in binding interactions occurred, with tMan 3OH now acting as a hydrogen acceptor for Asn365, and tMan 4OH serving as a hydrogen donor in interaction with Glu347. In **BmB**, Glu347 also engaged in a hydrogen bond with tMan 6OH during approximately 10% of the simulation time, aligning with the STD NMR results and providing further insight into the substantial contribution of H6 tMan to the interaction (Figure 4.8 and Figure 4.10E-F).

The internal Rha unit did not establish stable interactions with DC-SIGN, during the MD simulation, in agreement with experimentally observed low STD contribution and CORCEMA-ST calculations (Figure 4.9).

Importantly, the crucial role of $Ca^{2+}$ in mediating the binding was further confirmed by NMR. Upon the addition of perdeuterated EDTA to the sample, a complete fading of STD NMR signals was observed, indicating that no interaction occurred in the absence of calcium (data not shown).

In conclusion, a 3D complex of BVMPK O-antigen accommodated in the DC-SIGN binding site has been proposed, highlighting the essential role of the tMan unit for recognition and binding in both **BmA** and **BmB** (Figure 4.7 and Figure 4.8), giving the same result also for longer oligomer. Furthermore, SPR experiments were also performed to validate the models proposed by NMR experiments and MD calculation. SPR competition experiments were run using oligomers containing from one to 8 repeating units (2-Mer, 4-Mer, 8-Mer and 16-Mer) to evaluate their affinity towards DC-SIGN (Figure 4.11).[201]

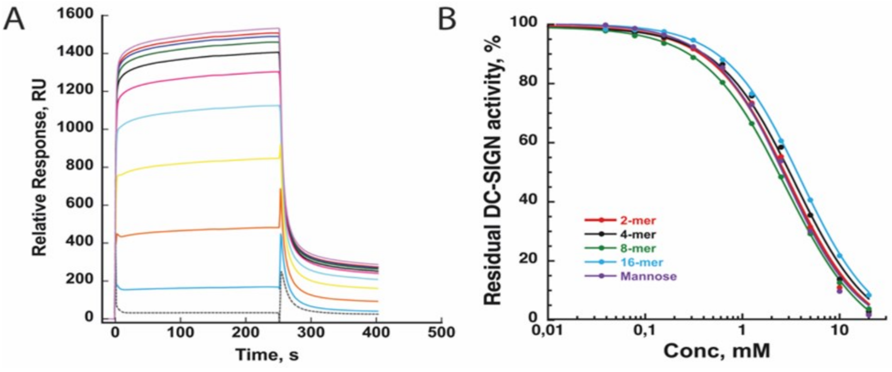


**Figure 4.11:** Inhibition of DC-SIGN interaction over BSA-Man surfaces with different oligomer of the O-Antigen repeating unit. A) Representative sensorgrams of the interaction and its inhibition with 2-Mer with increasing concentrations, by a multiplication factor of 2, starting from 40 µM up to 20 mM. B) Inhibition curve obtained for 2-Mer, 4-Mer, 8-Mer and 16-Mer.

As results, independently from 2-Mer to 16-Mer tested oligomers, the IC$_{50}$ obtained, were always in the same range, between 3 to 4.2 mM. Despite the increasing repeating units, the contribution to the binding was enterely given by the terminal unit. Moreover, the 2-Mer IC$_{50}$ of 3.3 mM was the same as a single Man in SPR competition test,[226] underlying the central role of tMan in the binding.

### 4.2.1.2 O-antigen oligomers with tRha – binding studies

As mentioned before, oligomers terminating with L-Rha were also *ad-hoc* synthesized (3-Mer and 5-Mer, Figure 4.3), to study their interaction with DC-SIGN, following the same approach as for tMan oligomers. Interestingly, NMR and biophysical techniques detected no binding with DC-SIGN (Figure 4.12). In order to further prove the absence of recognition, D-Man monosaccharide (in the same 1:30 ratio) was added to the sample and run STD NMR spectra. It was possible to observe the presence of STD NMR signals deriving from the monosaccharide, confirming the absence of the binding for 3-Mer and 5-Mer and, also, the functionality of the protein (data not shown).

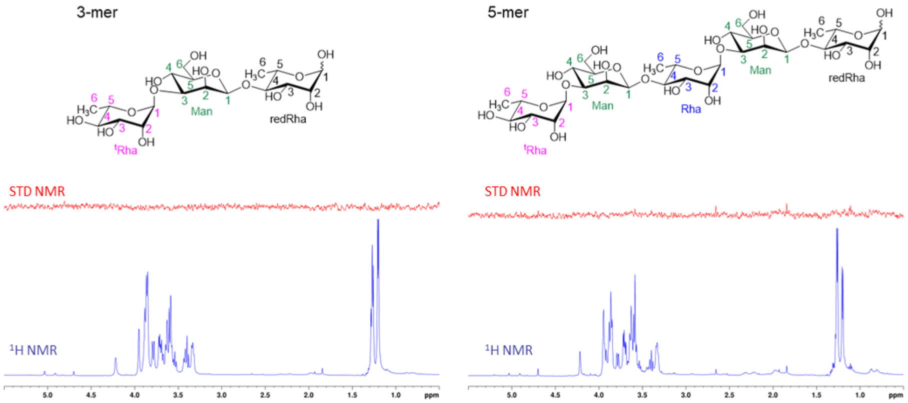


**Figure 4.12:** STD NMR studies of tRha oligosaccharides 3-Mer and 5-Mer with DC-SIGN where no interaction was observed.

### 4.2.2 BVMPK LPS core - DC-SIGN interaction

The presence in *B. vulgatus* core LPS of uncommon sugars like Fuc (Fucose) and Gal*f* (Galactofuranose) increased the interest in unveiling their interaction and recognition by DC-SIGN. Particularly, the three composing disaccharides, resulting from the dissection of the core, excluding the Kdo, were studied, after their *ad-hoc* synthesis, performed by the group of Prof. S. Kulkarni (Indian Institute of Technology, Bombay, India) (Figure 4.2).

The three disaccharides here analyzed were: (i) Rha-3βGlc, (ii) Fuc-2βGal and (iii) Gal*f*-3αRha (Figure 4.2A).

NMR binding studies (STD and tr-NOESY) on *B. vulgatus* core fragments showed how Fuc-2βGal disaccharide was the unique portion recognized by DC-SIGN (Figure 4.13). The highest STD NMR enhancement was attributed to the H1 of the Fuc residue, proposed as the main anchor for the recognition of the Fuc-2βGal disaccharide by DC-SIGN.

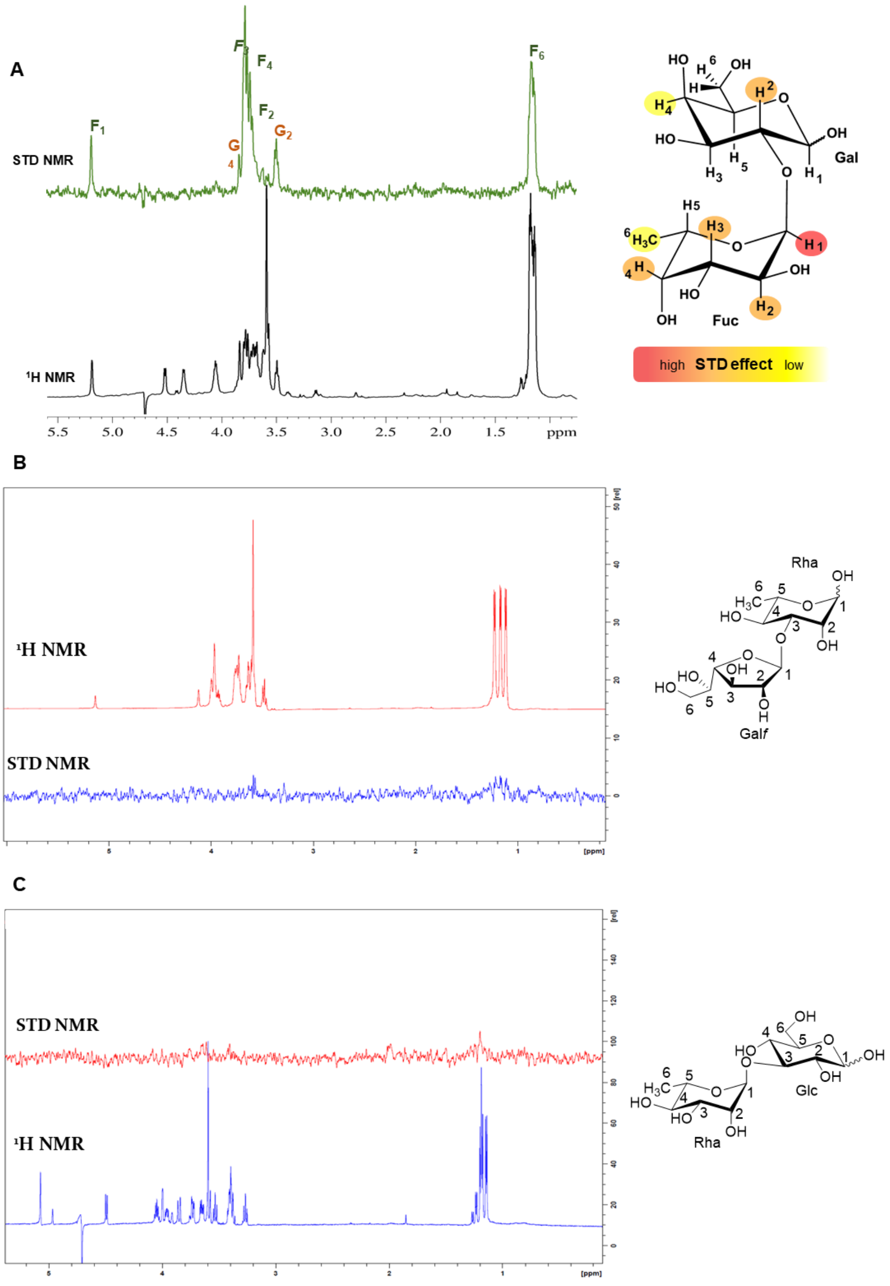
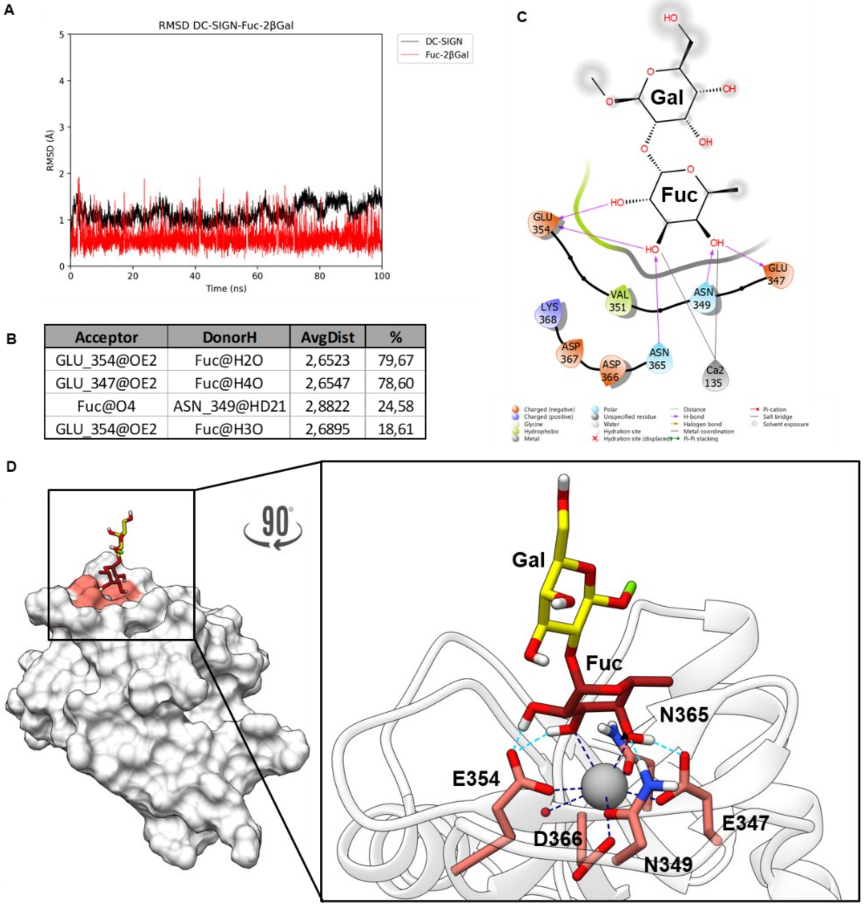


**Figure 4.13:** A) $^1$H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:20 of the DC-SIGN: Fuc-2βGal mixture. Interacting epitope map between Fuc-2βGal and DC-SIGN derived from STD-NMR data. B) Gal*f*-3αRha ligand structure and $^1$H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:20 mixture with DC-SIGN. C) Rha-3βGlc $^1$H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:20 mixture.

Additionally, the disaccharide was prepared by using Glycam builder utility (www.glycam.org)[227] to be modelled into the protein binding pocket. Docking studies were performed again to obtain a better starting point for the MD calculations. Therefore, a representative pose from the docking studies was used as initial step of MD simulation trajectory and the analysis of the MD confirmed the stability of the complex (Figure 4.14A).



| Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| GLU_354@OE2 | Fuc@H2O | 2,6523 | 79,67 |
| GLU_347@OE2 | Fuc@H4O | 2,6547 | 78,60 |
| Fuc@O4 | ASN_349@HD21 | 2,8822 | 24,58 |
| GLU_354@OE2 | Fuc@H3O | 2,6895 | 18,61 |

**Figure 4.14:** Computational studies of Fuc-2βGal interacting with DC-SIGN A) Protein (black) and ligand (red) RMSD. The ligand RMSD was calculated in reference to the protein. B) Protein-ligand H-bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD. C) Two-dimensional plot representing the interactions between Fuc-2βGal and the binding pocket residues of DC-SIGN. D) Representation of the core hexasaccharide onto DC-SIGN binding pocket with the Fuc residue coloured in red and the Gal residue in yellow. Fuc-2βGal is accommodated in the binding pocket

interacting through Fuc residue. Coordination interactions with $Ca^{2+}$ ion are depicted in dark blue while H-bond interactions with the protein are coloured in cyan.

The interaction was stabilized by the coordination of the Fuc residue with the $Ca^{2+}$ ion through the canonical 3OH-4OH coordination. Additionally, the Fuc residue, showed a stable hydrogen bonding network, interacting with DC-SIGN in the $Ca^{2+}$ dependent binding pocket through the hydroxyl moieties at positions 2, 3 and 4 (Figure 4.14). Actually, 4-OH acted as an H-bond donor to E347 and acceptor from N349, 3-OH was found as a donor to E354 and 2OH acted as donor to E354.

### 4.2.3 BVMPK full core LPS- DC-SIGN interaction

Finally, BVMPK LPS moiety containing the full core, three OPS repeating units and a tRha unit as terminal residue (tRha-LPS, Figure 4.15A) was evaluated in the interaction with DC-SIGN. STD-NMR experiments (Figure 4.15) supported the above findings and confirmed how the OPS portion containing the terminal tRha of tRha-LPS did not interact with DC-SIGN. As for the core region, Fuc was the main sugar unit accommodated in the protein binding site; interestingly, also the Gal*f* residue was significantly involved (Figure 4.15B-C), and, in addition, also Gal and Kdo units were located nearby the protein surface. The recognition took place mainly through the Fuc unit, in agreement with the previous binding experiments described for the Fuc-2βGal core disaccharide. The strongest STD effect was attributed to H1 to H4 protons of Fuc (above 70%); in addition, almost all Gal*f* protons, mainly H2 Gal proton and the side chain protons of Kdo (H6, H7 and H8), gave moderate to good STD signals, highlighting their location inside the DC-SIGN binding pocket. So, these results showed the key involvement of Fuc in the binding and the presence inside the binding pocket also of Gal, Gal*f* and Kdo; likely the rest of the core was solvent exposed.

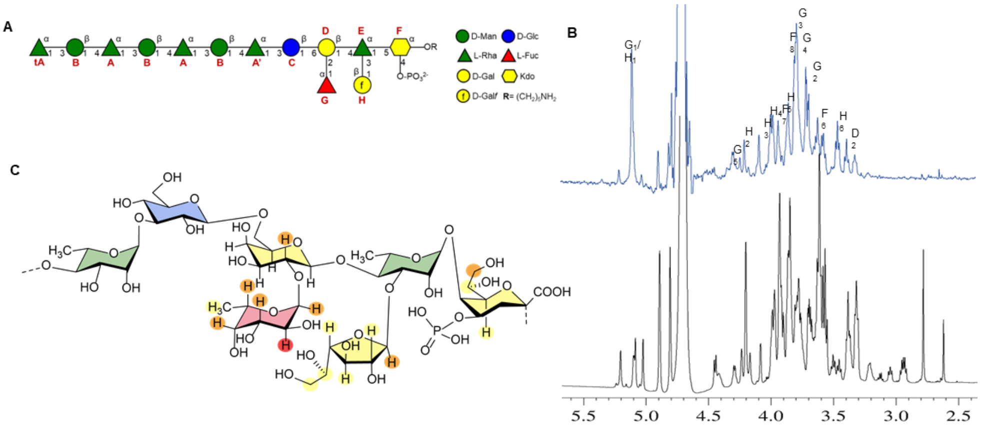


**Figure 4.15:** NMR interaction studies between tRha-LPS and DC-SIGN. A) tRha-LPS tridecasaccharide schematic structure sketched according to SNFG nomenclature. B) STD NMR zoomed spectrum. At the bottom $^1$H NMR reference spectrum whereas on top 1D STD NMR spectrum of the 1:30 mixture of DC-SIGN: tRha-LPS. C) 2D representation of the interacting epitope map of the interaction between tRha-LPS and DC-SIGN derived from STD NMR data.

Then, the bioactive conformation of the tRha-LPS was inferred by tr-NOESY experiments, performed on a DC-SIGN: tRha-LPS mixture (Figure 4.16), allowing to derive the key *inter*-protons distances, crucial to evaluate the conformational behavior in the bound state (Table 6).

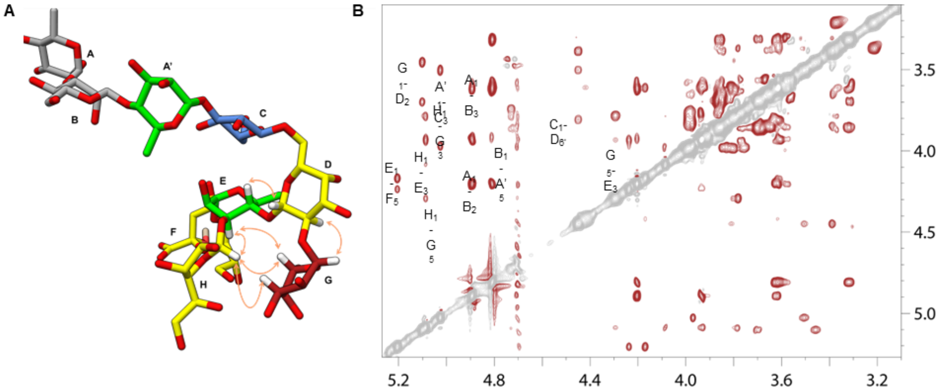


**Figure 4.16:** A) Schematic representation of the bioactive conformation with the core oligosaccharide sugars coloured in according to the SNFG nomenclature and the OPS sugars in grey. B) tr-ROESY NMR on the 1:30 mixture of DC-SIGN:tRha-LPS.

**Table 6:** DC-SIGN: core LPS experimental and calculated key *inter*-residue distances.

| | **Exp. (tr-NOESY)** | **Calc. (MD derived)** |
|---|---|---|
| **G1-D2** | 2.65 | 2.84 ± 0.22 |
| **H1-G3** | 2.95 | 2.82 ± 0.65 |
| **H1-E3** | 2.76 | 2.73 ± 0.34 |
| **H1-G5** | 3.25 | 2.84 ± 0.42 |
| **A'1-C3** | 2.76 | 2.92 ± 0.29 |
| **A1-B2** | 2.19 | 2.10 ± 0.24 |
| **A1-B3** | 2.74 | 2.68 ± 0.26 |
| **B1-A'5** | 3.80 | 4.13 ± 0.15 |
| **D1-E4** | 2.89 | 2.80 ± 0.17 |
| **C1-D6** | 3.02 | 3.74 ± 0.21 |
| **C1-D4** | 3.64 | 3.50 ± 0.30 |
| **C1-D6'** | 2.70 | 3.02 ± 0.22 |
| **G5-E3** | 3.06 | 3.38 ± 0.34 |

The presence of uncommon sugars, from a computational point of view, like Gal*f* and Kdo, required as first step their parametrization, using AMBER18 package,[228] and, then, they were used for MD simulations, in order to depict a 3D model of tRha-LPS. First, the disaccharides constituting the core region (Figure 4.6) were already built to evaluate the energetically accessible conformational regions *via* MM calculations using standard protocols. The adiabatic energy maps, which illustrate the $\phi/\psi$ and $\phi/\omega$ glycosidic torsion angles (Figure 4.17), confirmed that the global energy minimum aligned with the *exo*-anomeric effect.

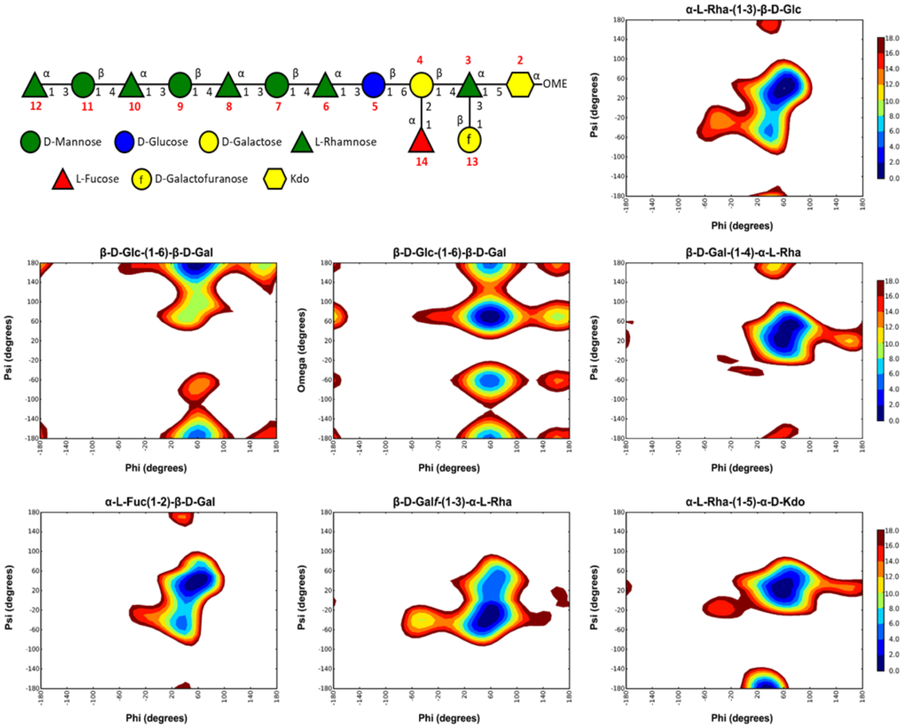
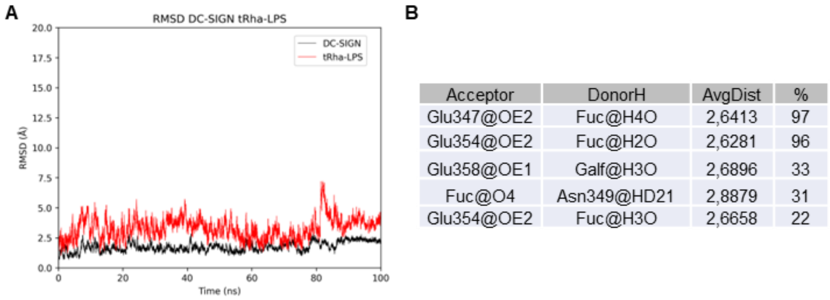


**Figure 4.17:** Adiabatic energy maps, illustrating the glycosidic torsion angles ϕ/ψ and ϕ/ω, of the basic constituent disaccharides of tRha-LPS, obtained by molecular mechanics simulations (cf. also Figure 4.6A).

Thus, the tRha-LPS tridecasaccharide in its bioactive conformation, was manual docked into DC-SIGN binding site (PDB:1SL5),[229] to run MD simulation in explicit solvent with AMBER. RMSD calculations conducted on the protein backbone and ligand residues, using the first frame as a reference, provided evidence of the complex stability (Figure 4.18).



| Acceptor | DonorH | AvgDist | % |
|---|---|---|---|
| Glu347@OE2 | Fuc@H4O | 2,6413 | 97 |
| Glu354@OE2 | Fuc@H2O | 2,6281 | 96 |
| Glu358@OE1 | Galf@H3O | 2,6896 | 33 |
| Fuc@O4 | Asn349@HD21 | 2,8879 | 31 |
| Glu354@OE2 | Fuc@H3O | 2,6658 | 22 |

**Figure 4.18:** DC-SIGN: tRha-LPS complex MD simulation analysis. A) RMSD of DC-SIGN (black) and tRha-LPS (red) being the ligand RMSD calculated in reference to the protein. B) Protein-ligand H-bonds described as acceptor, donor, distance and stability (in percentage) during the 100 ns MD simulation.

The glycosidic torsion angles were sampled during the MD simulations in both free and bound states, to assess the conformational behavior (data not shown). The corresponding average distances obtained for the simulation from ⟨$r^{-6}$⟩ values were compared to those collected experimentally and an excellent accordance was found (Table 6). MD results showed that trajectories matched those predicted by the MM calculation, with the glycosidic linkages adopting $\phi$ values in accordance with the *exo*-syn anomeric conformation. Moreover, MD simulation confirmed the NMR binding data, showing how the Gal*f* residue was located close to the protein surface, and therefore established interactions with DC-SIGN, with most of the Kdo unit, the rest of the core and the OPS solvent exposed.

Furthermore, a detailed analysis of computational studies showed how the Fuc residue was accommodated in the protein binding site and acted as anchor, establishing key interactions with DC-SIGN and stabilizing the whole ligand by coordination with the $Ca^{2+}$ ion in the canonical way, through hydroxyl moieties at positions 3 and 4 (Figure 4.15). Additionally, the Fuc residue was stabilized by further H-bonds interactions, with OH moieties at positions 2, 3 and 4 establishing strong interactions with DC-SIGN. The most stable interaction was observed between Fuc 4-OH as H-bond donor with E347 (97% of the MD, Figure 4.18), and as H-bond acceptor with N349 (31%); E354 was found as H-bond acceptor in both, the interaction with 2-OH (96%) and 3-OH (22%). In addition, the conformation adopted by tRha-LPS supported the location of Gal*f* close to the binding pocket in the 3D complex, confirmed by both STD and trNOESY data. Interestingly, MD simulation also proved that Gal*f* 3-OH acted as H-bond donor to E358 (Figure 4.18) and that the conformational features of tRha-LPS also moved H6, H7 and H8 of the Kdo residue, close to the binding pocket, supporting the experimental STD effects. Furthermore, CORCEMA-ST analysis was also run on the DC-SIGN: tRha-LPS complex finding a good agreement between the experimental and the predicted calculations (Figure 4.19).

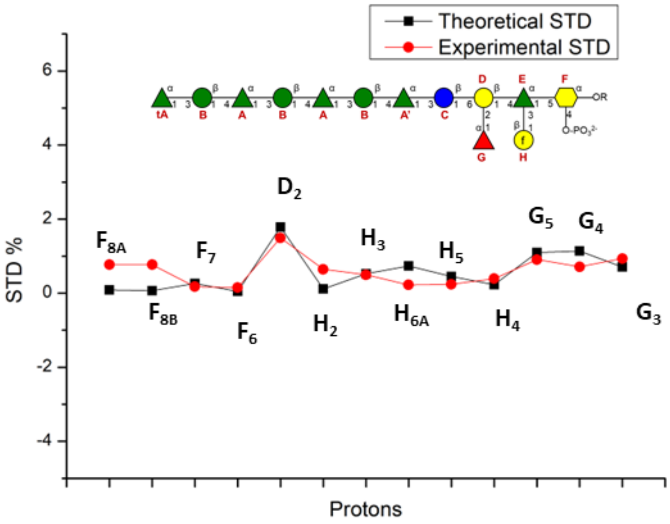


**Figure 4.19:** CORCEMA-ST analysis derived from the most representative poses of the DC-SIGN/tRha-LPS complex.

Hence, computational results fully matched and integrated NMR data, and provided information on the further accommodation of *Bacteroides* core LPS inside DC-SIGN binding pocket.

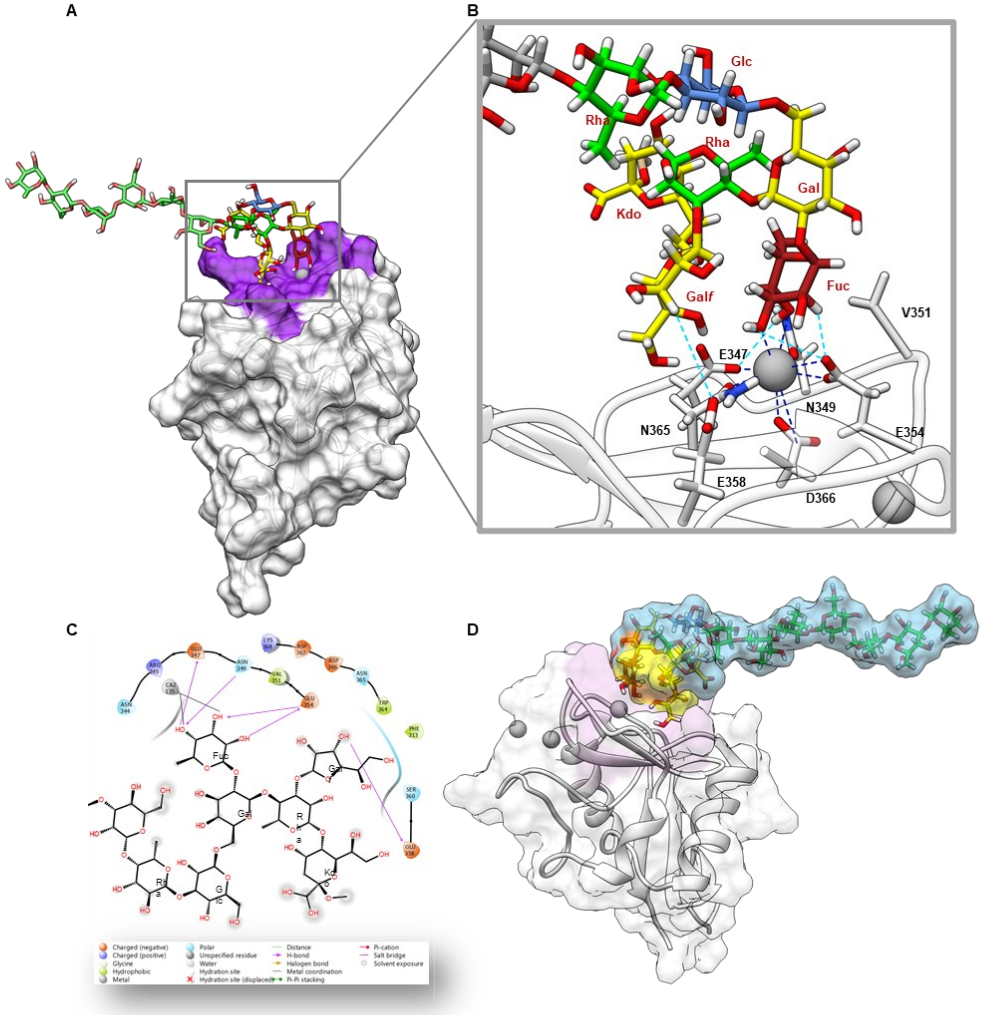


**Figure 4.20:** DC-SIGN: tRha-LPS complex MD simulation analysis. A) 3D model of DC-SIGN – tRha-LPS complex with the protein amino acids involved in the interaction highlighted in purple. B) Close-up view of a representative pose from the most populated MD family. In grey the $Ca^{2+}$ ions. The ligand is coloured according to SNFG nomenclature. $Ca^{2+}$ coordination bonds are shown as dashed blue lines while the observed H-bond interactions are indicated by dashed cyan lines. C) Two-dimensional plot representation of the most stable interactions between tRha-LPS and DC-SIGN binding pocket residues. D) 3D representation of the DC-SIGN: tRha-LPS complex with the ligand surface coloured according to the STD NMR results.

## 4.3 Discussion

In the evolving landscape of bacterial research, there has been a growing interest in understanding the role and immune perception of bacterial envelope glycoconjugates. These compounds, acting as MAMPs, play a crucial role in triggering the host immune response.[60, 72, 88, 230] However, numerous questions remain unanswered, particularly in the context of the GM and its modulation of the host immune system.[216] The host's ability to discern between beneficial and harmful bacterial species suggests that commensal bacteria express chemically distinct MAMPs, thereby modulating the immune response in various ways. The immune activation driven by LPS is highly structure-dependent, exhibiting a spectrum from potent to moderate stimulation and even inhibition of the immune response in antagonistic LPS.

CTLs, a diverse family of antigen uptake membrane receptors, located on the plasma membrane of APCs (cf. §1.5.1.1), emerge as crucial players in the recognition of commensals. The interaction between host lectins and commensals permits to promote mutualism and prevent systemic inflammation. The CTL DC-SIGN plays a multifaceted role in immune-related functions, encompassing antigen uptake, processing, immune regulation, and serving as an immune homeostatic receptor. Furthermore, DC-SIGN-dependent cross-presentation is enhanced by the simultaneous triggering of TLRs, being described as its ability to modify TLR signaling in a pro- or anti-inflammatory way depending on the ligand recognized (cf. Chapter III).[100, 220] Due to the DC-SIGN dual role in mediating both tolerance and immune response, depending on the context, pathogens and tumors may exploit this duality by upregulating the expression of DC-SIGN ligands, using DC-SIGN-dependent tolerogenic signaling to evade the immune system.[220] DC-SIGN preferentially binds to Fuc-exposing[101] and Man-exposing ligands, stimulating the TLR-induced cytokine response. Both, Man and Fuc ligands stimulate IL-10 secretion, but while Fuc ligands stimulate anti-inflammatory response, reducing LPS-induced IL-6 and IL-12 secretion, Man ligands act in a pro-inflammatory way, increasing both interleukins production.[102] Nevertheless, recently, DC-SIGN ability to recognize GlcNAc or Gal residues among others has been proved.[103, 104, 192, 204, 220, 229] Indeed, in the Chapter III, it has been described that DC-SIGN binds *E. coli* R1-type core oligosaccharide, by recognizing its outer core pentasaccharide, acting as a cross-linker between two tetrameric units of DC-SIGN, in a pro-inflammatory way.

In this chapter, an interdisciplinary approach, combining NMR spectroscopy, computational and biophysical studies, has been employed to propose a 3D complex of the interaction between DC-SIGN and BVMPK LPS, confirmed also by predictions of CORCEMA-ST program. Understanding the recognition of LPS by DC-SIGN opens new avenues for therapeutic interventions against GM-related inflammatory diseases. The study reveals that the interaction between DC-SIGN and BVMPK LPS O-antigen occurs through its terminal portion in a sugar-dependent manner and in an independent length-manner. Specifically, the presence of a tMan residue resulted to be crucial for recognition and binding. The research evidenced the complexity of the binding process, highlighting two distinct binding modes (**BmA** and **BmB**) and underscoring the dynamic nature of the interaction. Notably, the tMan emerges as a key player in the interaction, as confirmed by experimental NMR results and computational analyses. Furthermore, the study delves into the bioactive conformation of the LPS core region, uncovering how BVMPK LPS interacts with DC-SIGN. The investigation identifies Fuc, Gal, Gal*f*, and Kdo as crucial residues in the LPS core recognition by DC-SIGN. The study describes a novel, complex binding epitope involving a terminal β-Man linked to an internal L-Rha and a terminal α-Fuc located in the inner core region.

This newfound understanding of the interaction between *B. vulgatus* LPS and DC-SIGN at a molecular level contributes to a broader comprehension of host-microbe interactions and offers potential applications for gut health. Particularly, the discovery of this beneficial interaction could lead to the development of therapeutic interventions for conditions related to gut dysbiosis, such as allergy, autoimmunity, and gastrointestinal inflammatory disorders.[231] In literature, a study has been reported in which EPS produced by strains of *Lactobacillus* and *Bifidobacterium* can exert various beneficial functions, such as changing microbiota compositions, interacting with the intestinal epithelial barrier, and modulating the intestinal immune system.[232, 233] In this context, the obtained findings about the anti-inflammatory interaction between *B. vulgatus* LPS and DC-SIGN could be exploited for the formulation of tailored dietary LPS mixtures, acting as a prebiotic carbohydrates. Furthermore, this interaction might be leveraged to design diagnostic tools for monitoring the composition of the gut microbiota. Detecting changes in DC-SIGN's interaction with *B. vulgatus* LPS could be indicative of shifts in the microbial community, allowing for the early diagnosis of gut-related diseases. However, further exploration of DC-SIGN interactions with other members of the gut microbiota could contribute to a deeper understanding of the complex relationships between the host immune system and commensal bacteria.

# Chapter 5 – Molecular recognition of eukaryotic glycans by *Mycoplasma pneumoniae* and *Mycoplasma genitalium* cytoadhesins

## 5.1 Introduction

Bacterial adhesion to host target cells constitutes a pivotal initial stage of infection, playing a crucial role in pathogenesis. Dedicated bacterial surface structures facilitate this process, enabling microorganisms to target specific biomolecular signatures on host cells.[234] Mycoplasmas are Gram-positive derived facultative anaerobic bacteria, cell-wall less, occasionally adopting a parasitic lifestyle.[235]
Among *mycoplasma* species, *Mycoplasma pneumoniae* (*Mpn*) and *Mycoplasma genitalium (Mge)* promote *mycoplasma*-associated infectious diseases,[116, 117] through the adhesion to target host cells, mediated by the interaction between peculiar adhesins and specific sialylated glycans, exposed on host cell surfaces. *Mpn* is an exclusively human parasite,[118] primarily affecting the conducting airways and being the leading cause of upper and lower respiratory tract infections.[116, 119] It is responsible for up to 40% of community-acquired pneumonias (CAPs) in children and adults worldwide. Atypical CAPs, caused by *Mpn*, may also involve extrapulmonary organs, including, for example, head, eyes, ears, skin or gastrointestinal tract.[236] *Mge* is a sexually transmitted bacterium associated with various reproductive tract syndromes, including cervicitis, pelvic inflammatory disease (PID), infertility[121] in women, and urethritis in men.[122] Additionally, *Mge* has been associated with preterm birth, spontaneous abortion, and HIV acquisition.[237] Both *Mpn* and *Mge* employ cytoadherence and gliding motility, mediated by a complex attachment organelle, with a transmembrane and tetrameric adhesion complex, called Nap, playing a fundamental role in infectivity (cf. §1.5.2.1).[127]

The current chapter offers a distinct perspective on host-guest interactions with respects to those previously described in this thesis, focusing on the analysis of eukaryotic glycans and bacterial proteins. Specifically, the recognition profiles and binding modes of different eukaryotic sialoglycans (Figure 5.1), commonly found on the host cell surface, in interaction with bacterial cytoadhesins, namely, P110 from *Mge* and P40/P90 from *Mpn,* located in the Nap complex, are dissected at atomical level.

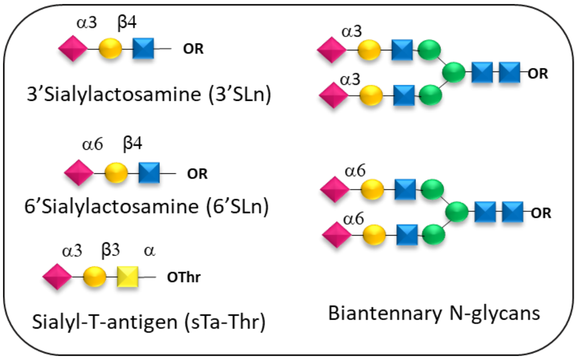


**Figure 5.1:** Overview of sialoglycans studied in interaction with P110 and P40/P90 represented using SNFG nomenclature.

The 3D complexes of different cytoadhesins interacting with common N- and O-glycans, found on host cells, have been delineated, employing techniques such as NMR spectroscopy, fluorescence analysis and computational studies, as depicted in the workflow represented in Figure 5.2.

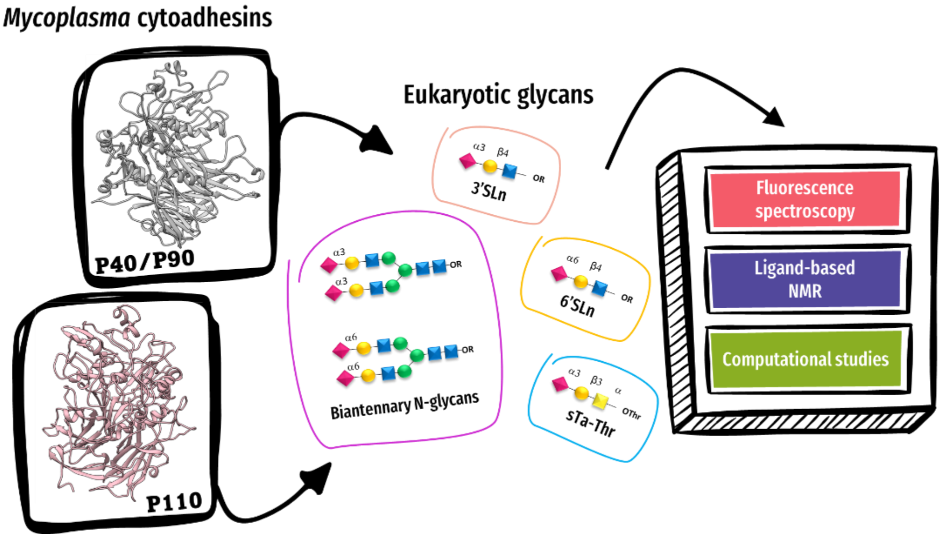


**Figure 5.2:** Workflow of the experimental activities.

The obtained results provide, on one side, precise ligands epitope maps together with the bioactive conformations and, on the other side, protein structural features governing sialoglycans recognition and the corresponding binding affinities. These findings serve as a foundational basis for developing alternative therapeutic strategies aimed at inhibiting bacterial adhesion to counteract *mycoplasma*-associated infections.

## 5.2 Fluorescence spectroscopy

The interaction of *mycoplasma* cytoadhesins, P110 (from *Mge)* and P40/P90 (from *Mpn)*, with the ligands 3'SLn, 6'SLn and sTa-Thr, was first examined using fluorescence spectroscopy (Figure 5.3).

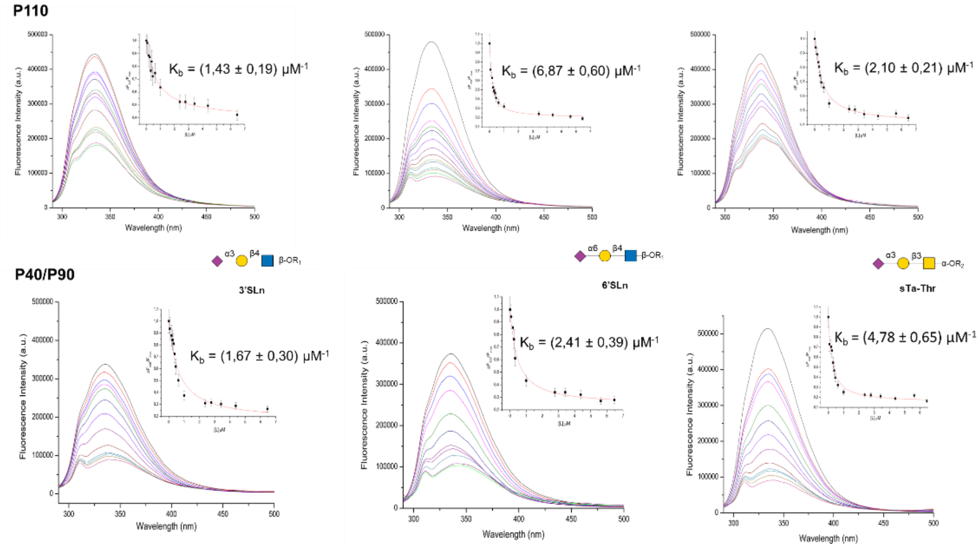


**Figure 5.3:** Fluorescence titrations. Fluorescence spectra of *Mge* P110 (upper panel) and *Mpn* P40/P90 (lower panel) in the presence of increasing amounts of the ligands 1-3, respectively. The insets report the binding isotherm and the values of the binding constants ($K_b$) obtained as described in the text.

A constant concentration of both proteins was titrated with increasing quantities of each sialoglycan to generate binding isothermal curves for all combinations of protein and ligand and to obtain information regarding the binding affinity. As results, both P110 and P40/P90 exhibited similar recognition capabilities for various Neu5Ac-containing ligands and all the obtained binding constants ($K_b$) were in the micromolar range. (Figure 5.3). Interestingly, these results were in accordance with the $K_D$ previously obtained by SPR experiments.[125, 126]

The molecular recognition of each sialylated ligand by *Mge* and *Mpn* cytoadhesins have been subsequently examined using complementary techniques to elucidate the mechanisms underlying bacterial adhesion and gliding. In particular, the ligand binding epitopes were obtained using STD NMR;[140, 141] tr-NOESY experiments[238, 239] were employed to explore the conformational behavior of the ligands. Additionally, MD simulations were performed in order to generate 3D models of the complexes, to support the NMR results, as described as follows.

## 5.3 Interaction of 3'SLn with P110 and P40/P90

The analysis of the STD NMR spectra of the mixture P110/3'SLn indicated the selective affinity of the protein binding pocket toward the Neu5Ac moiety (Figure 5.4A). As suggested by the STD enhancements, the entire Neu5Ac was recognized by P110, receiving a good magnetization transfer. The acetyl group of the Neu5Ac unit exhibited the highest relative STD effect, H6 and H7 protons registered an STD percentage exceeding 40%; while H4, H5, H8 and H9 Neu5Ac protons showed lower STD signals. The other sugar residues (Gal and GlcNAc) did not contribute to the recognition process. These results confirmed the interactions observed in the crystal structure of P110 in complex with 3'SLn.[125]

It is noteworthy that, despite the hypothesis of a sialidase catalytic site in proximity to the protein binding pocket,[125, 127] no hydrolysis of the Neu5Ac from the ligand was observed in the presence of the protein. Therefore, the NMR data strongly indicate the absence of a catalytic site in the protein crown region.

Tr-NOESY experiments and MD simulations were employed to elucidate the conformational behavior of the ligand and describe how it is accommodated in the protein binding pocket. The presence of key tr-NOEs between the diastereotopic protons at position 3 of Neu5Ac and the proton H3 of the Gal residue (Figure 5.4B) suggested a conformer selection upon binding, with 3'SLn adopting a *t* conformation when accommodated in the protein binding pocket. The NMR results were confirmed by MD simulations. Indeed, the comparison of glycosidic torsion trajectories between the free (data not shown) and the bound state revealed that, in the absence of the protein, the ligand could adopt three main conformations: *t, g*, and -*g*, based on the different values of the Φ(C1-C2-O-C3') dihedral angle around the Neu5Ac-α-(2,3)-Gal glycosidic linkage, at 180°, 60° and -60°, respectively.[138, 240] On the other hand, upon binding, the $\phi$ torsion angle mainly adopted the value of 180°, indicating that the t conformer was preferentially accommodated into the protein binding pocket (Figure 5.5).

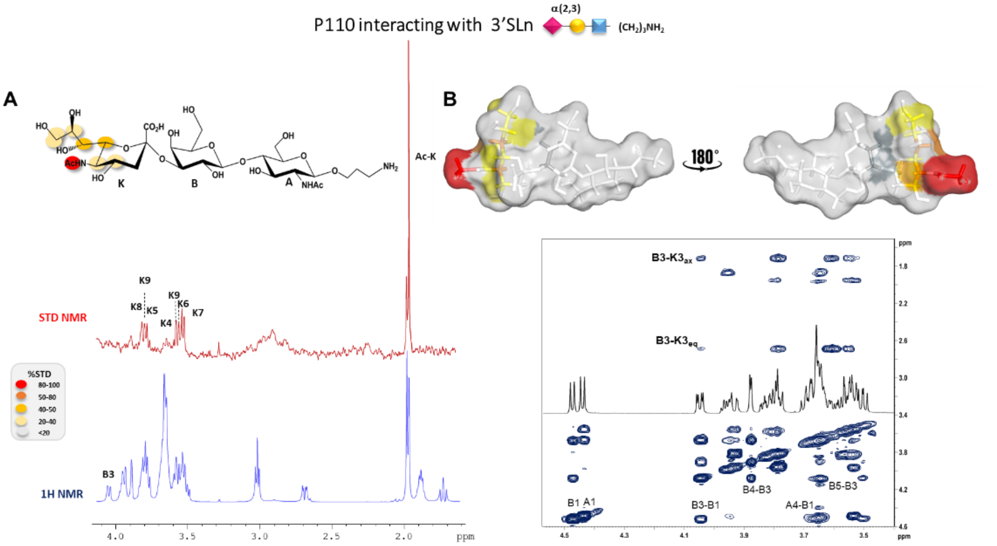


**Figure 5.4:** NMR analysis of 3'SLn bound to P110. A) STD-NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand interacting with P110. B) Tr-NOESY spectra of P110-3'SLn and STD-derived epitope mapping of 3'SLn in its bioactive conformation with the surface coloured according to the STD effects.

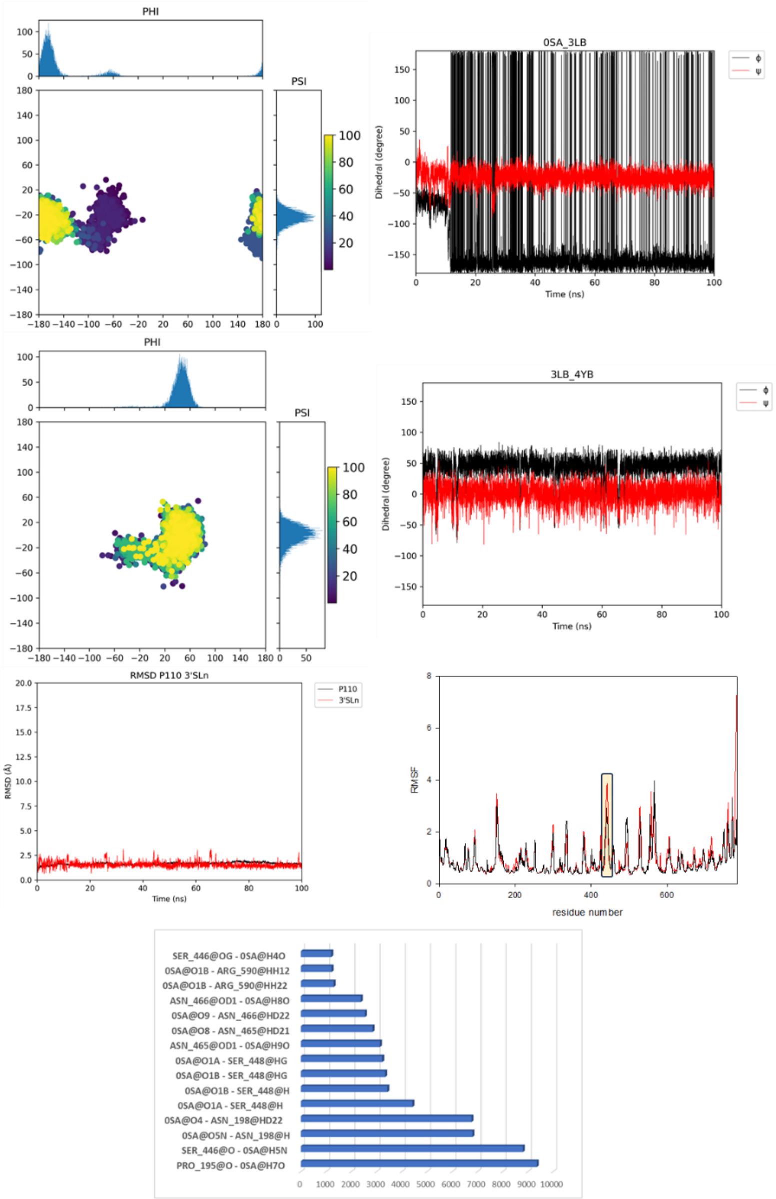


**Figure 5.5:** MD analysis of P110 in interaction with 3'SLn. 3'SLn dihedral angles analysis in the bound state with P110 are represented as scatter plots of the ϕ torsion against ψ, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) are reported. The ligand RMSD was calculated in reference to the protein. The overall flexibility of the protein in the free (red line) and bound (black line) state was measured via the average root mean square fluctuation (RMSF) values during the MD trajectory of P110 bound to 3'SLn. RMSF measures the average amplitude of each atom's motions relative to a mean reference position. The higher value of RMSF represents higher flexibility. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

The cluster analysis of MD simulations allowed an accurate description of the binding mode and the main molecular interactions between P110 and 3'SLn (Figure 5.6). Numerous hydrogen bonds were observed at the protein-ligand interface: specifically, H-bonds were established between Neu5Ac carboxyl group and Arg600 and Ser458. Asn200 exhibited two interactions with Neu5Ac: one with the OH at position 4 and the second with its acetamide moiety, while Pro197 formed a H-bond with the hydroxyl proton 7 of Neu5Ac. In addition, Asn475 established crucial H-bonds with hydroxyl protons 8 and 9 of the glycerol chain of Neu5Ac, which strongly contributed to stabilize the complex and resulted in the flexible loop region (471-482 aa) moving closer to the Neu5Ac binding site. As shown in Figure 5.6, the ligand was positioned with the acetamide group of the Neu5Ac in a hydrophobic pocket close to Phe457, contributing to stabilize the binding. Another significant contact was observed between the amide nitrogen of Neu5Ac and Ser456, forming a hydrogen bond, stable during the MD simulation, in accordance with the high STD contribution.

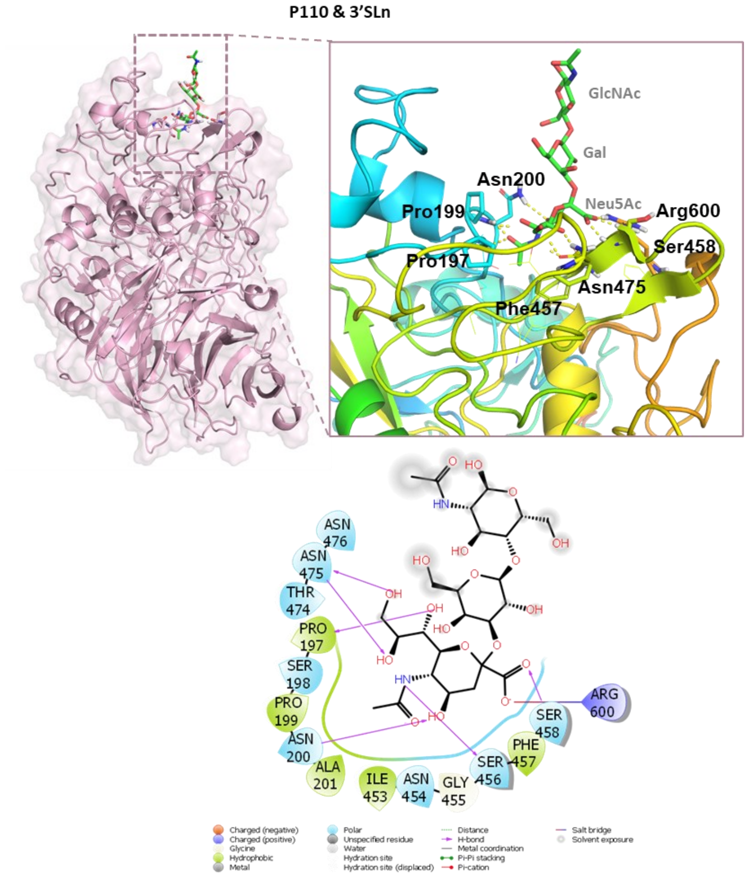


**Figure 5.6:** 3D view of the P110-3'SLn complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. The flexible loop moving closer to the binding site upon binding is represented in yellow. Two-dimensional plot highlighting the main protein-ligand interactions is also reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

Similarly, the interaction between the *Mpn* cytoadhesin P40/P90 and 3'SLn was investigated. Also in this case, STD NMR analysis underlined a selective binding for the Neu5Ac moiety (Figure 5.7). The acetyl group of Neu5Ac exhibited the highest magnetization transfer and the H7 of Neu5Ac showed a substantial STD effect, exceeding 50%. Protons H5, H6, and H9 also contributed to the binding, even though lower STD percentages ranging from 50% to 20%. Slight STD enhancements were observed for the diastereotopic protons of Neu5Ac and the H4 of the Gal unit.

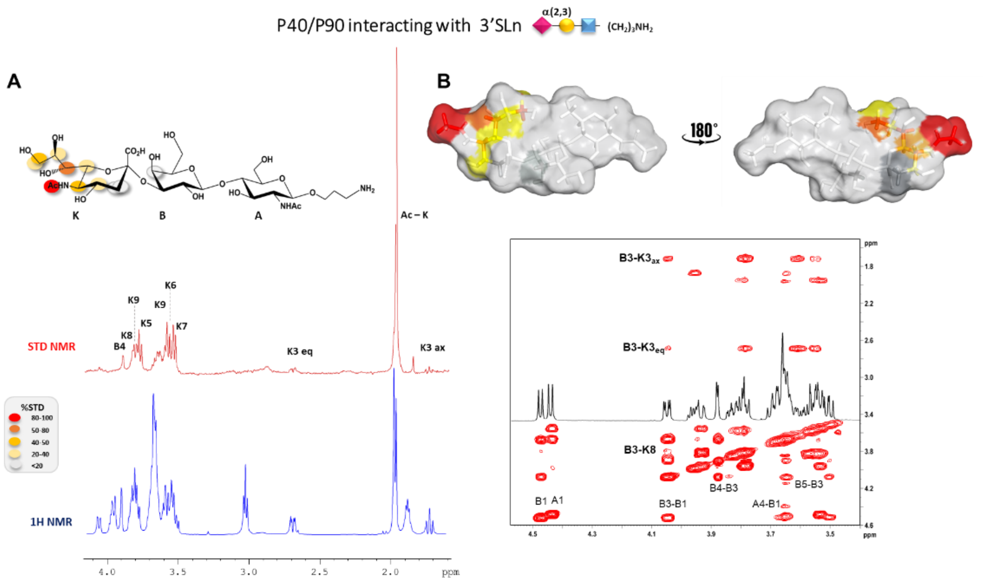


**Figure 5.7:** NMR analysis of 3'SLn bound to P40/P90. A) STD-NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand interacting with P40/P90. B) Tr-NOESY spectra of P40/P90-3'sLn and STD-derived epitope mapping of 3'SLn in its bioactive conformation with the surface coloured according to the STD effects.

Computational studies were also carried out to provide insights into the conformational behavior of 3'SLn when bound to P40/P90. Consistent with tr-NOESY data (Figure 5.7B), MD results indicated that 3'SLn could adopt two distinct conformations in the binding pocket: *t* and *-g*. These conformations corresponded to Φ dihedral angles around the glycosidic linkage between Neu5Ac and Gal at 180° and -60°, respectively (Figure 5.8). This variability is likely attributed to an increased flexibility and a distinct orientation of a loop region (642-653 aa), which remained distant from the binding site, resulting in a wider binding pocket compared to that of P110, allowing a higher ligand flexibility. However, accordingly to the STD NMR data, the main interactions were exclusively established with the Neu5Ac moiety also in the complex with P40/P90.

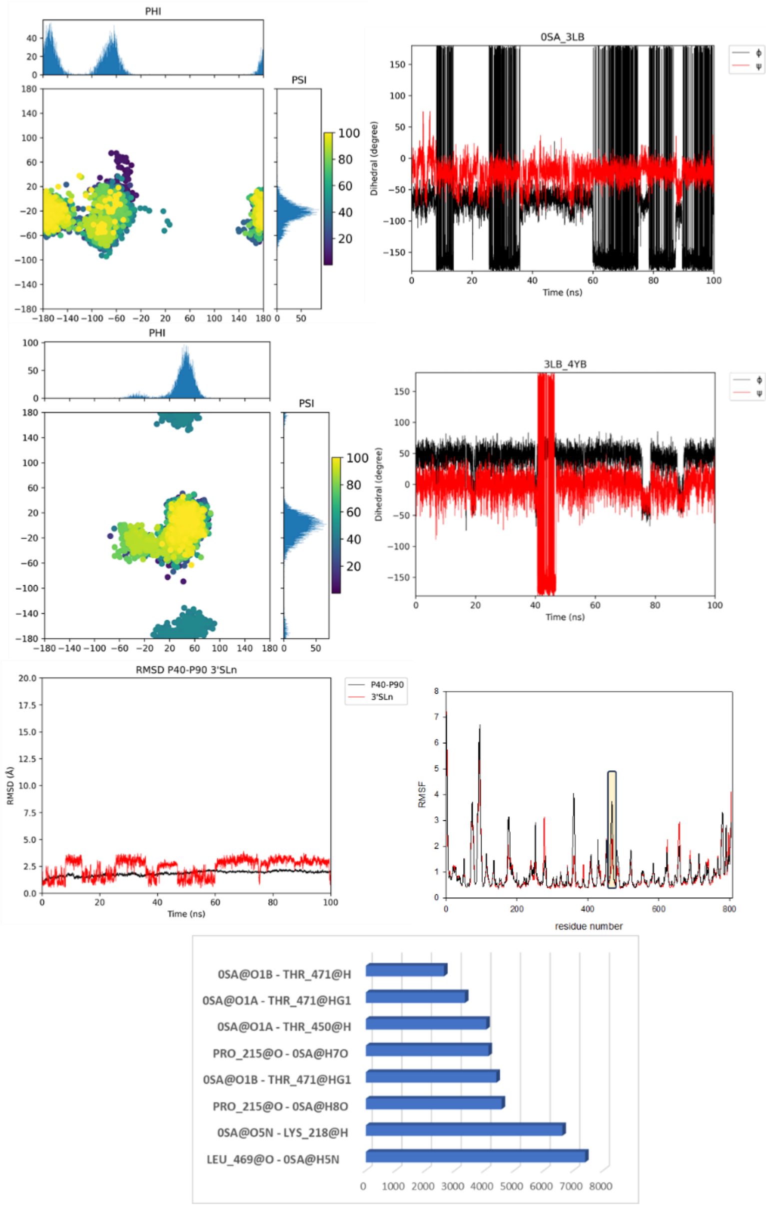


**Figure 5.8:** MD analysis of P40/P90 in interaction with 3’SLn. 3’SLn dihedral angles analysis in the bound state with P40/P90 is represented as scatter plots of the ϕ torsion against ψ, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) are reported. The ligand RMSD was calculated in reference to the protein. The overall flexibility of the protein in the free (red line) and bound (black line) state was measured via the average root mean square fluctuation (RMSF) values during the MD trajectory of P40/P90 bound to 3’SLn. RMSF measures the average amplitude of each atom's motions relative to a mean reference position. The higher value of RMSF represents higher flexibility. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

As shown in the main representative pose of the complex (Figure 5.9), Neu5Ac was the only sugar residue completely accommodated in the protein binding pocket, with Gal and GlcNAc residues positioned further away from the protein surface. Specifically, a stable contact was observed between the amide group at position 5 of Neu5Ac and Leu630, persisting for approximately 70% of the MD simulation time. The Neu5Ac carboxyl group formed H-bonds with the side chain of Thr632. In accordance with STD NMR findings, the hydroxyl group at position 7 interacted with Pro235, while the oxygen of the acetyl moiety was in contact with Lys238. The existence of hydrophobic residues near the binding site, like Phe631, played a role in stabilizing the complex, allowing the interaction of Neu5Ac with the protein, through hydrophobic interactions.

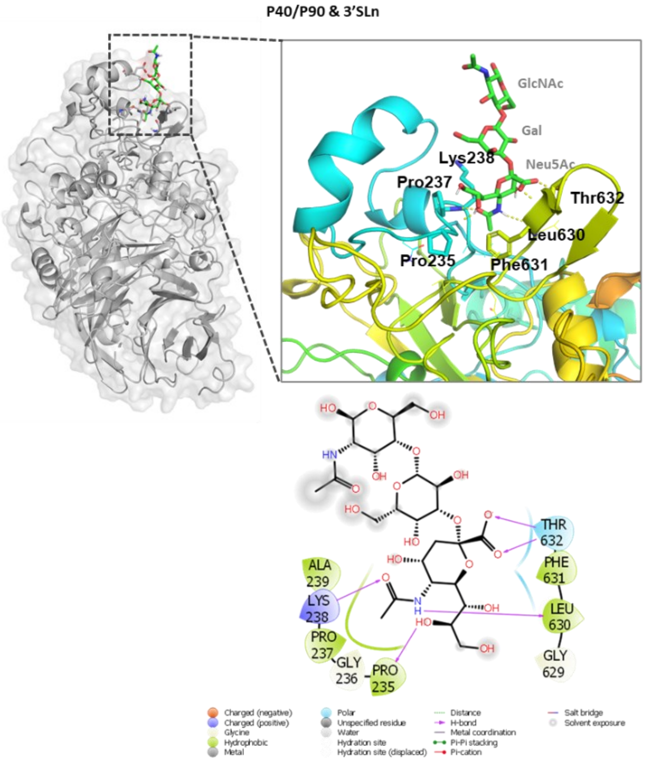


**Figure 5.9:** 3D view of the P40/P90-3'SLn complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. The flexible loop moving closer to the binding site upon binding is represented in yellow. Two-dimensional plot highlighting the main protein-ligand interactions is also

reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

## 5.4 Interaction of 6’SLn with P110 and P40/P90

The binding mode and conformational behavior of 6’SLn interacting with P110 and P40/P90 were investigated using a combination of NMR experiments and computational approaches. Once again, the portion of the ligand most involved in binding with each cytoadhesin was the Neu5Ac. The analysis of STD NMR and the resulting epitope map of 6’SLn bound to P110 indicated that the acetyl group of Neu5Ac had the highest STD NMR effect (Figure 5.10). Additionally, proton H7 significantly contributed to the interaction, with an STD NMR relative percentage close to 50%, while H6, H8, and H9 protons of Neu5Ac made a minor contribution.

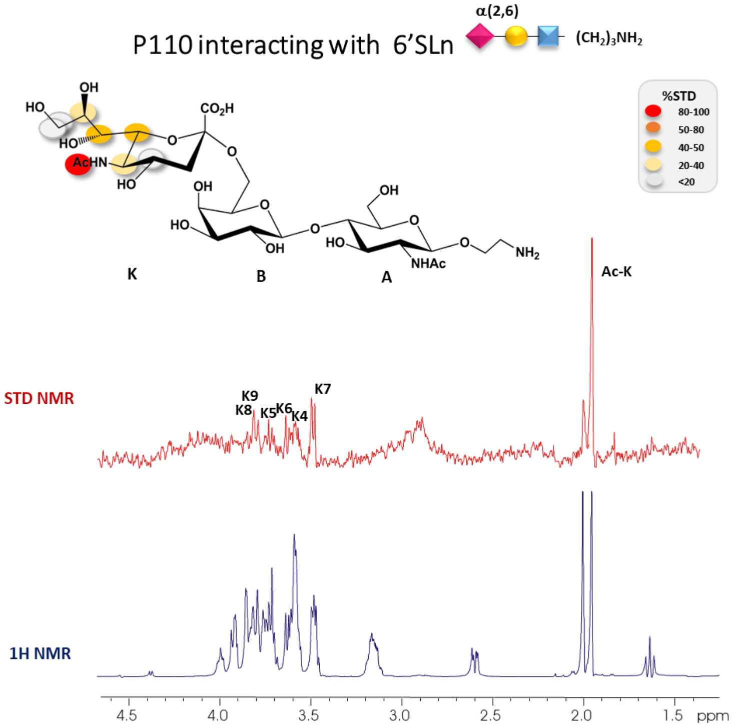


**Figure 5.10:** STD-NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand 6’SLn interacting with P110.

MD data revealed the formation of numerous H-bonds between P110 amino acids and the terminal sugar of 6’SLn, further stabilized by hydrophobic contacts with Phe457 (Figure 5.11). Pro197 and the hydroxyl group at position 7 of Neu5Ac established interaction for more than 90% of the simulation time, as well as the amide group forming an interaction with the carbonyl oxygen of Ser456. Other hydrogen bonds were established, including those between Asn200 and the carbonyl oxygen of the acetyl moiety, as well as between Asn200 and the hydroxyl group at position 4. The hydroxyl protons at positions 8 and 9 of the glycerol chain of Neu5Ac interacted with Asn475. Moreover, the hydroxyl group of Ser458 played a crucial role in interacting with the carboxyl group of Neu5Ac. Despite Neu5Ac was the residue mainly involved in the interaction, there were also different transient interactions between protein amino acids and some protons of the GlcNAc unit throughout the simulation.

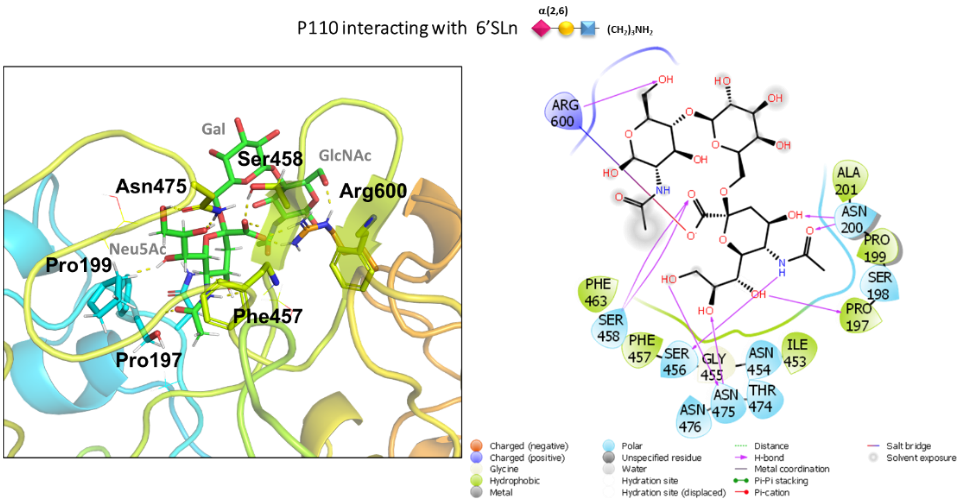


**Figure 5.11:** 3D view of the P110-6’SLn complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. Two-dimensional plot highlighting the main protein-ligand interactions is also reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

the results observed for the cytoadhesin P110, the STD-NMR experiments conducted on the P40/P90-6’SLn mixture indicated that the protein selectively recognized the Neu5Ac residue (Figure 5.12). The epitope map revealed that the most involved group was the acetyl moiety of Neu5Ac, with protons H6 and H7 exhibiting high STD enhancements of approximately 50%, while the glycerol chain and other protons of Neu5Ac contributed to a lesser extent.

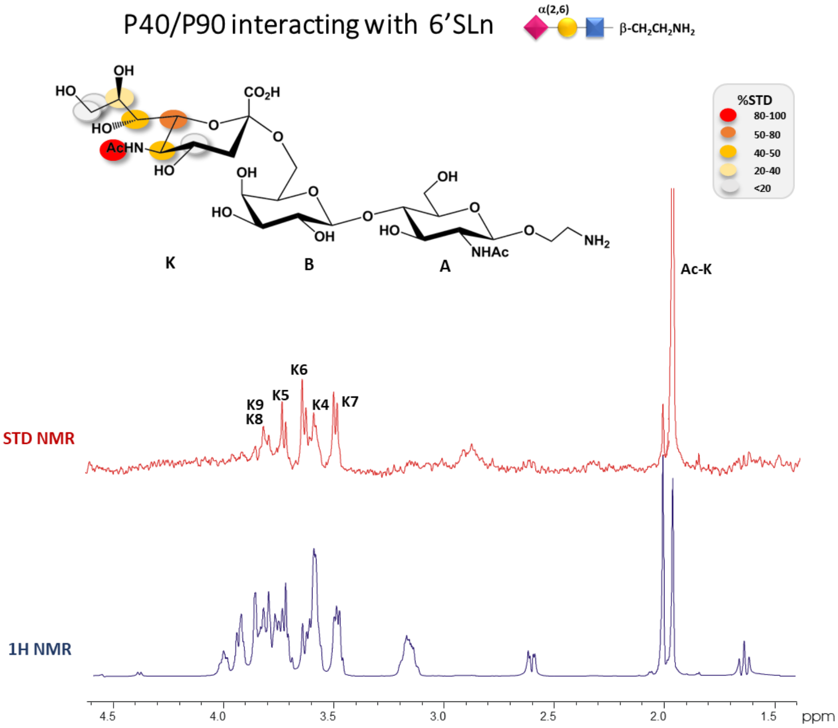


**Figure 5.12:** STD-NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand 6'SLn interacting with P40/P90.

These findings aligned with MD results, highlighting that 6'SLn established a crucial interaction with amino acids in the binding pocket primarily through the Neu5Ac residue. Specifically, the carboxyl group of Neu5Ac formed a H-bond with Thr632, consistent with the crystal structure of P40/P90 in complex with 3'SLn.[126] The acetamide group of Neu5Ac interacted with Leu630 and Lys238, while Pro235 displayed an interaction with the hydroxyl group at position 8 of the glycerol side chain (Figure 5.13).

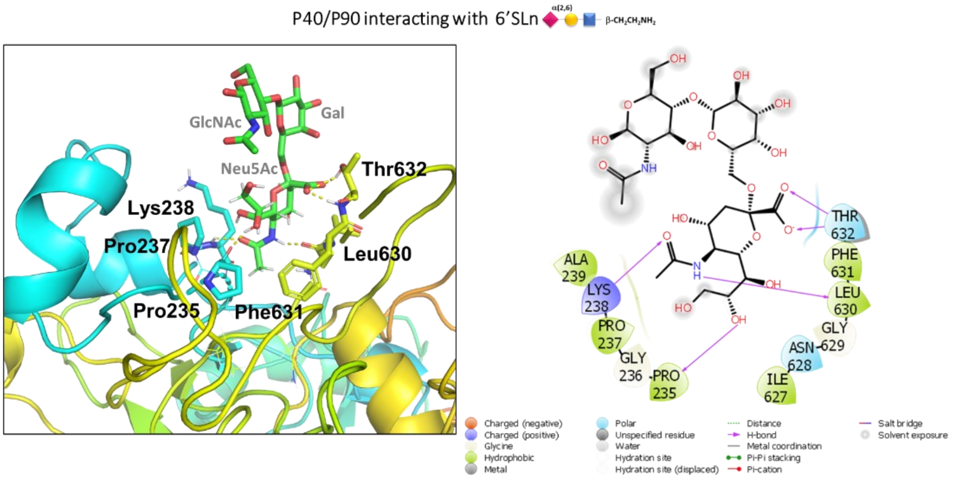


**Figure 5.13:** 3D view of the P40/P90-6'SLn complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. Two-dimensional plot highlighting the main protein-ligand interactions is also reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

The analysis of MD simulations also indicated a higher flexibility of the ligand when bound to P40/P90 compared to the complex with P110, as resulted previously for the analysis of the complexes with 3'SLn. Due to the additional ω torsion angle in 6'SLn compared to 3'SLn, the ligand could adopt different conformations according to different values of the ω angle (60°, 180°, and -60°), resulting in an equilibrium between different rotamers, namely *gt*, *tg*, and *gg*, respectively. Specifically, in the free state (data not shown), the *gt* and *tg* rotamers were the most populated. Interestingly, MD results revealed that upon binding to P110, the ligand was predominantly accommodated in the *gt* conformation (Figure 5.14). Conversely, when 6'SLn bound to P40/P90, an equilibrium between different rotamers, similarly accommodated in the protein binding pocket, was observed, in which the *tg* conformation was the most populated (Figure 5.15).

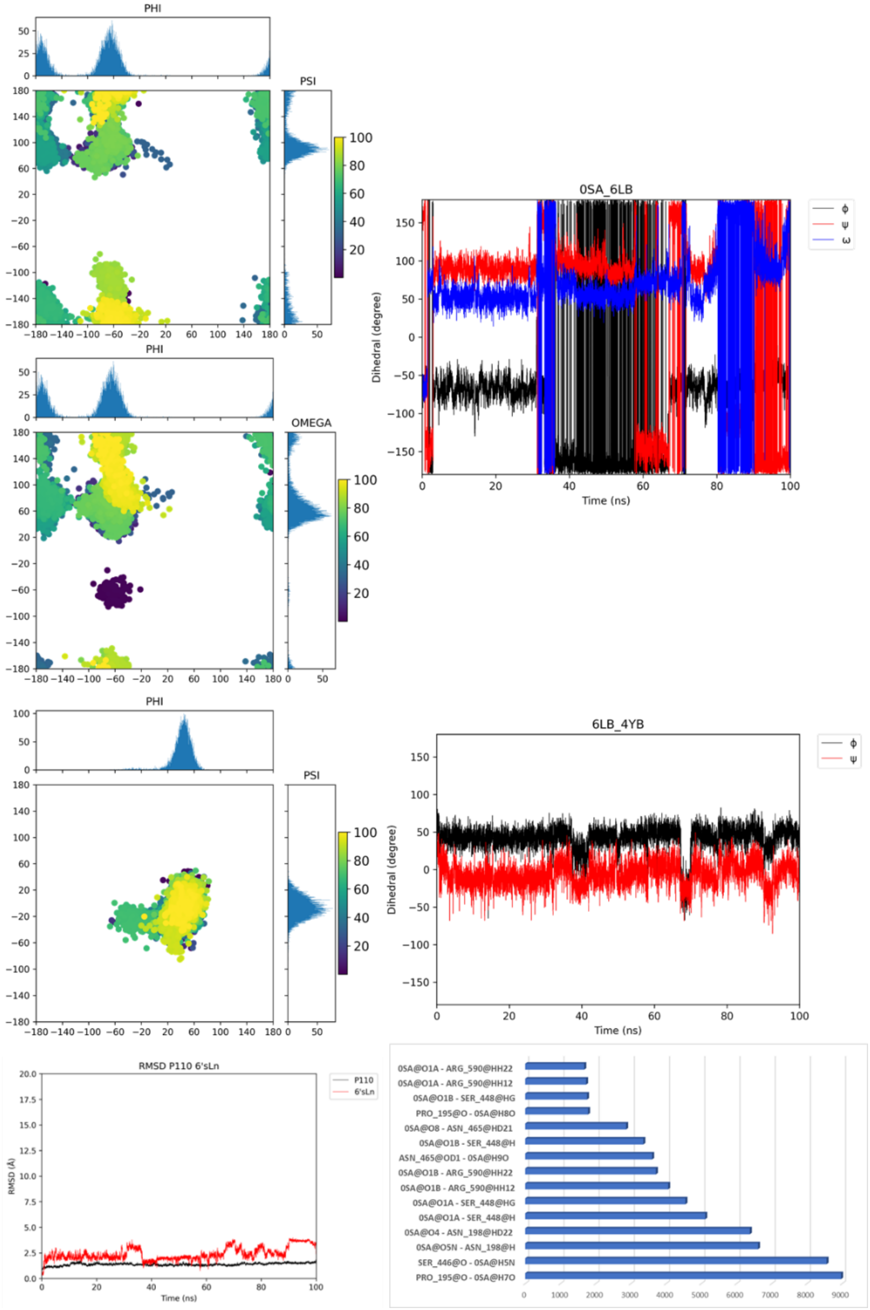


**Figure 5.14:** MD analysis of P110 in interaction with 6’SLn. 6’SLn dihedral angles analysis in the bound state with P110 is represented as scatter plots of the ϕ torsion against ψ or ω, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) is reported. The ligand RMSD was calculated in reference to the protein. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

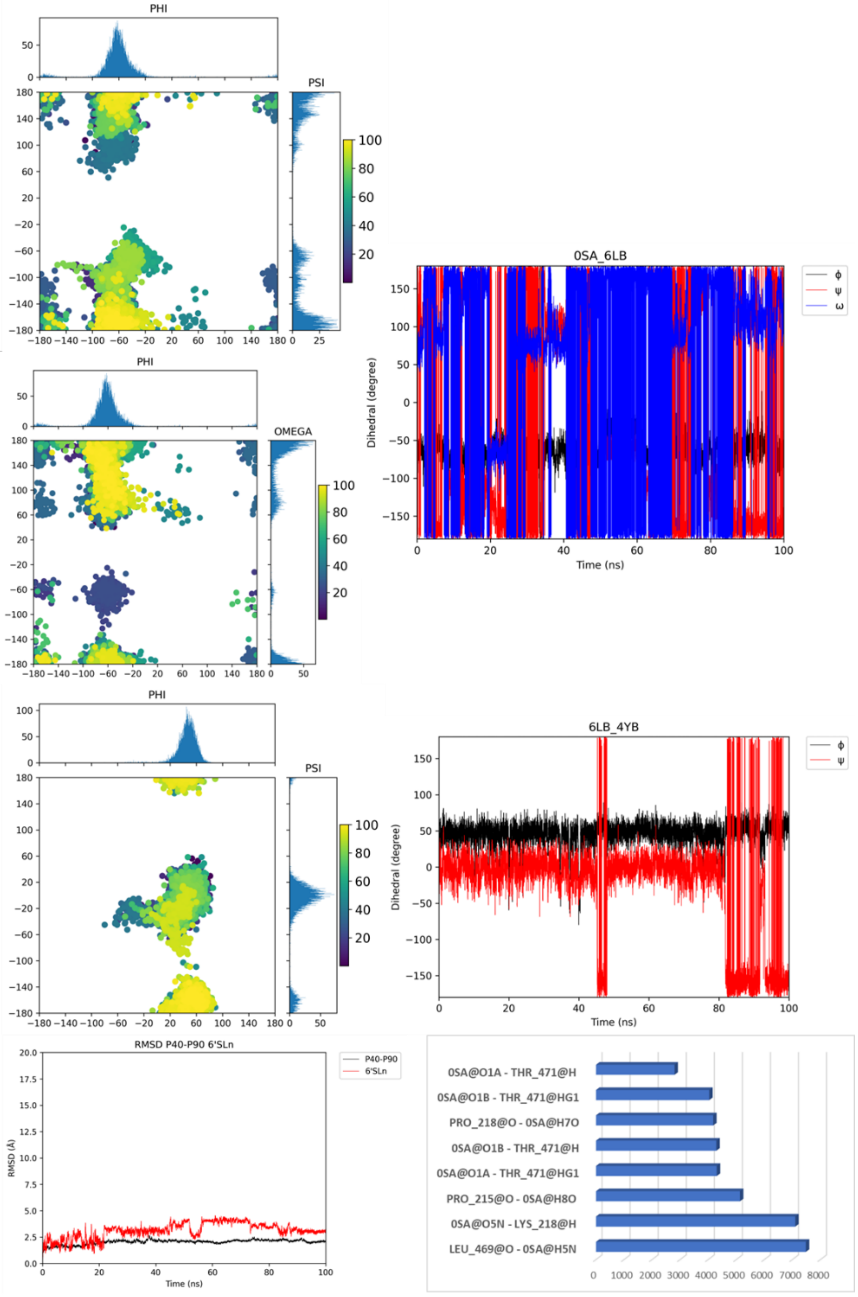


**Figure 5.15:** MD analysis of P40/P90 in interaction with 6’SLn. 6’SLn dihedral angles analysis in the bound state with P40/P90 is represented as scatter plots of the ϕ torsion against ψ or ω, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) is reported. The ligand RMSD was calculated in reference to the protein. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

## 5.5 Interaction of complex type N-glycans with P110 and P40/P90

P110 and P40/P90 were also tested in interaction with longer N-glycans, particularly two undecasaccharides, containing the Neu5Ac-α-(2,3)-Gal or Neu5Ac-α-(2,6)-Gal epitopes at their terminal end, mimicking natural complex-type biantennary N-glycans. Experiments performed on P40/P90 interacting with branched α-(2,3)-undecasaccharide have been reported as example in this thesis. STD-NMR experiments (Figure 5.16) revealed a selective protein recognition for the Neu5Ac residues, while the other sugar units did not contribute to the binding process, as happened with smaller trisaccharides. As already observed in the case of trisaccharide ligands, the acetyl group of the Neu5Ac displayed the highest STD enhancement, although STD signals of strong intensity also came from H6 and H7 of the Neu5Ac. In contrast, the resonances of the protons belonging to the other sugar units were not visible in the spectra, indicating that they were solvent exposed.

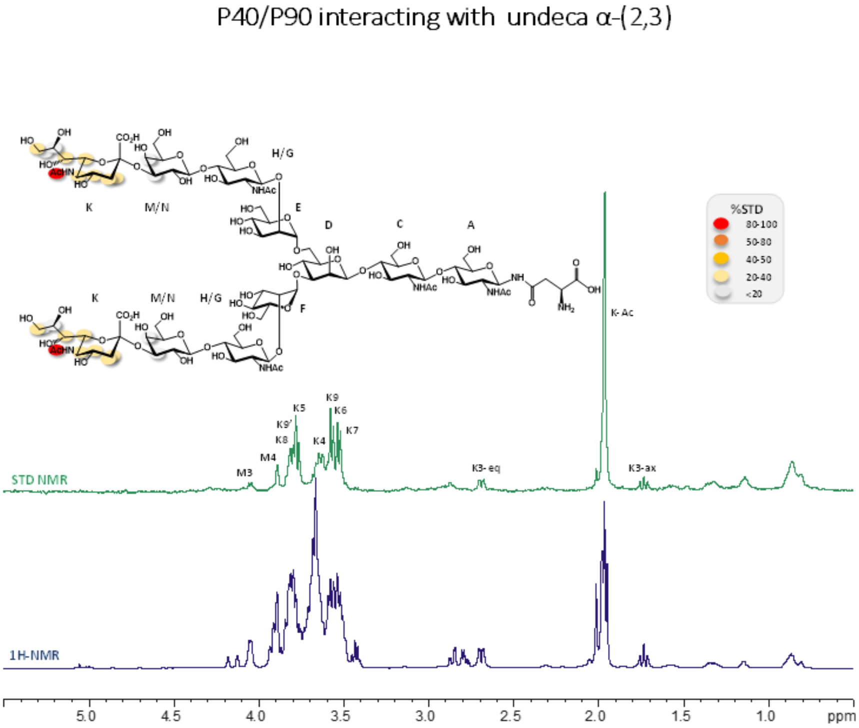


**Figure 5.16:** NMR analysis of complex glycan bound to P40/P90. STD-NMR spectrum (green) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand interacting with P40/P90.

These findings were also confirmed by MD simulations. As shown in Figure 5.17, the longer complex N-glycan was anchored to the protein surface, with the Neu5Ac on one antenna mainly establishing polar contacts with Leu634 and Thr632 and hydrophobic interactions with Phe631, as previously reported for the trisaccharide ligand. Additionally, more interactions were observed between the Neu5Ac unit positioned at the terminal end of the other glycan branch and Arg226 and Ser227 residues. Thus, for the first time, these results showed the interaction between *Mge* and *Mpn* cytoadhesins and complex biantennary N-glycans.

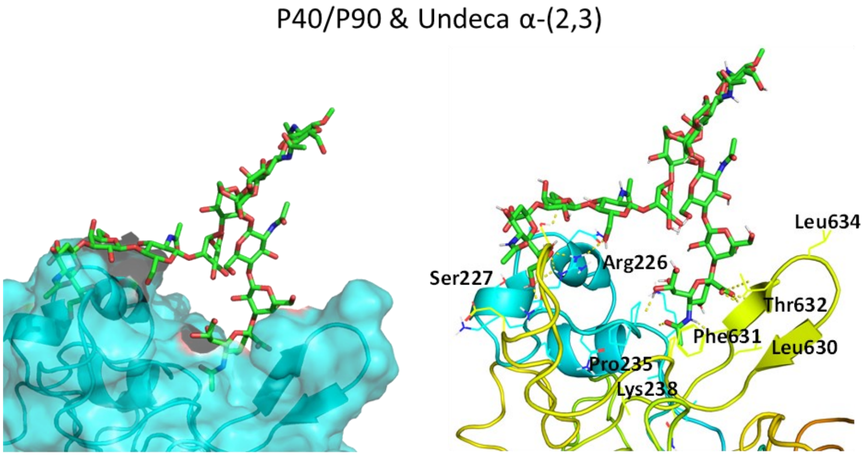


**Figure 5.17:** 3D view of the P40/P90-undecasaccharide α-(2,3) complex.

## 5.6 Interaction of sTa-Thr with P110 and P40/P90

The molecular interaction between *Mpn* and *Mge* cytoadhesins and the *ad hoc* synthetized sialyl-T-antigen linked to a Thr residue (sTa-Thr), was also investigated. As revealed by the STD NMR analysis, both P110 and P40/P90 proteins were able to recognize the O-glycan in a similar manner, where the Neu5Ac exhibited the highest STD contribution, the Gal unit slightly contributed to the interaction and the reducing GalNAc, along with the Thr residue, did not receive any magnetization transfer from the protein (Figure 5.18 and Figure 5.19).

In the case of P110, the acetyl group of Neu5Ac exhibited a significant contribution to the binding, followed by H7. H5, H6, H8, and H9 protons showed STD NMR relative percentages above 50%. In contrast, H4 of Neu5Ac and H4 and H6 of the Gal contributed to the binding event in a lesser extent, displaying %STD values below 40% (Figure 5.18).

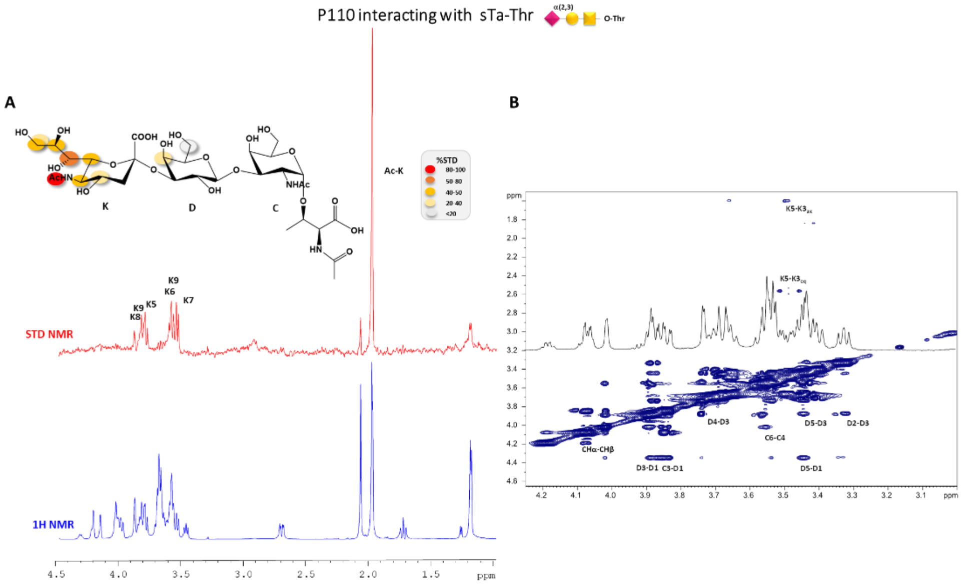


**Figure 5.18**: NMR analysis of sTa-Thr bound to P110. A) STD NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand interacting with P110. B) Tr-NOESY spectra of P110-sTa-Thr mixture.

Similarly, for P40/P90, a robust saturation transfer was observed for the acetyl group and H7 of Neu5Ac, showing the highest STD NMR effects. H4, H5, and H9 contributed to the binding with STD NMR enhancements greater than 50%, while H3, H6 and H8 exhibited %STD NMR lower than 50%. Moreover, slight STD effects were observed for the Gal moiety, particularly, H2, H3, H4 and H6 protons displayed STD NMR percentages lower than 20% (Figure 5.19).

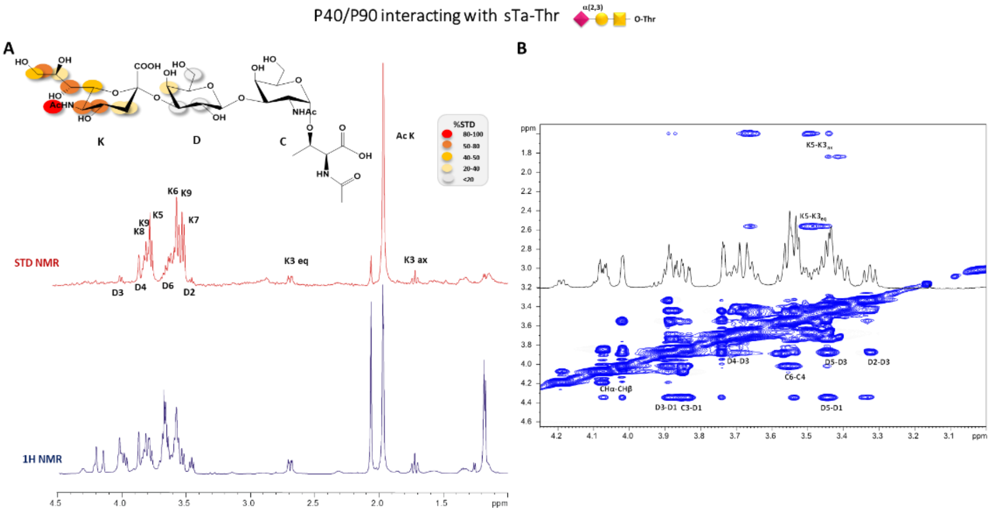


**Figure 5.19:** NMR analysis of sTa-Thr bound to P40/P90. A) STD NMR spectra (red) and the unsaturated reference spectrum (blue) together with the epitope map of the ligand interacting with P40/P90. B) Tr-NOESY spectra of P40/P90-sTa-Thr mixture.

Once the ligand-interacting epitopes were established, the binding profile of the P110 and P40/P90 complexes with sTa-Thr were also investigated using computational approaches and tr-NOESY analysis. In general, sTa-Thr was accommodated in the same protein binding pocket as the previously observed ligands. In the P110/sTa-Thr complex (Figure 5.20), interactions occurred between Pro197 and the hydroxyl group at position 7 of Neu5Ac, as well as between Ser456 and the amide nitrogen of the Neu5Ac. The Asn200 residue also interacted with the OH at position 4 of Neu5Ac, and Ser458 established two H-bonds with the carboxyl group of this residue, as seen in the crystal structures.[125, 127]

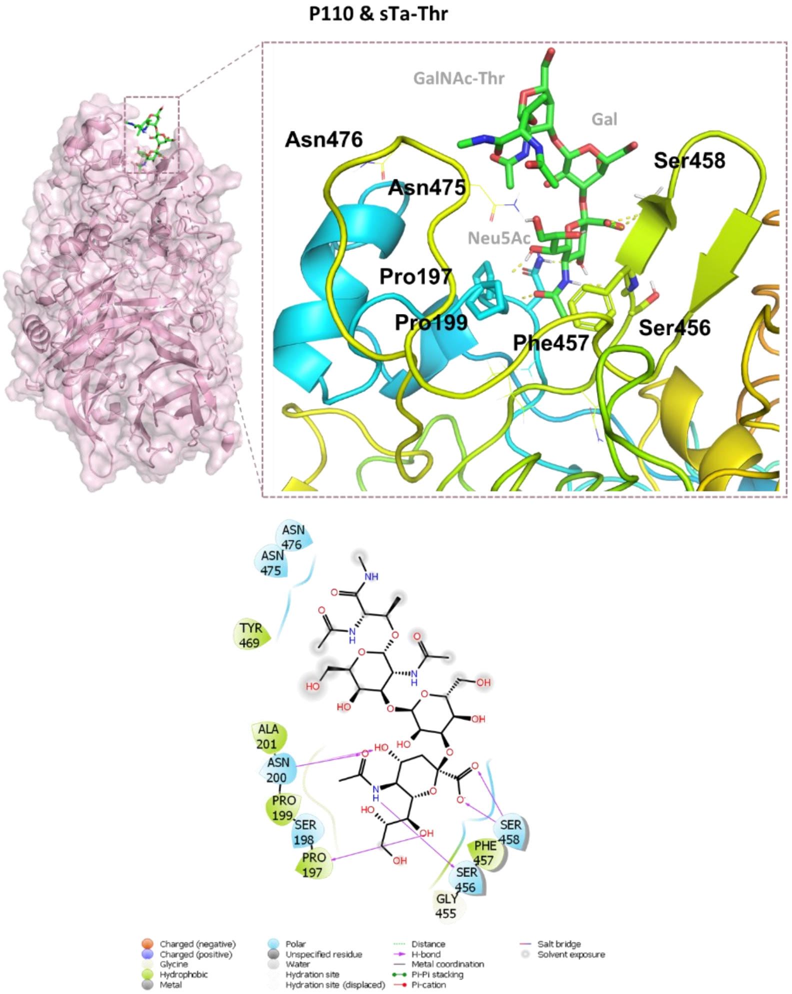


**Figure 5.20:** 3D view of the P110-sTa-Thr complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. Two-dimensional plots highlighting the main protein-ligand interactions is also reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

MD simulation of the P40/P90 complex with sTa-Thr (Figure 5.21) also revealed selective recognition toward Neu5Ac. The most stable contact occurred between Leu630 and the amide group of Neu5Ac, in accordance with the observed high STD NMR value. The Neu5Ac carboxyl moiety also established stable H-bonds with Thr632. Other significant contacts were observed for OH at C7 of Neu5Ac, forming a H-bond with Pro235, accordingly with the high STD contribution of H7 (around 80%) and, also, Lys218 with the acetyl group.

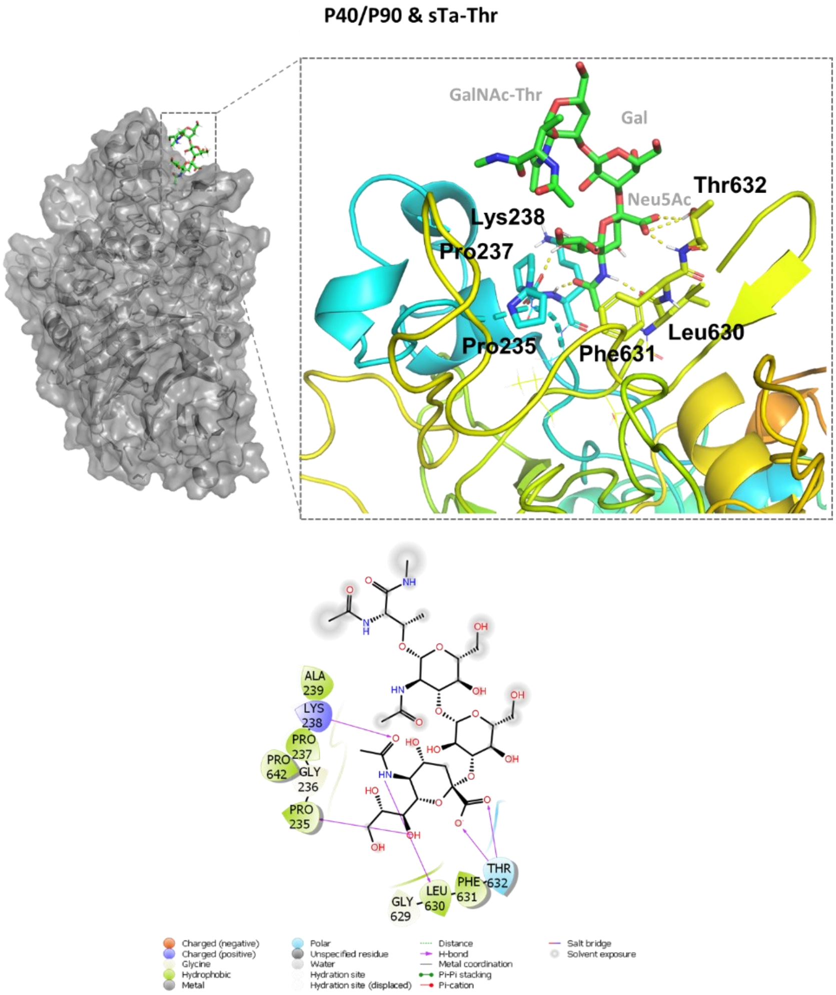


**Figure 5.21:** 3D view of the P40/P90-sTa-Thr complex. The aa of the binding pocket involved in the binding with Neu5Ac are represented as sticks. Two-dimensional plots highlighting the main protein-ligand interactions is also reported. Solid arrows represent hydrogen bonds with the functional groups of the backbone; the other residues in the binding pocket participate in polar and hydrophobic interactions.

MD simulations of sTa-Thr in free (data not shown) and bound states to both P110 and P40/P90 proteins were also run to compare the conformational behavior of the ligand. The shape and conformation of sTa-Thr were influenced by the glycosidic torsion angles, namely $\phi$ (C1-C2-O-C3')/$\psi$(C2-O-C3'-H3') around Neu5Ac-$\alpha$-(2,3)-

Gal and ϕ (H1–C1–O–C3')/ ψ (C1–O–C3'–H3') around the Gal-β-(1,3)–GalNAc bond. Interestingly, in the binding with P110, the *-g* conformer selection was observed, with the ligand populating only the minimum characterized by the value of ϕ torsional angle at -60° (Figure 5.22). Instead, regarding the interaction of sTa-Thr with P40/P90, no significant conformational differences were observed between the free and bound state, with the ligand mainly adopting the two *t* and *-g* conformations with the ϕ torsion angle along the Neu5Ac-α-(2,3)-Gal unit at 180° and -60°, respectively (Figure 5.23).

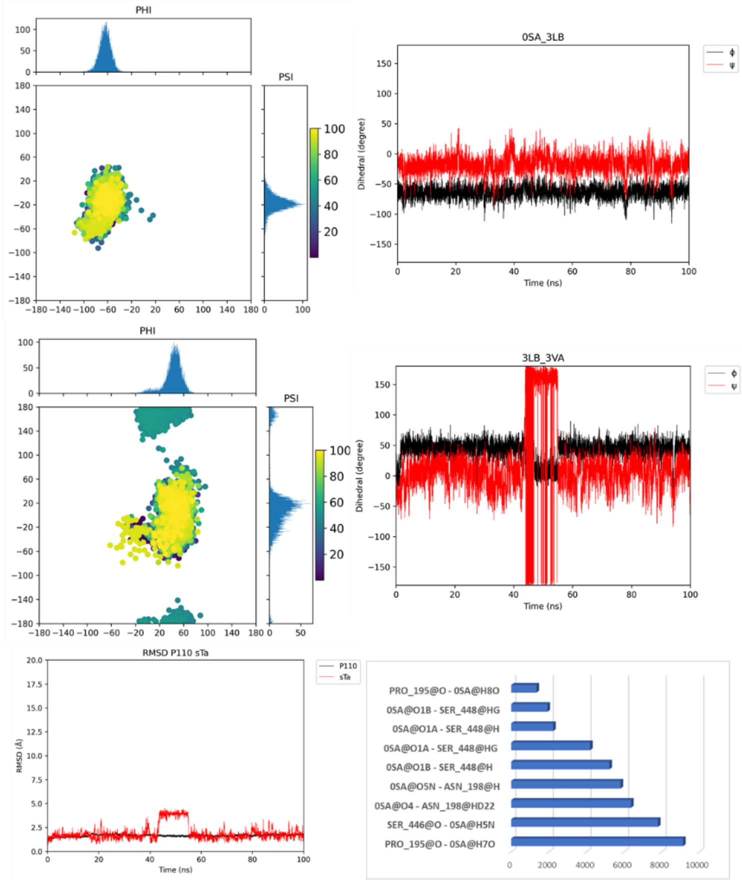


**Figure 5.22:** MD analysis of P110 in interaction with sTa-Thr. sTa-Thr dihedral angles analysis in the bound state with P110 is represented as scatter plots of the ϕ torsion against ψ, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) is reported. The ligand RMSD was calculated in reference to the protein. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

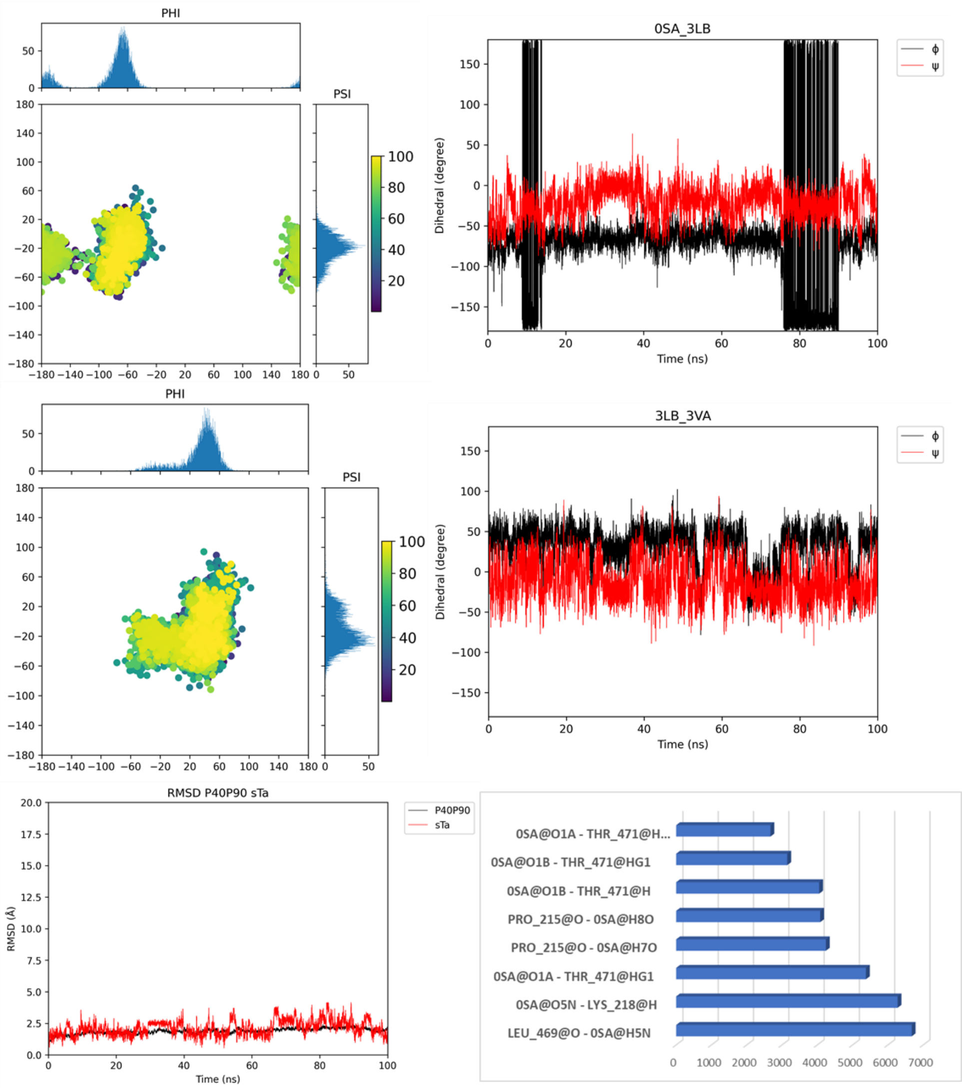


**Figure 5.23:** MD analysis of P40/P90 in interaction with sTa-Thr. sTa-Thr dihedral angles analysis in the bound state with P40/P90 is represented as scatter plots of the ϕ torsion against ψ, during the MD simulation with the relative histograms to represent the most populated energies. RMSD of the protein (black) and the ligand (red) is reported. The ligand RMSD was calculated in reference to the protein. The protein-ligand H-bonds (acceptor – donor H) established during the 100 ns MD are also shown.

## 5.7 Discussion

Sialic acid-containing glycans are abundantly present on eukaryotic host cell surfaces, playing crucial roles in diverse biological functions, relevant to both health and disease. They modulate cell division, signaling, and cell-cell recognition, influencing a wide spectrum of immune responses. Furthermore, several pathogenic organisms, including various *Mycoplasma* species, have developed mechanisms to utilize host sialoglycans for association and entry into target cells, thereby initiating bacterial infection.

Despite recent efforts to understand the mechanism of *mycoplasma* adhesion to host cells, the molecular basis of the protein-glycan interactions remains not fully understood. Considering the increasing antibiotic resistance observed in *mycoplasma* species, especially in the respiratory pathogen *Mpn* and urogenital pathogen *Mge*, there is an urgent need for alternative therapeutic strategies to prevent and combat *mycoplasma*-associated infections.

In this chapter, a comprehensive molecular characterization of the recognition modes of *mycoplasma* cytoadhesins, P110 and P40/P90, interacting with different host sialylated N- and O-glycans, have been performed, by using a combination of biophysical techniques and computational methods. This integrated approach allowed to identify ligand epitopes in the protein binding site, explore the conformational behavior of N- and O-glycans interacting with bacterial cytoadhesins and construct 3D models of protein-ligand complexes.

The obtained results demonstrated that both P110 and P40/P90 recognize and bind various N- and O-sialoglycans. Particularly noteworthy is the first experimental evidence showing that *Mycoplasma* cytoadhesins can bind to naturally membrane-exposed O-glycans. Despite the cytoadhesins having different binding sites and shapes, Neu5Ac is the primary contributor to the binding in all complexes. Hydrogen bonds involving the Neu5Ac residue, coupled with hydrophobic interactions, anchor the entire sialoglycans to the binding sites of both cytoadhesins. Notably, the more open binding site and a higher flexibility of some loops near the binding site of

P40/P90 allow the accommodation of sialoglycans maintaining higher conformational freedom, unlike the recognition by P110, which induces conformer selections. Furthermore, NMR data revealed that neither P110 nor its orthologue P40/P90 demonstrated sialidase activity, despite the presence of a putative sialidase catalytic site.[125]

Worth noting, the mode of action of *Mycoplasma* cytoadhesins differs significantly from another group of bacterial adhesins recently analyzed in our research group, the Siglec-like adhesins, SLBR-H and SLBR-B from *S. gordonii (*cf. §1.5.2). Despite some similarities in the sugar binding side of the two classes of proteins, differently from *mycoplasma* cytoadhesins, SLBR-H and SLBR-B, accommodate not only the Neu5Ac residue of N- and O-glycans (with a preference for sialoglycans containing Neu5Ac-α-(2-3)-Gal), but also other sugar residues present in the ligands.[114] Instead, as discussed in this chapter, *mycoplasma* cytoadhesins show a binding mode very similar to mammalian Siglecs, which recognize prevalently Neu5Ac residues and, rarely, Gal residues.[22, 115]

In conclusion, the obtained results provide a more comprehensive understanding of the structural features governing the recognition of different sialoglycans by the main cytoadhesins from *Mpn* and *Mge*. These insights into the molecular basis of the ligands' specificity and the knowledge of the specific binding sites lay the groundwork for designing species-specific anti-adhesive compounds. This approach has been used, for example, in the synthesis of different heteroglycoclusters to target two *P. aeruginosa* lectins, LecA and LecB, to counteract either the bacterial adhesion in the human tissues and organs, and the development of antibiotic resistance by biofilm formation.[241] This potential development could lead to novel therapeutic strategies for treating or preventing *Mycoplasma* infections by inhibiting cytoadherence. Furthermore, comparative studies involving different *Mycoplasma* species and their cytoadhesins can offer insights into variations in binding specificity and mechanisms. This comparative approach may reveal commonalities and differences among various *Mycoplasma* strains.

# Chapter 6 – Isolation and characterization of CPS from an *Acinetobacter baumannii* clinical isolate

## 6.1 Introduction

*Acinetobacter baumannii*, ubiquitous aerobic Gram-negative opportunistic pathogen, is responsible for both community and health care – associated infections (HAIs). Particularly, it causes pulmonary, urinary tract, bloodstream and surgical wound infections.[242] It is one of the ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacter* spp.), a group of feared microorganisms holding significant importance due to their propensity to accumulate multiple antimicrobial resistance (AMR) mechanisms and their ability to spread within healthcare facilities and communities.[243]As in other bacteria, their cell envelope shield is the first line of defence against stress conditions and is crucial to resistance to antibiotics.

*A. baumannii* produces peculiar CPSs that cover its surface to protect itself against environmental stress, desiccation, phagocytosis and complement-mediated killing. Typically, the biosynthesis of CPS in *A. baumannii* includes genes present within the chromosomal K locus (KL) involved in capsule initiation (*itr*), glycosyltransfer (*gtr*), translocation (*wzx*), polymerization (*wzy*), and the incorporation of *non*-carbohydrate components.[244] *A. baumanni* CPS represents one of the major virulence factors and notably, some bacterial strains exhibit CPS containing complex atypical sugars, like nonulosonic acids which are structurally related to a typical monosaccharide present in the eukaryotic cell, the sialic acid (cf. §1.2.1). Thanks to their unique structures, these bacterial CPSs are able to mimic SAMPs, interact with inhibitory host receptors and elude immune response.

Build on these screens, the CPS from a clinical isolate of *A. baumannii* has been isolated and characterized, mainly by NMR spectroscopy and Mass spectrometry. Its conformational behavior in the free state has been studied. Moreover, after the conjugation of the characterized CPS to the carrier protein $CRM_{197}$, its immunogenic activity has been tested *in vivo* on mice, as depicted in the workflow represented in Figure 6.1.

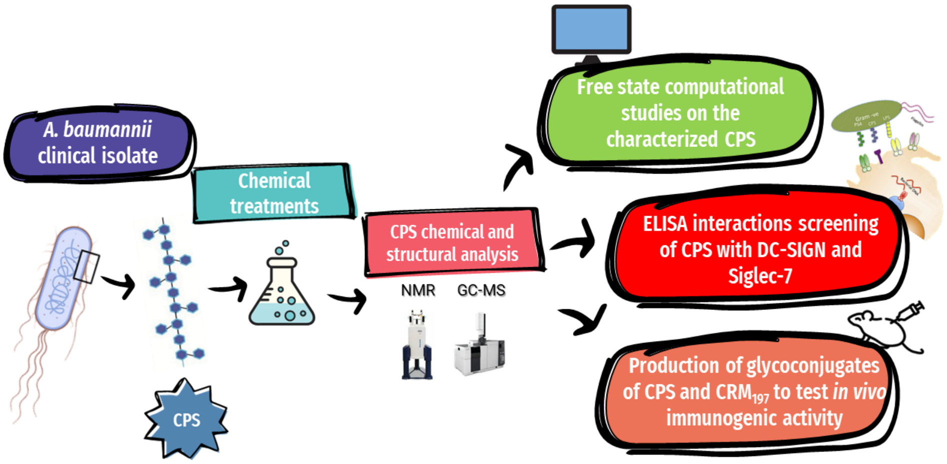


**Figure 6.1:** Workflow of the experimental activities.

This work lays the foundation for new antimicrobial strategies, different from the use of antibiotics, and sets the basis for possible glycoconjugates vaccines against *A. baumannii* infections.

## 6.2 Extraction, purification and characterization of CPS from an *A. baumannii* clinical isolate

### 6.2.1 CPS extraction and purification

Dried cells from an *A. baumannii* clinical isolate were provided by Prof. E. De Gregorio (Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II). After cell washing in water to remove eventual contaminants derived from cell growth, CPS has been extracted using the PCP and hot phenol-water extractions and purified following the procedures introduced in Chapter II (cf. also Chapter VII for methods). The determination of CPS structure has been performed by using MS and NMR techniques as follows.

### 6.2.2 CPS chemical and structural analysis by GC-MS and NMR spectroscopy

The compositional analysis of CPS from *A. baumannii* by means of AMG (cf. Chapter II) revealed the prevalence of hexoses and hexosamines, specifically Gal, GalNAc and GlcNAc, and the presence of the legionaminic acid (Leg). The presence of Kdo indicated a contamination of the sample with LOS (Figure 6.2).

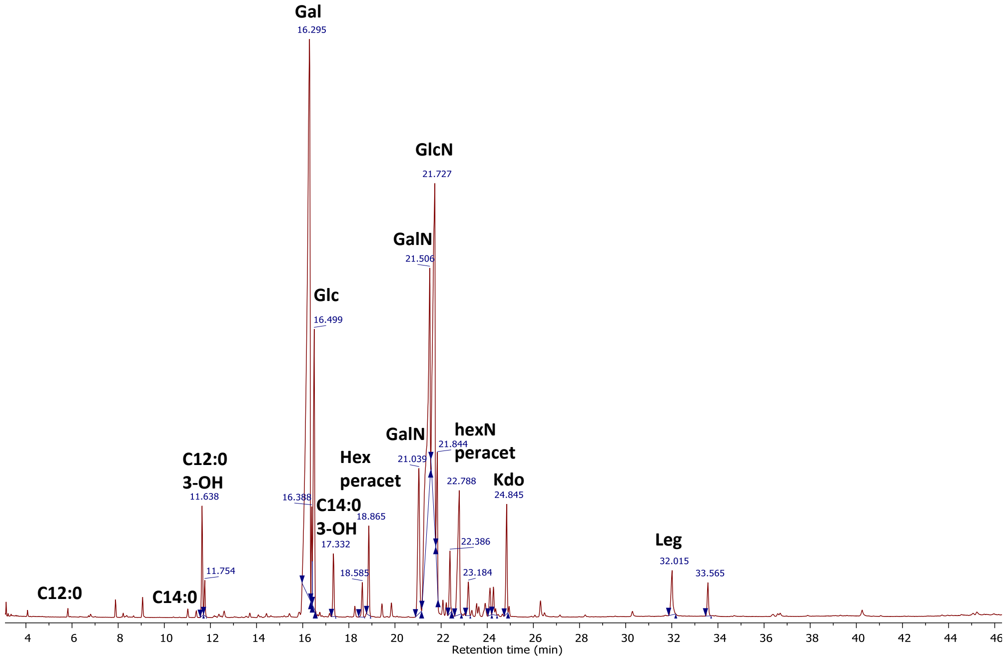


**Figure 6.2:** AMG analysis of *A. baumannii* CPS and assignment of the peaks by comparison with AMG standards.

Monosaccharides linkages analysis of pure CPS from *A. baumannii* by means of PMAA revealed the prevalence of 4,6-linked hexose, 3,6-linked hexose and 3-linked hexosamine (Figure 6.3).

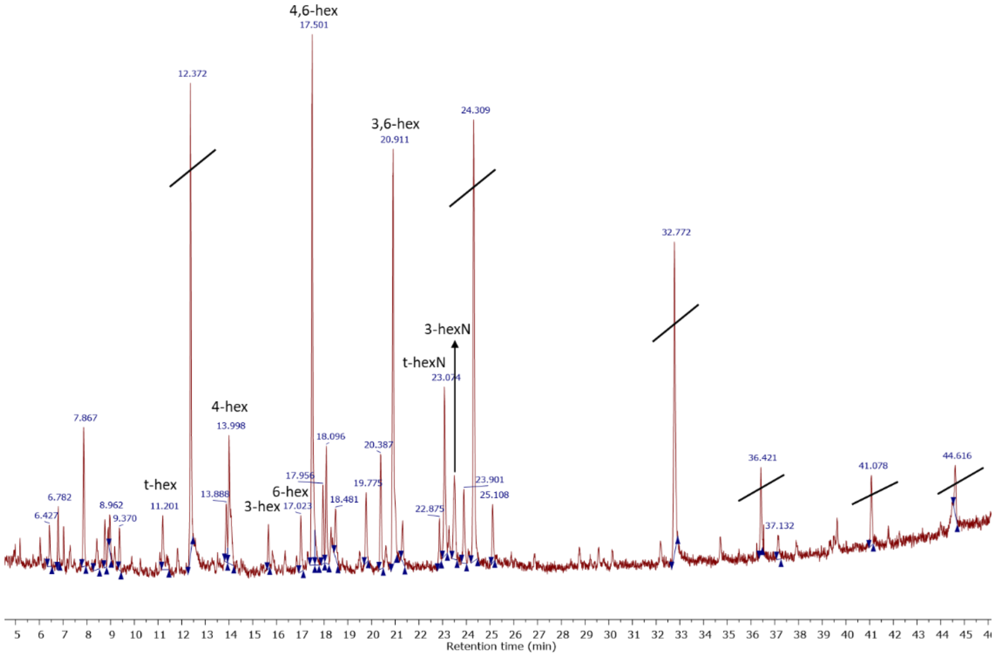


**Figure 6.3:** PMAA analysis of *A. baumannii* CPS.

Then, a combination of mono- and bi-dimensional NMR experiments (cf. Chapter II) has been performed to derive the structure of the CPS repeating unit. As example, the superimposition of HSQC and $^1$H spectra of the isolated CPS have been reported in Figure 6.4. As showed in figure, four anomeric signals were identified, together with the characteristic signals of the methylene protons at position 3 of Leg residue. Thus, the CPS repeating unit was composed by a pentasaccharide composed of β-D-GalNAc, β-D-Gal, β-D-Gal, α-D-GlcNAc and α-Leg *non*-stoichiometrically substituted by an acetyl group and an hydroxybutyryl group (Figure 6.4 and Table 7).

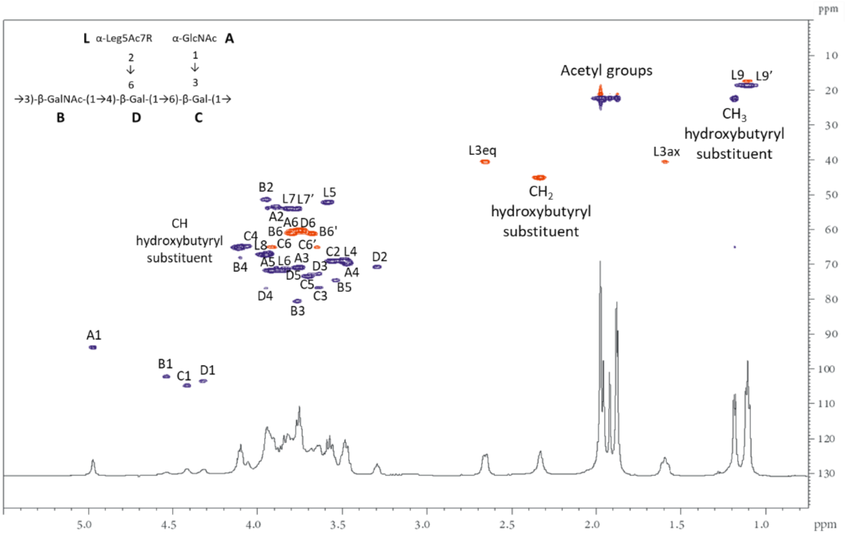


**Figure 6.4:** HSQC and $^{1}$H spectra of CPS from *A. baumanni*.

**Table 7:** $^{1}$H, $^{13}$C chemical shift values of the CPS from *A. baumannii*. (cf. Figure 6.4).

| ***A. baumannii CPS*** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Unit** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | 4.97 | 3.89 | 3.75 | 3.46 | 3.90 | 3.79 | | | |
| **α-GlcNAc** | 93.8 | 53.6 | 70.8 | 69.6 | 71.8 | 60.8 | | | |
| | | | | | | | | | |
| **B** | 4.53 | 3.94 | 3.76 | 4.10 | 3.53 | 3.67 | | | |
| **β-GalNAc** | 102.1 | 51.4 | 80.5 | 68.0 | 74.6 | 61.1 | | | |
| | | | | | | | | | |
| **C** | 4.41 | 3.55 | 3.64 | 4.06 | 3.70 | 3.91/3.64 | | | |
| **β-Gal** | 104.9 | 69.3 | 76.7 | 64.9 | 73.4 | 65.1 | | | |
| | | | | | | | | | |
| **D** | 4.32 | 3.29 | 3.63 | 3.94 | 3.71 | 3.73 | | | |
| **4-β-Gal** | 103.5 | 70.8 | 72.7 | 76.9 | 73.4 | 60.3 | | | |
| | | | | | | | | | |
| **L** | | | 2.65/1.60 | 3.48 | 3.57 | 3.84 | 3.81 | 3.94 | 1.10 |
| **α-Leg** | | | 40.5 | 68.8 | 52.1 | 71.5 | 54.0 | 67.1 | 18.9 |
| | | | | | | | | | |

Given the difficulties in understanding how Leg residue was linked to the CPS backbone, due to the absence of key NOEs and crucial signals in the HMBC spectra, the CPS was cleaved by mild acidic hydrolysis with an aqueous solution of AcOH 6%, taking the advantage of the lability of the Leg glycosidic linkage. The polysaccharide was further purified by performing a bio-gel P10, obtaining two fractions then analyzed by GC-MS and NMR: the first containing the CPS backbone without Leg (Figure 6.5, Figure 6.6 and Table 8) and the second containing only Leg in solution (Figure 6.7 and Table 9).

The comparison of the chromatogram of PMAA of the entire CPS structure with the one of hydrolized CPS lacking the Leg residue, helped to fully define the structure. Indeed, in the first chromatogram, the predominant species was a 4,6-linked hexose unit (Figure 6.3), while the one acquired on the hydrolyzed CPS showed that the predominant species was a 4-linked hexose (Figure 6.5). This evidence suggested that α-Leg was linked to the central β-Gal with a 2,6 linkage.

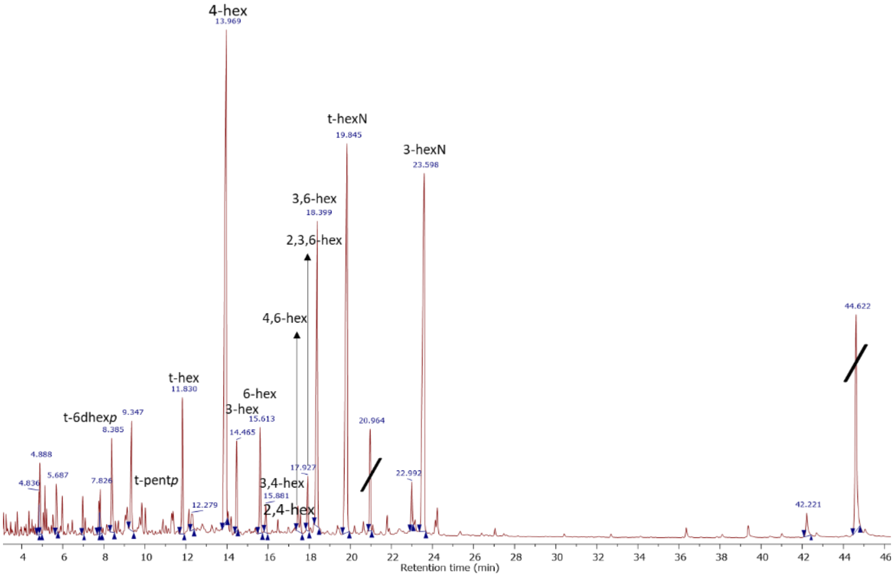


**Figure 6.5:** PMAA analysis of hydrolyzed *A. baumannii* CPS.

These results were also confirmed by NMR analysis (Figure 6.6 and Table 8).

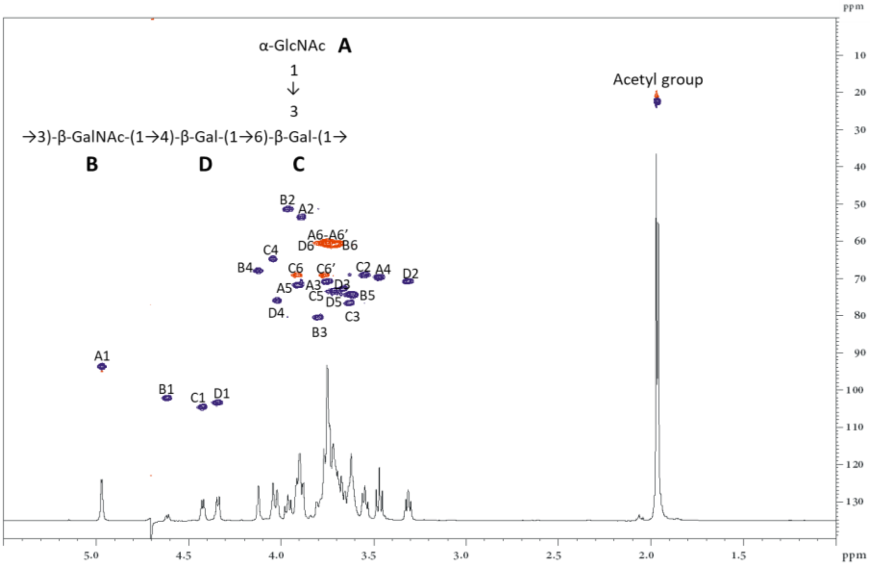


**Figure 6.6:** HSQC and $^{1}$H spectra of hydrolyzed CPS from *A. baumannii*.

**Table 8:** $^{1}$H, $^{13}$C chemical shift values of the hydrolyzed CPS from *A. baumannii*. (cf. Figure 6.6).

| ***A. baumannii* hydrolyzed CPS** | | | | | | |
|---|---|---|---|---|---|---|
| **Unit** | **1** | **2** | **3** | **4** | **5** | **6** |
| **A** | 4.97 | 3.89 | 3.74 | 3.47 | 3.90 | 3.74/3.68 |
| **α-GlcNAc** | 93.8 | 53.5 | 70.8 | 69.7 | 71.8 | 60.2 |
| | | | | | | |
| **B** | 4.61 | 3.96 | 3.79 | 4.12 | 3.61 | |
| **β-GalNAc** | 102.2 | 51.4 | 80.6 | 68 | 70.4 | |
| | | | | | | |
| **C** | 4.42 | 3.54 | 3.62 | 4.04 | 3.70 | 3.91/3.70 |
| **β-Gal** | 104.7 | 69.2 | 76.7 | 64.7 | 73.4 | 69.12 |
| | | | | | | |
| **D** | 4.34 | 3.31 | 3.66 | 4.02 | 3.64 | 3.76/3.71 |
| **4-β-Gal** | 103.5 | 70.7 | 72.8 | 75.9 | 74.4 | 60.4 |
| | | | | | | |

NMR experiments on the Leg residue, alone in solution, were also performed confirming the configuration of its stereogenic centers (Figure 6.7 and Table 9).

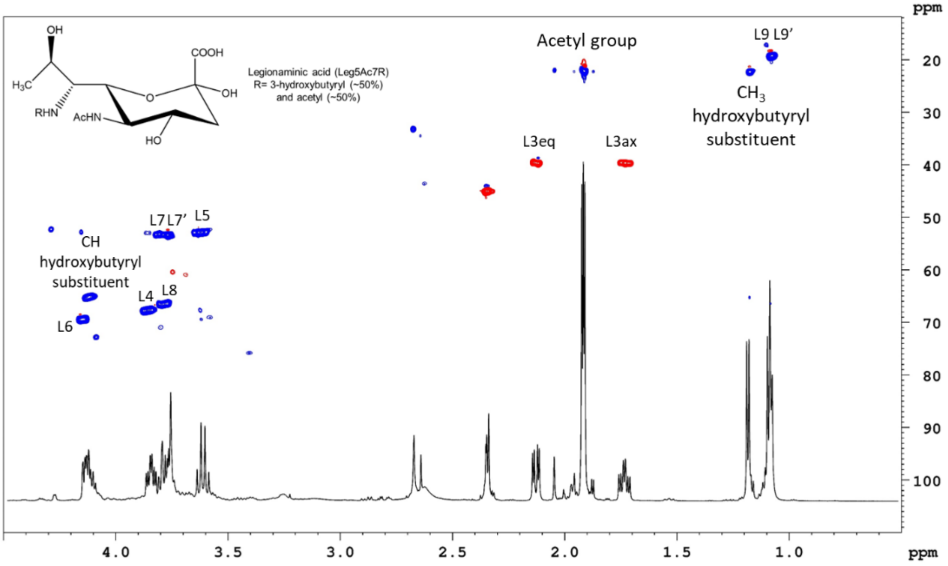


**Figure 6.7:** HSQC and $^1$H spectra of free Leg.in solution.

**Table 9:** $^1$H, $^{13}$C chemical shift values of the reducing Leg (cf. Figure 6.7).

| **Reducing Leg** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Unit** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **L** | 1.72/2.11 | 3.85 | 3.62 | 4.14 | 3.76 | 3.78 | 1.07 |
| **α-Leg** | 39.6 | 67.6 | 52.8 | 69.4 | 53.3 | 66.4 | 19.2 |
| | | | | | | | |

Altogether the obtained results showed that the CPS from this *A. baumanni* clinical isolate is made up of a pentasaccharide repeating unit (Figure 6.8).

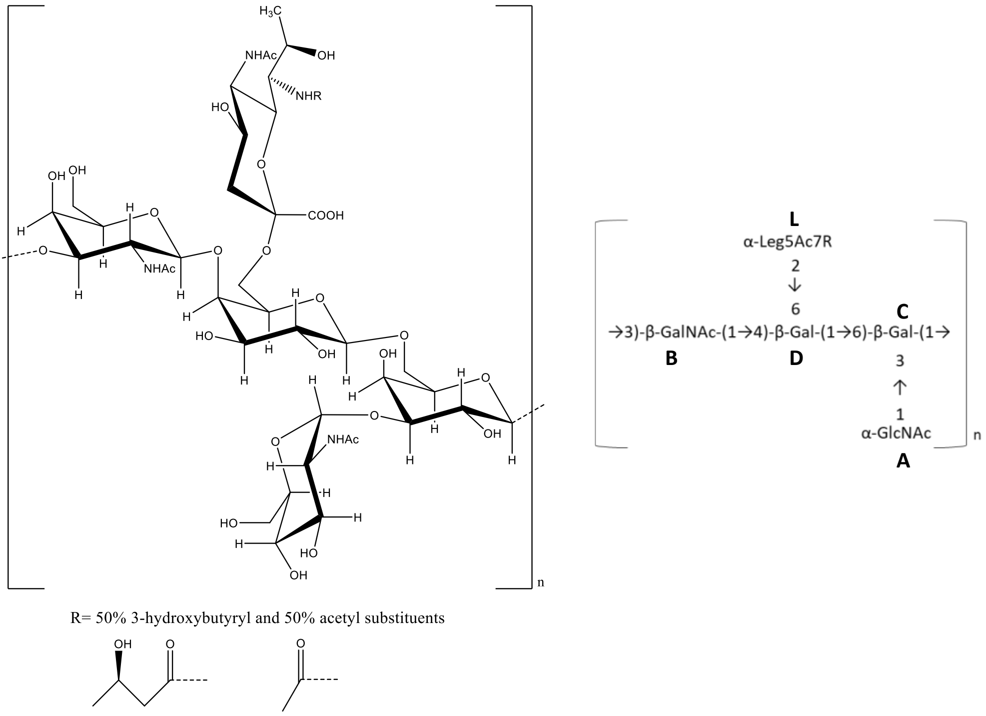


**Figure 6.8:** CPS repeating unit from the *A. baumannii* clinical isolate, letter code in agreement with NMR assignment.

## 6.3 Computational studies on characterized CPS from *A. baumannii*

Once determined the CPS chemical structure from the *A. baumannii* clinical isolate, computational studies were performed to explore its conformational behavior by using AMBER18 software.[168] In particular, a comparison of the 3D behavior of the CPS repeating unit before and after the hydrolysis (thus, in the presence and in the absence of Leg) has been made.

### 6.3.1 Molecular mechanics and molecular dynamics of CPS without Leg

Firstly, the analysis of an oligosaccharide composed by the single CPS repeating unit but not bearing Leg, has been performed.

Initially, the conformational preferences of all the constituting disaccharides repeating unit were determined by building their potential energy surfaces as a function of the glycosidic torsional angles Φ and Ψ. The calculated maps suggested that all the disaccharide mainly adopted conformations according to the *exo*-anomeric orientation (Figure 6.9).

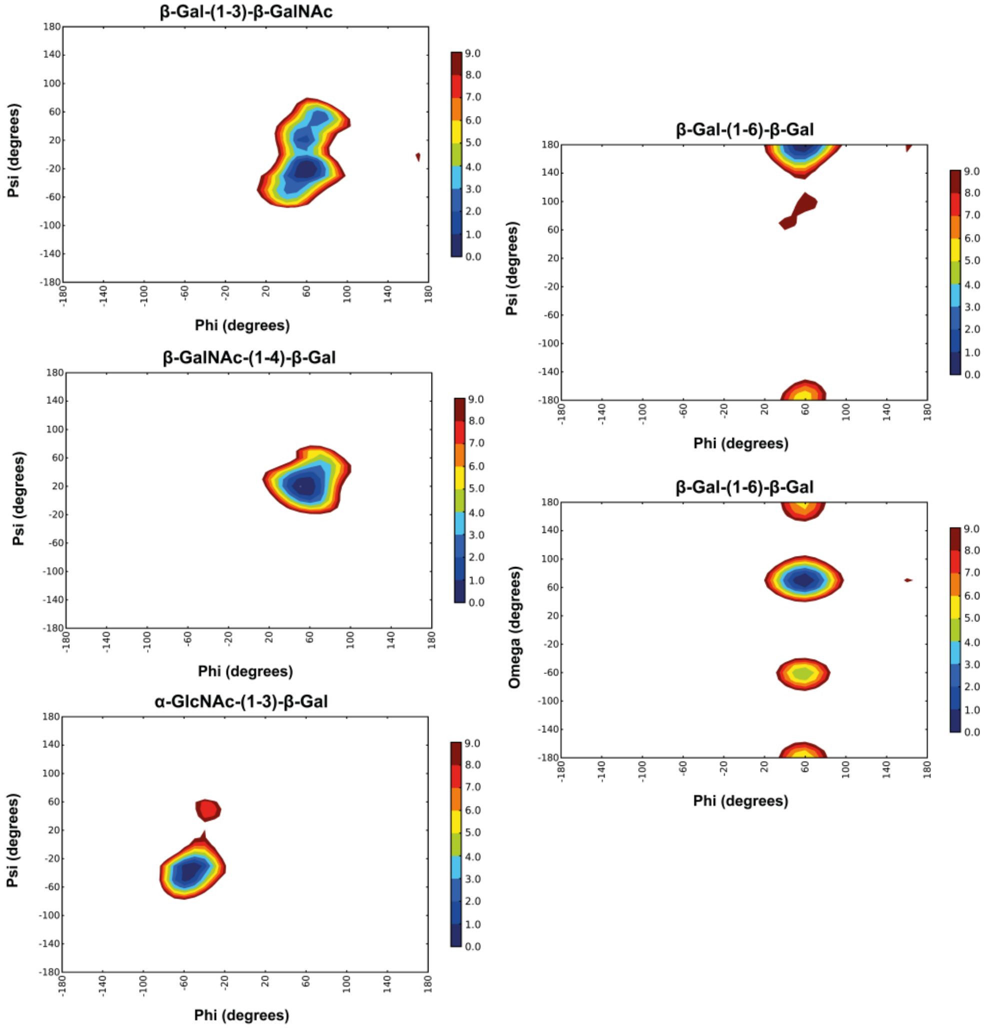


**Figure 6.9**: Adiabatic maps representing the conformational spaces of phi (φ) vs psi (ψ) and phi (φ) vs omega (ω) torsional angles of the composing disaccharides studied by MM. The colour scheme reflets the energy distribution within the conformational landscape, with red indicating energy maxima and blue representing energy minima. The color gradient provides insight into the stability and energetics of different torsional conformations, in some cases, showing even various minimums, aiding in the analysis of carbohydrate structural dynamics.

Subsequently, starting from the favorite global minima, a hexasaccharide fragment representing the CPS repeating unit with capping ends was built and its conformational behavior was disclosed by using molecular dynamic simulations in explicit solvent with AMBER. (Figure 6.10).

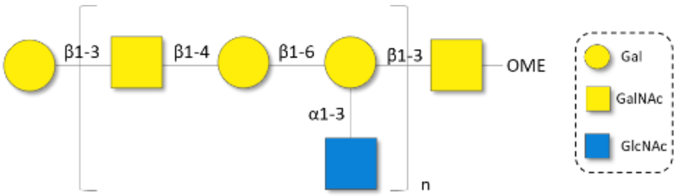
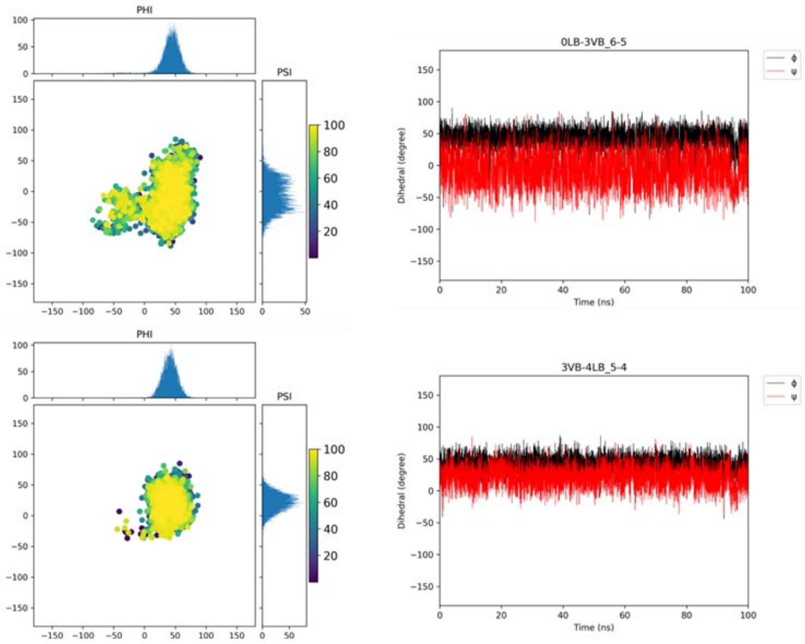


**Figure 6.10:** Schematic representation of the repeating unit of CPS from the *A. baumannii* clinical isolate without Leg residue, coloured according to SNFG nomenclature.

MD simulations were run in explicit solvent, using a water box around the ligand in order to study its free state conformational behavior for 100 ns. Indeed, the analysis of the dihedral angles fluctuations along the 100 ns allowed to better understand the flexibility of the ligand (Figure 6.11).



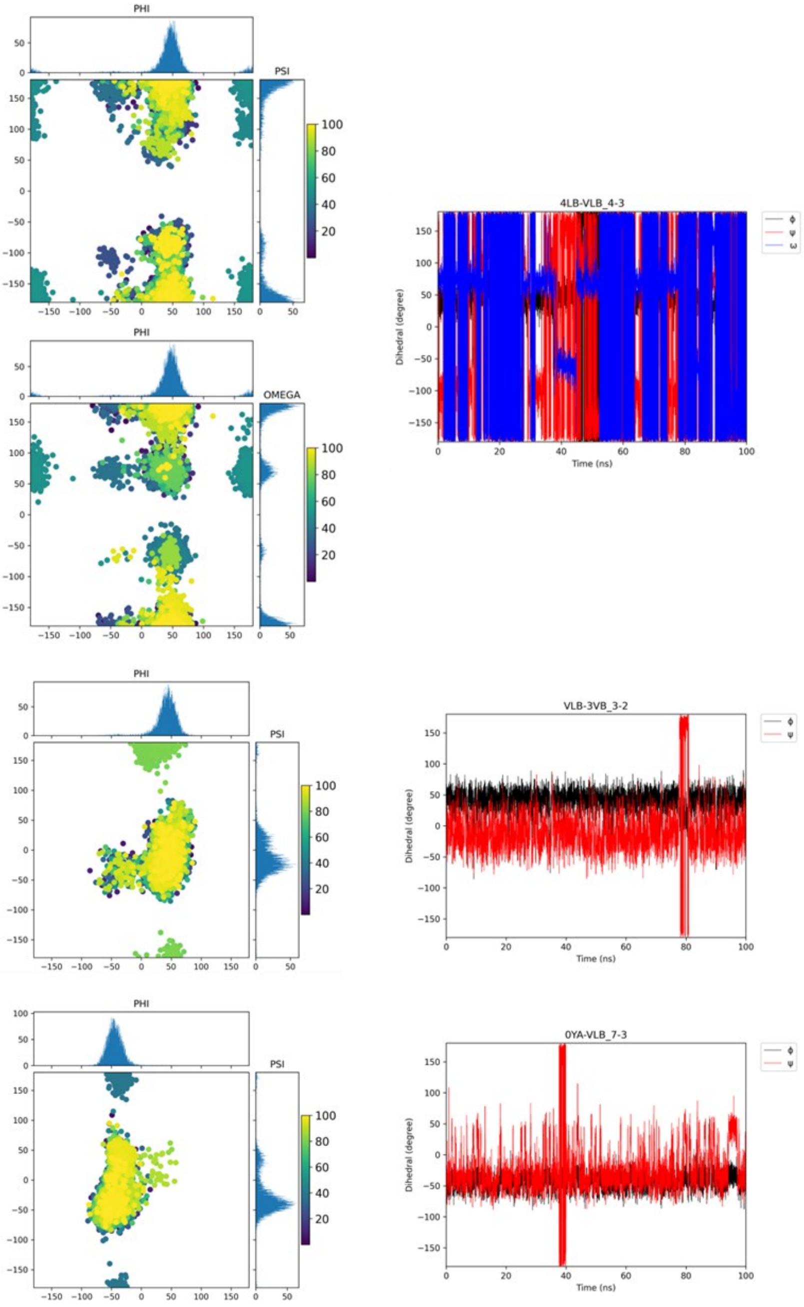


**Figure 6.11:** Free state MD analysis of CPS repeating unit without Leg residue. Left panel: Dihedral angles analysis, represented as scatter plots of the ϕ torsions against ψ and ϕ torsions against ω, during the MD simulation with the relative histograms to represent the most populated energies. Right panel: Fluctuations of the ϕ (black), ψ (red) and ω (blue) dihedral torsions around the glycosidic linkages along the 100ns MD simulation in the free state.

As showed in Figure 6.11, the torsion angles along β-GalNAc-(1-4)-β-Gal and α-GlcNAc-(1-3)-β-Gal glycosidic linkages were stable. Two different families were observed for the β-Gal-(1-3)-β-GalNAc and β-Gal-(1-3)-β-GalNAc glycosidic linkages with ϕ = 50° and Ψ = -20°/20° (with a 50:50 and 70:30 ratio, respectively), as often occur for the capping end residues. More importantly, fluctuations of the ω torsion were observed around the β-Gal-(1-6)-β-Gal glycosidic linkage. Indeed, two possible minima, characterized by an omega value of 180° (*tg* conformation) or 60° (*gt* conformation), with a 70:30 ratio, were observed (Figure 6.12), thus conferring a higher degree of flexibility to the ligand.

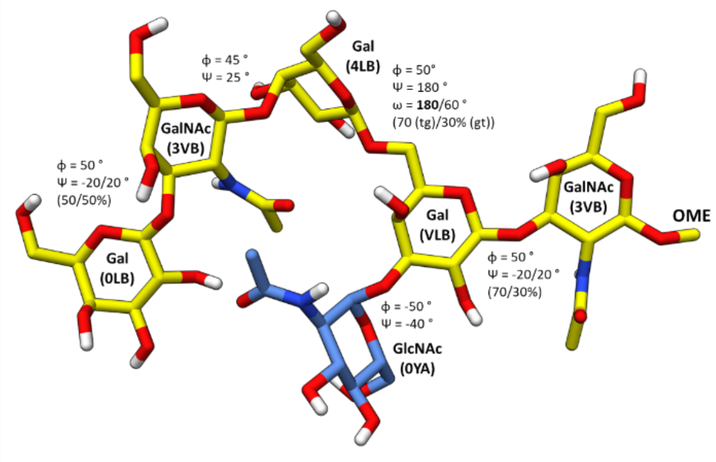


**Figure 6.12:** 3D representative pose of the most populated cluster coming from the 100 ns MD simulation for the free state of one repeating unit of the CPS without Leg residue, coloured according to the SNFG nomenclature. The values of the torsion angles ϕ, ψ and ω for each glycosidic linkage are also shown.

### 6.3.2 Molecular mechanics and molecular dynamics of entire CPS

The conformation of the CPS repeating unit, including Leg, was then explored to evaluate if its negative charge could affect the tridimensional arrangement of the ligand. Since no parameters are available for Leg on GLYCAM website, as for most bacterial monosaccharides, it was necessary to parametrize it (Figure 6.13). Parametrization means defining specific parameters like atomic charges, bond lengths, bond angles, torsional angles, and *non*-bonded interactions within a sugar molecule, integrating experimental data, quantum mechanical calculations and empirical fitting. The aim is to ensure that sugar molecules behave realistically in simulations by accurately defining their potential energy surfaces. This was achieved through a protocol for *non*-common bacterial sugars, optimized in my research group.

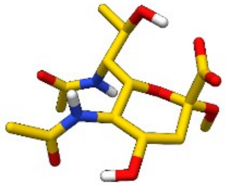


**Figure 6.13:** Chemical structure of α-Leg5Ac7Ac, coloured according to SNFG nomenclature. α-Leg5Ac7Ac has been parametrized by Dr. F. Nieto-Fabregat, in our group at Department of Chemical Sciences, University of Naples Federico II.

MM was performed to obtain adiabatic maps of the disaccharide linkage α-Leg5Ac7Ac-(2-6)-β-Gal. (Figure 6.14).

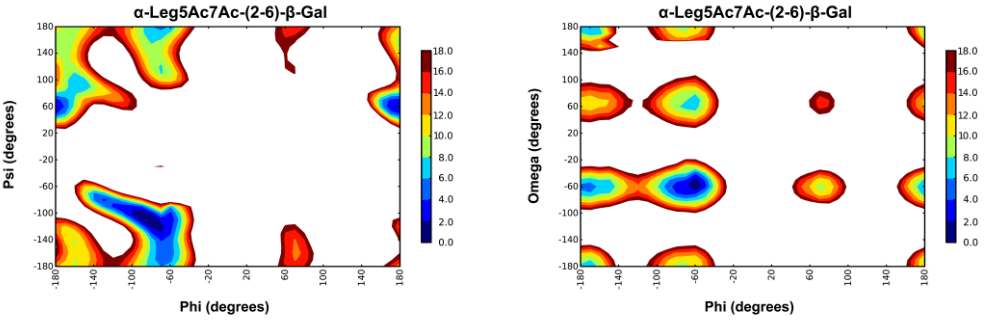


**Figure 6.14:** Adiabatic maps representing the conformational spaces of phi (φ) vs psi (ψ) and phi (φ) vs omega (ω) torsional angles of the α-Leg5Ac7Ac-(2-6)-β-Gal disaccharide studied by MM. The colour scheme reflets the energy distribution within the conformational landscape, with red indicating energy maxima and blue representing energy minima. The colour gradient provides insight into the stability and energetics of different torsional conformations, showing even various minima, aiding in the analysis of carbohydrate structural dynamics.

To explore in detail the preferred conformation of the entire CPS repeating unit (Figure 6.15) in its free state, a MD simulation in explicit solvent, using a water box around the entire ligand, was run for 100 ns. The analysis of the dihedral angles fluctuations along the 100 ns are reported in Figure 6.16.

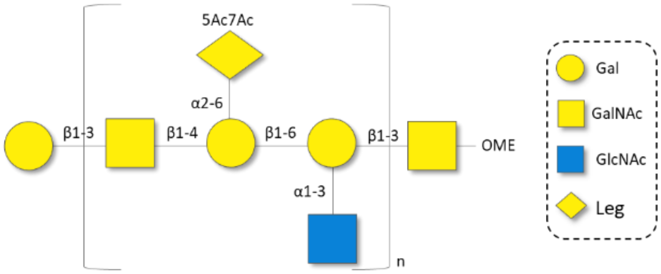
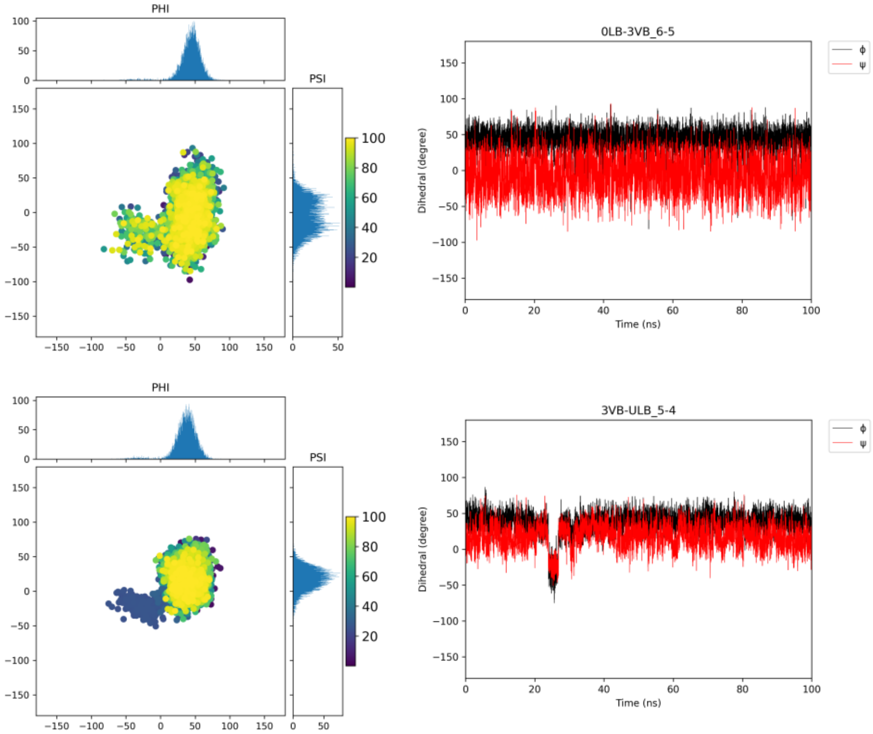
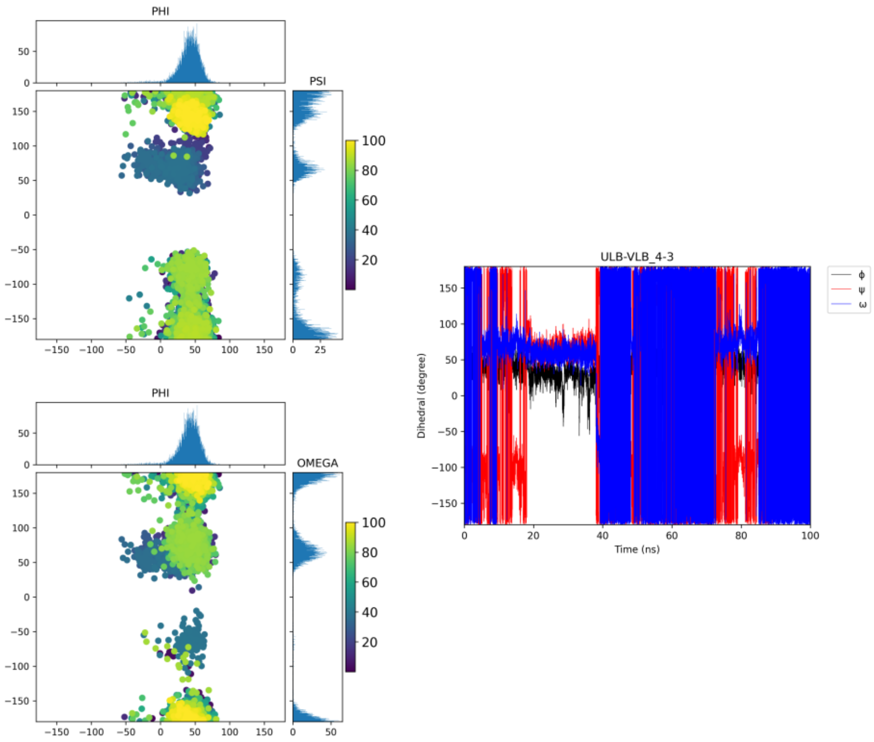
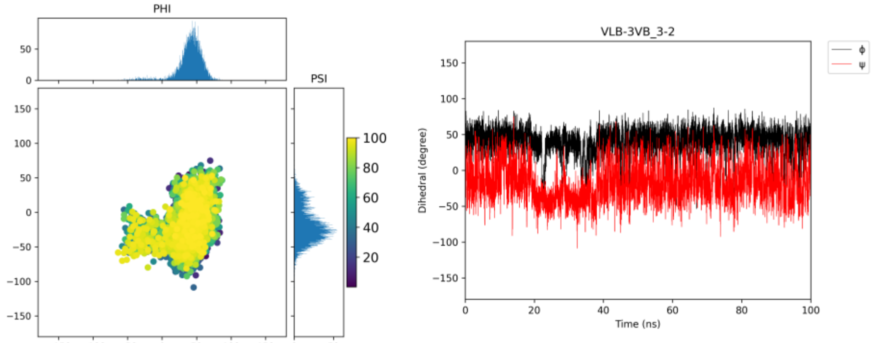


**Figure 6.15:** Schematic representation of the repeating unit of CPS from the *A. baumannii* clinical isolate, coloured according to SNFG nomenclature.




PHI
PSI
PHI
OMEGA
ULB-VLB_4-3
ϕ
ψ
ω
Dihedral (degree)
Time (ns)
100
80
60
40
20


PHI
PSI
VLB-3VB_3-2
ϕ
ψ
Dihedral (degree)
Time (ns)
100
80
60
40
20

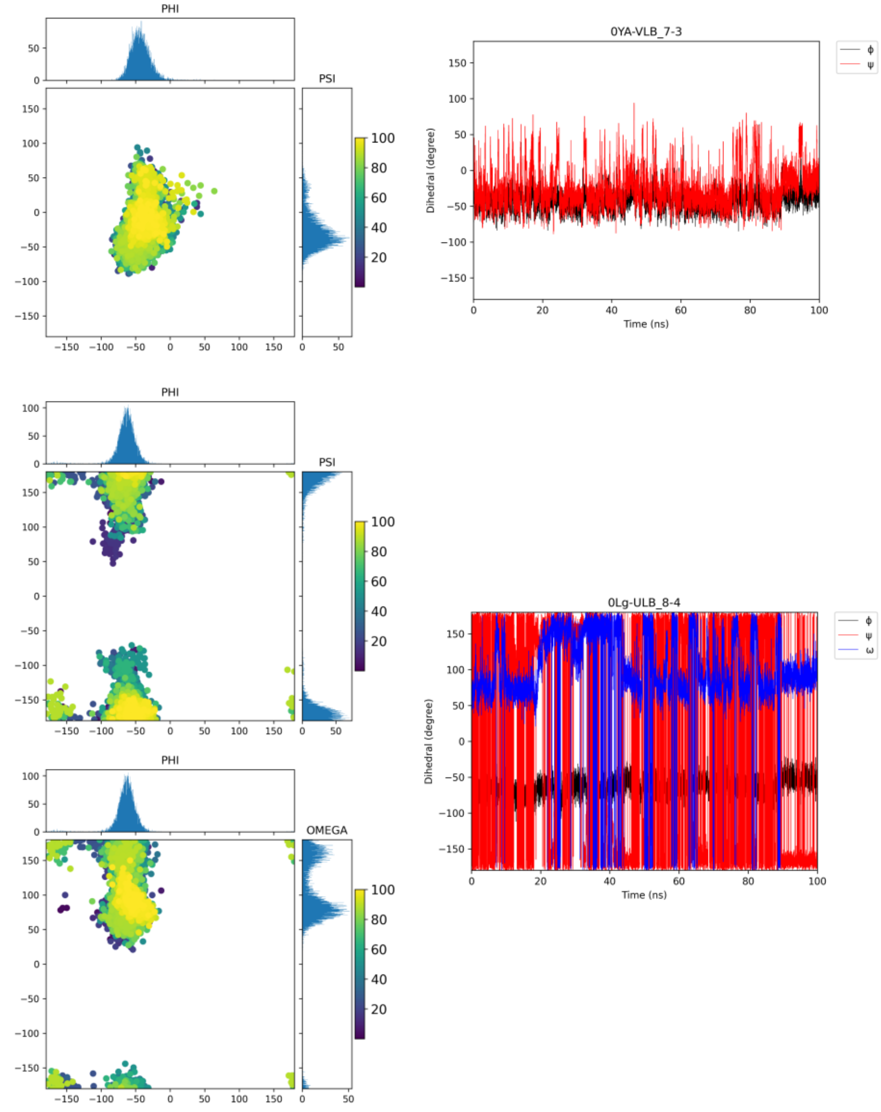


**Figure 6.16:** Free state MD analysis of CPS repeating unit. Left panel: Dihedral angles analysis, represented as scatter plots of the ϕ torsions against ψ and ϕ torsions against ω during the MD simulation with the relative histograms to represent the most populated energies. Right panel: Fluctuations of the ϕ (black), ψ (red) and ω (blue) dihedral torsions around the glycosidic linkages along the 100ns MD simulation in the free state.

As showed in Figure 6.16, also in this case, the torsion angles along β-GalNAc-(1-4)-β-Gal and α-GlcNAc-(1-3)-β-Gal glycosidic linkages remained stable. On the contrary, for the β-Gal-(1-3)-β-GalNAc and β-Gal-(1-3)-β-GalNAc torsions, as expected, there were variations of the ψ torsions, as they were the capping ends of the ligand. Fluctuations through the β-Gal-(1-6)-β-Gal glycosidic linkage were also observed, having stability for both the ϕ and ψ torsions with values of 50° and 180° respectively. Nonetheless, two possible minima were populated due to the flexibility of the ω torsion, that could adopt a value of 180° (*tg* conformation) or 60° (*gt* conformation), with a 60:40 ratio, respectively. The presence of an additional 6-linked sugar, further contributed to confer flexibility to the ligand. Indeed, the ω torsion around α-Leg5Ac7Ac-(2-6)-β-Gal linkage, could adopt values of 60° (*gt* conformation) and 180° (*tg* conformation) with a 60:40 ratio, respectively. All this information can be found sum up in Figure 6.17.

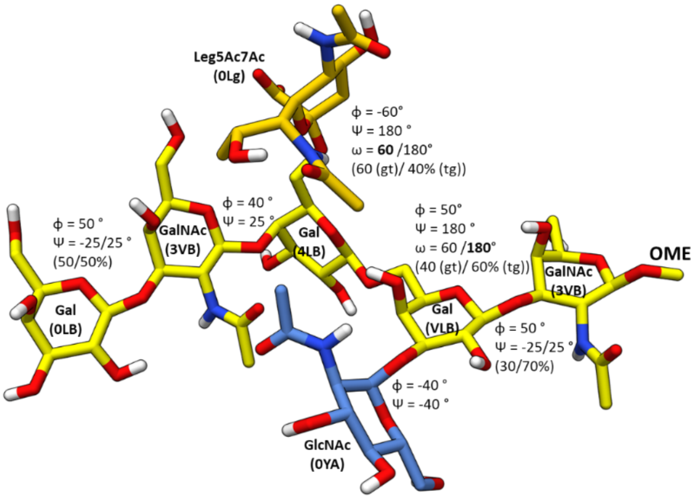


**Figure 6.17:** 3D representative pose of the most populated cluster coming from the 100 ns MD simulation for the free state of one repeating unit of the entire CPS with Leg5Ac7Ac residue, coloured according to the SNFG nomenclature. The values of the torsion angles ϕ, ψ and ω for each glycosidic linkage are also shown.

## 6.4 ELISA interactions screening: CPS from *A. baumannii* can interact with DC-SIGN and Siglec-7

Preliminary studies were carried out in order to identify potential binder of *A. baumanni* CPS. ELISA was conducted to evaluate the interaction between *A.*

*baumannii* CPS and various recombinant human lectins. Thanks to a collaboration with Dr. F. Chiodo (Amsterdam Universitair Medische Centra, Netherlands), three CPS samples were tested, namely: (i) CPS1, ultracentrifuged sample, (ii) CPS2, sample after acetic acid hydrolysis, and (iii) CPS3, further purified sample through gel filtration chromatography. ELISA plate wells were coated with 50 µg/ml of each CPS in PBS and compared to the typical lectin's interaction motifs used as positive controls. Positive controls were covalently bound to polyacrylamide to ensure their adsorption into the plate wells. Interactions were evaluated by measuring the Abs at 450 nm.

Fc-DC-SIGN and Fc-Siglec-7 were able to interact with CPS3 (Figure 6.18A and B).

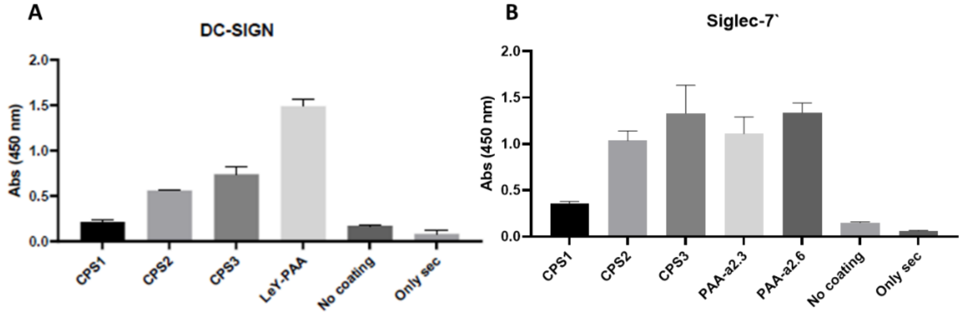


**Figure 6.18:** ELISA analysis between Fc-DC-SIGN (A) and Fc-Siglec-7 (B) and three purification steps of CPS. LeY-PAA, PAA-α2,3 and PAA-α2,6 are the positive controls.

These results will pave the way for further interaction studies between CPS and host receptors, based on NMR spectroscopy, computational methods, and other biophysical techniques.

## 6.5 Production of glycoconjugates to test *in vivo* immunogenic activity

Glycoconjugate vaccines have revolutionized the fight against bacterial infections.[245] Carbohydrates on bacterial surfaces, like CPS, play a vital role in infection and have been targeted for vaccine development. Glycoconjugate vaccines consists of carbohydrates covalently linked to carrier proteins, triggering T-cell-dependent immune responses.[246] Their success is evident in preventing diseases caused by pathogens like *N. meningitidis*,[247] and recent approvals of pneumococcal vaccines underscore their ongoing public health impact.[248] Furthermore, in the era of post COVID-19 pandemic, vaccine development remains a priority for governments to combat future pandemics, being identified as critical tools against evolving pathogens.[249] Traditional antibiotics are increasingly ineffective due to bacterial evolution, making glycoconjugate vaccines a promising solution to combat diseases caused by AMR pathogens, including the ESKAPE group, which encompasses *A. baumannii*.[250]

Advances in glycan chemistry have enabled the development of glycoconjugate vaccines with well-defined and highly pure carbohydrate antigens, able to strongly stimulate immune responses. In the following paragraphs, insights into the production of glycoconjugates starting from the characterized CPS from an *A. baumannii* clinical isolate will be provided, then, their immunogenic activity will be evaluated *in vivo*.

This work has been carried out during my stay at GSK (Siena, Italy), under the supervision of Dr. M.R. Romano and Dr. R. Adamo.

### 6.5.1 Analytical characterization of CPS by SEC-HPLC

Before starting with the conjugation, analytical studies to characterize CPS in terms of MW and LOS contamination have been performed by using Size Exclusion Chromatography–High Performance Liquid Chromatography (SEC-HPLC). In these procedures, two CPS samples were used: (i) CPS post extraction and (ii) CPS post ultracentrifugation (UC), to eventually control differences between two different steps of purification (cf. §1.2.1).

#### 6.5.1.1 MW evaluation of the CPS by SEC-HPLC

The evaluation of the MW of the two CPS samples was performed by comparison with pullulans in a range of MW from 20 kDa to 2500 kDa (Figure 6.19).

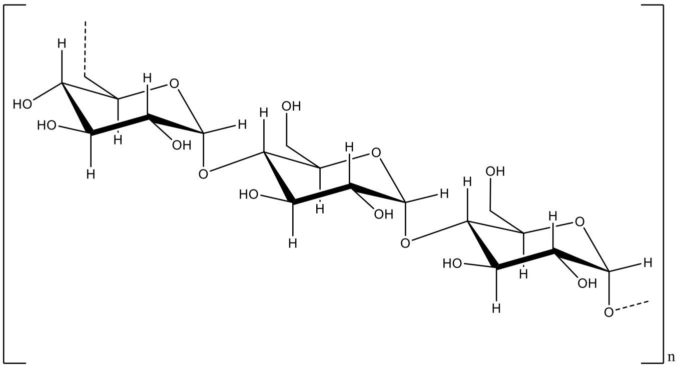


**Figure 6.19:** Pullulan chemical formula. It is a polysaccharide polymer consisting of maltotriose units. Three glucose units, connected by an α-1,4 glycosidic bond compose a maltotriose unit, whereas consecutive maltotriose units are connected to each other by an α-1,6 glycosidic bond.

The two samples and the pullulans in buffer 10 mM NaPi pH 7.2 were injected in a TSK6000PW column. The resulting chromatogram indicated that both the CPS samples possessed a MW higher than 2500 kDa (Figure 6.20).

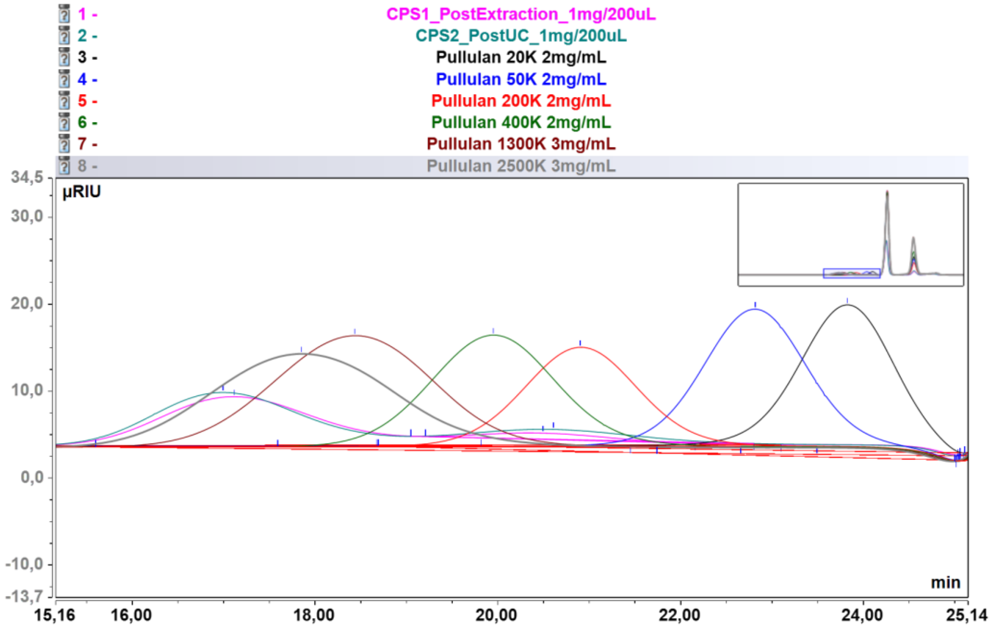


**Figure 6.20:** Chromatogram of MW evaluation of CPS samples in comparison with pullulans by SEC-HPLC.

### 6.5.1.2 Evaluation of LOS contamination in CPS by SEC-HPLC

The evaluation of LOS contamination in CPS has been performed by quantifying reactive carbonyl groups of different molecular size populations, after derivatization with semicarbazide (SCA) (Figure 6.21).

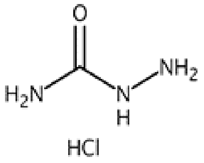


**Figure 6.21:** Semicarbazide chemical formula.

Also in this case, derivatized SCA-CPS post extraction (indicated as CPS1 in Figure 6.22) and derivatized SCA-CPS post UC (indicated as CPS2 in Figure 6.23) were used, compared to the blanks (not derivatized CPS1 and CPS2) and injected in a TSK3000PW. The presence of a peak related to LOS contaminant was observed only in the derivatized SCA-CPS post extraction (CPS1), as shown in Figure 6.22 and Figure 6.23. This result confirmed the utility of the ultracentrifugation step in LOS removal.

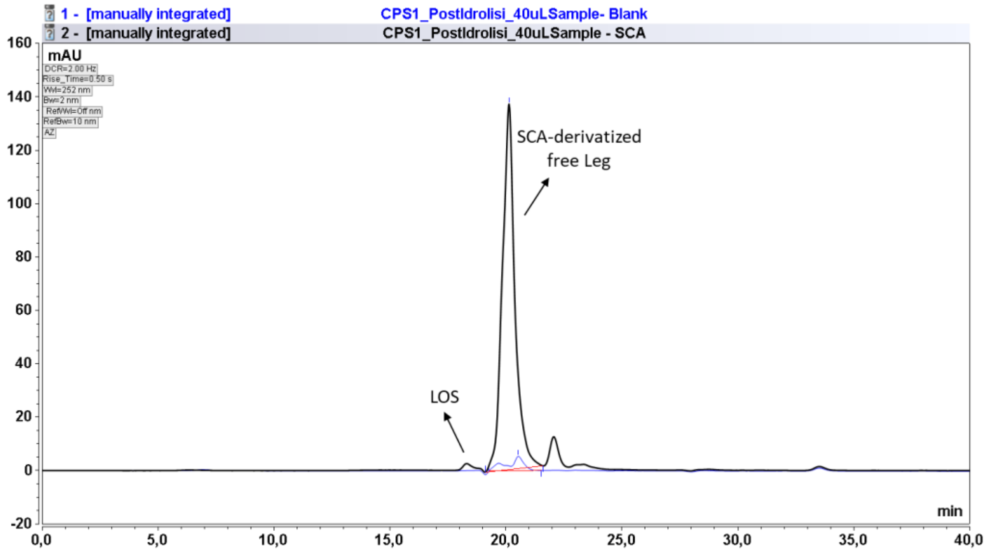


**Figure 6.22:** Chromatogram of derivatized SCA-CPS post extraction (black line) compared to blank (blue line).

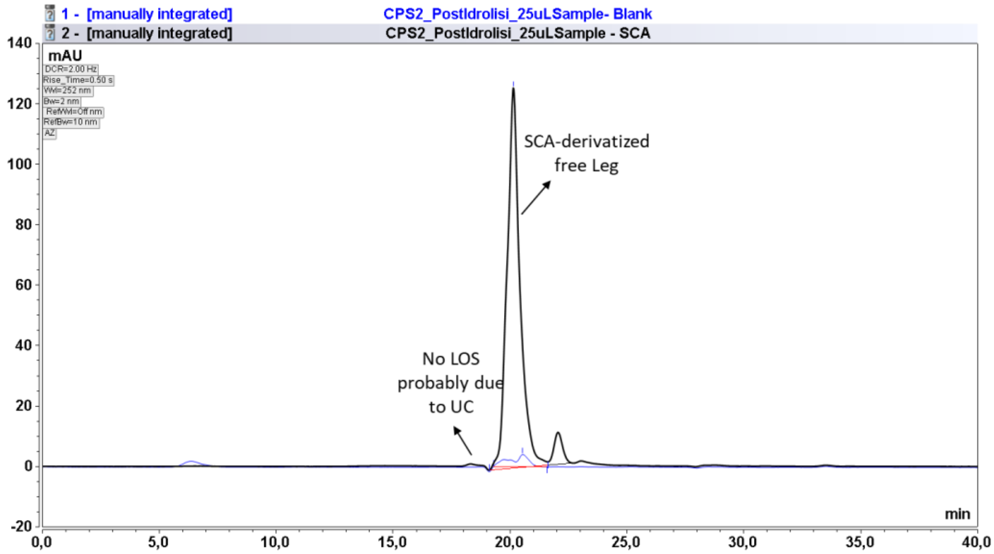
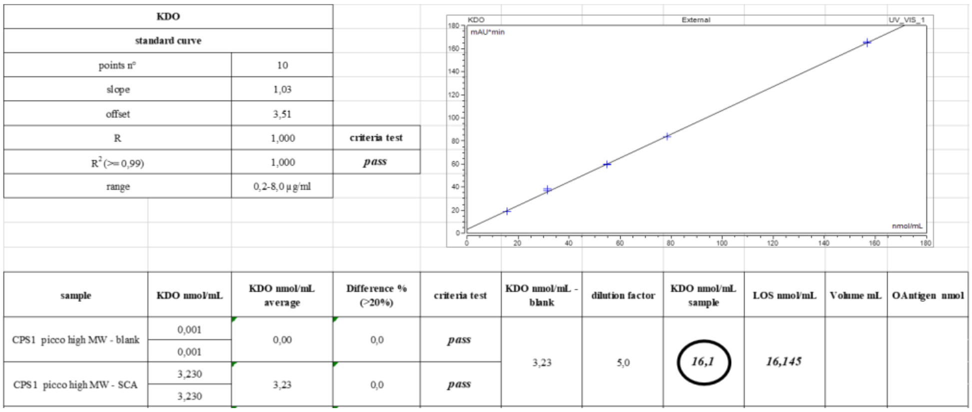


**Figure 6.23:** Chromatogram of derivatized SCA-CPS post UC (in black) compared to blank (in blue).

LOS contamination in CPS1 has been quantified through a derivatized SCA-Kdo calibration curve, measuring the Abs at 252 nm (Figure 6.24). In particular, being the ratio SCA-Kdo:LOS in nmol/mL 1:1, it was possible to assess that due to the presence of 3 Kdo molecules in CPS sample from *A. baumannii,* the total quantity of LOS contaminant, 16,1 nmol/mL reported in Figure 6.24, could be divided for 3, giving as result 3,4 nmol/mL.



| KDO | | |
|---|---|---|
| **standard curve** | | |
| points n° | 10 | |
| slope | 1,03 | |
| offset | 3,51 | |
| R | 1,000 | **criteria test** |
| $R^2$ (>=0,99) | 1,000 | ***pass*** |
| range | 0,2-8,0 µg/ml | |

| sample | KDO nmol/mL | KDO nmol/mL average | Difference % (>20%) | criteria test | KDO nmol/mL - blank | dilution factor | KDO nmol/mL sample | LOS nmol/mL | Volume mL | OAntigen nmol |
|---|---|---|---|---|---|---|---|---|---|---|
| CPS1 picco high MW - blank | 0,001<br>0,001 | 0,00 | 0,0 | ***pass*** | 3,23 | 5,0 | ***16,1*** | ***16,145*** | | |
| CPS1 picco high MW - SCA | 3,230<br>3,230 | 3,23 | 0,0 | ***pass*** | | | | | | |

**Figure 6.24:** SCA-Kdo calibration curve and relative calculations for quantification of LOS in nmol/mL in CPS1.

## 6.5.2 Random conjugation of CPS to CRM197 carrier protein, using CDAP chemistry

To produce the glycoconjugates, two different ultra centrifugated CPS samples (cf. §1.2.1) from the *A. baumannii* clinical isolate were used: (i) the entire CPS and (ii) hydrolyzed CPS, in order to remove Leg residue and evaluate, also in this case, the effects related to its presence, in the immunological response. Both CPS samples were conjugated to Cross reacting material ($CRM_{197}$), a nontoxic variant of diphtheria toxin, containing a single amino acid substitution (G52E), widely used carrier protein for conjugate vaccines against encapsulated bacteria (MW=58.48 kDa)[251, 252].

Starting from the first sample, the entire CPS was previously activated using 1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP), an organic cyanylating reagent (Figure 6.25).[253]

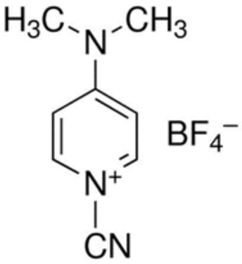


**Figure 6.25:** Chemical structure of CDAP.

The random binding of the cyanide groups of CDAP to the hydroxyl groups of CPS allowed for the subsequent binding of the activated polysaccharide to the amine groups of the Lys present in the $CRM_{197}$ protein (Figure 6.26).[254]

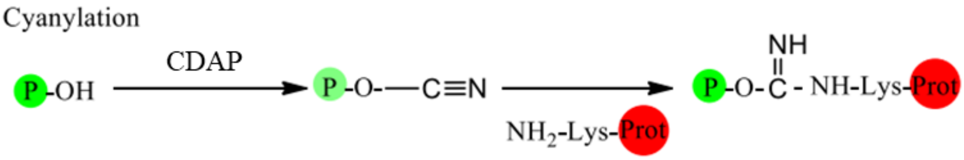


**Figure 6.26:** Chemistry used for the conjugation of CPS (P=polysaccharide) to the $CRM_{197}$ (Prot), through CDAP activation (extracted from ref.[254])

The successful conjugation of the entire CPS sample activated with CDAP to $CRM_{197}$ could be observed in the gel below, appearing as a peculiar ‘smear’. The final sample, reported in Figure 6.27, was obtained after two purification steps. The first in Amicon 100K in PBS 1X to remove not conjugated protein and the second using $(NH_4)_2SO_4$ to precipitate the CPS in excess. The obtained quantity was 1.8 mL of pure glycoconjugate at concentration of 0.1 mg/mL (Figure 6.28).

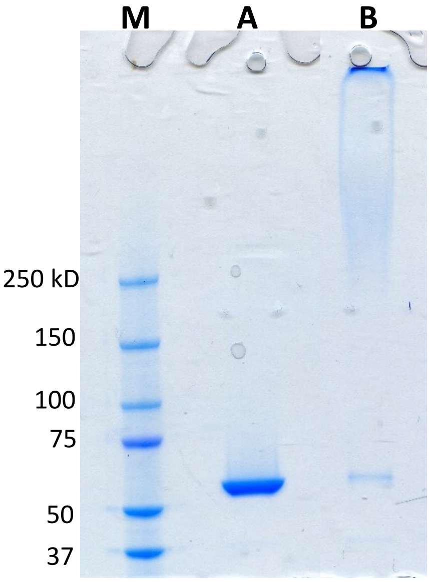
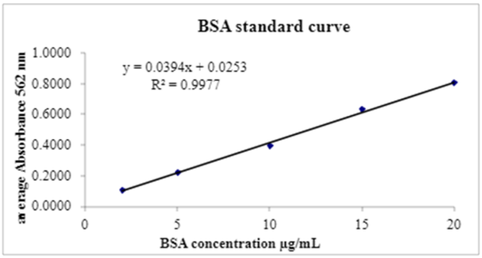


**Figure 6.27:** NuPAGE 3-8% with Coomassie staining, containing Precision Plus Protein Standards (M), 5 ug of CRM$_{197}$ (A) and 5 ug of the glycoconjugate of interest (B).

| µg/mL | Abs 562 nm | Abs 562 nm | Abs 562 nm average |
|---|---|---|---|
| 2 | 0.1430 | 0.0810 | **0.1120** |
| 5 | 0.2320 | 0.2120 | **0.2220** |
| 10 | 0.4210 | 0.3750 | **0.3980** |
| 15 | 0.5960 | 0.6690 | **0.6325** |
| 20 | 0.7890 | 0.8350 | **0.8120** |



| Sample | Abs 562 nm | | Abs 562 nm average | difference % | Protein µg/mL | Dilution | Total protein mg/mL | Volume mL | Quantity mg |
|---|---|---|---|---|---|---|---|---|---|
| **Conjugate_CPS UC+CRM_FinalSample** | 0.2180 | 0.2240 | 0.2210 | 2.7 | 4.977 | 20.0 | **0.100** | **1.80** | **0.179** |

**Figure 6.28:** microBCA assay performed on the glycoconjugate of CPS and CRM$_{197}$, using BSA standard curve.

The second glycoconjugate was produced by following the already described procedure, upon the hydrolysis in an aqueous solution of AcOH 1% for 2h at 100°C in order to remove the Leg residue. The hydrolysed CPS was firstly activated with

CDAP and then, conjugated to $CRM_{197}$. Also in this case, a precipitation in $(NH_4)_2SO_4$ and a purification by Amicon 30K in PBS 1X was performed (Figure 6.29). Notably, due to the lower MW of the hydrolysed sample, the conjugation was facilitated, allowing to obtain 1.8 mL of the glycoconjugate at concentration of 0.2 mg/mL (Figure 6.30).

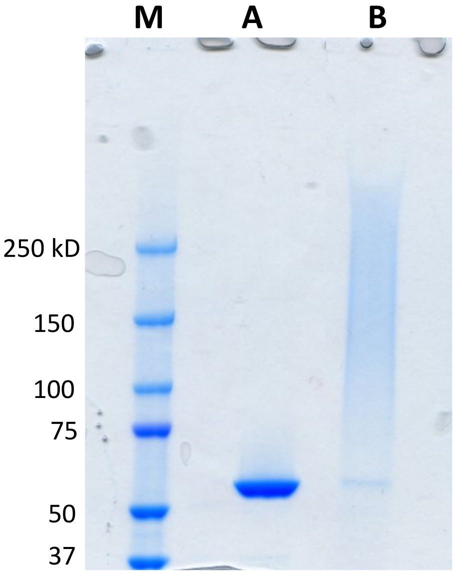
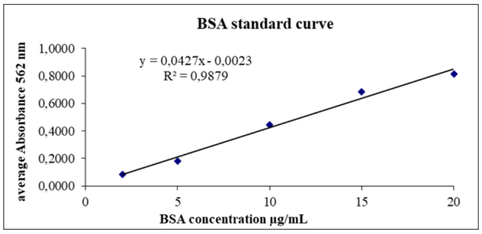


**Figure 6.29:**NuPAGE 3-8% with Coomassie staining containing Precision Plus Protein Standards (M), 5 ug of $CRM_{197}$ (A) and 5 ug of the glycoconjugate of interest (B).

| µg/mL | Abs 562 nm | Abs 562 nm | Abs 562 nm average |
|---|---|---|---|
| 2 | 0.0880 | 0.0760 | **0.0820** |
| 5 | 0.1900 | 0.1770 | **0.1835** |
| 10 | 0.4440 | 0.4440 | **0.4440** |
| 15 | 0.6970 | 0.6710 | **0.6840** |
| 20 | 0.8300 | 0.7960 | **0.8130** |



| Sample | Abs 562 nm | | Abs 562 nm average | difference % | Protein µg/mL | Dilution | Total protein mg/mL | Volume mL | Quantity mg |
|---|---|---|---|---|---|---|---|---|---|
| **IIConjugate_Hydrolyzed CPS UC+CRM_FinalSample** | 0.4570 | 0.4910 | 0.4740 | 7.2 | 11.157 | 20.0 | **0.223** | **1.80** | **0.402** |

**Figure 6.30**: microBCA assay performed on the glycoconjugate of hydrolysed CPS and CRM197, using BSA standard curve.

### 6.5.2.1 *in vivo* evaluation of the immunogenic activity of the glycoconjugates

Once evaluated their MW (Figure 6.31), the two produced glycoconjugates were administered to mice to evaluate their immunogenic activity and, thus, their ability to stimulate the antibody response.

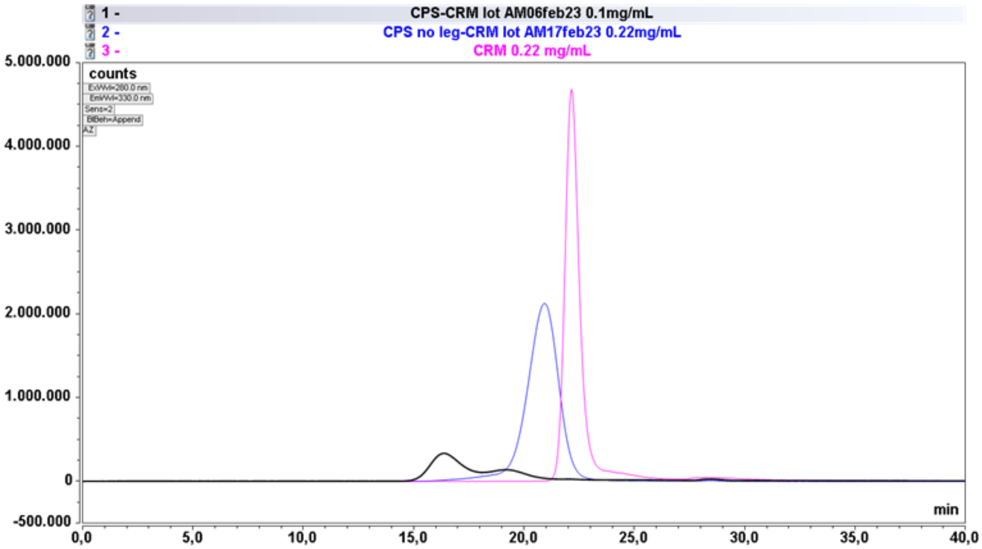


**Figure 6.31:** Chromatogram of SEC-HPLC analysis of produced glycoconjugates.

As expected, the elution of the samples on a TSK6000PWxl column occurred in order of decreasing MW: first, the glycoconjugate with the entire CPS (black line), second, the glycoconjugate with hydrolysed CPS (blue line), and finally, CRM197 alone (pink line). Upon obtaining the sera, ELISA was performed by coating the plates with the whole polysaccharide, using an anti-mouse IgG (whole molecule) as secondary antibody. Titers were calculated as the reciprocal of the dilution resulting in an OD of 1 (Figure 6.32).

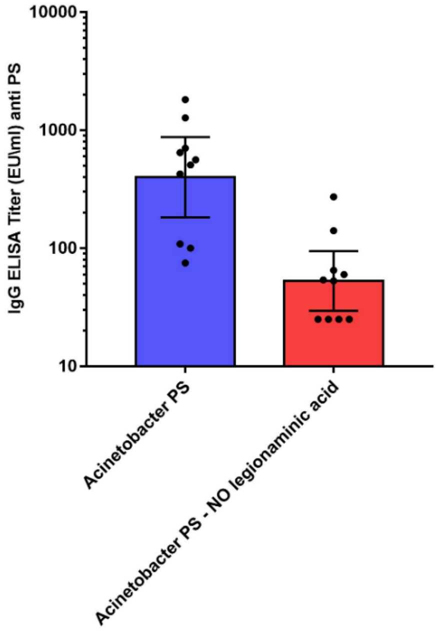


**Figure 6.32:** ELISA of the glycoconjugate containing the entire CPS (in blue) and the glycoconjugate containing hydrolysed CPS (in red).

The results indicated that the entire CPS was immunogenic, and the removal of the Leg residue significantly impacted the antibody response. This could be attributed to two possible explanations: (i) the containing Leg-epitope was the immunodominant epitope, and so its removal reduced the antibody response, (ii) the removal of Leg residue altered the 3D structure of the CPS, resulting in the loss of many conformational epitopes. Statistical analysis was also conducted between the two groups using the Mann-Whitney test, revealing a significant difference with a *p*-value of 0.0003.

## 6.6 Discussion

The extraction of glycoconjugates from an *A. baumannii* clinical isolate revealed the presence of a CPS on the surface of this Gram-negative bacterium. Mass spectrometry and NMR spectroscopy were conducted to elucidate its fine structure. Specifically, its repeating unit consists of a pentasaccharide containing (i) two Gal molecules and one GalNAc unit, all in β-configuration, (ii) a residue of α-GlcNAc and (iii) an α-Leg, not stoichiometrically substituted by a hydroxybutyryl group at position 7. Leg is a unique bacterial sugar, a nonulosonic acid structurally similar to Neu5Ac, a monosaccharide found in eukaryotic cells. Its study is essential because ESKAPE bacteria, including *A. baumannii*, have evolved to employ these sugars to mimic those of the host, evading the immune response by exploiting inhibitory proteins.

With the aim to achieve deep insights into the conformational features of *A. baumanni* CPS, once determined its chemical structure, computational studies were performed. To this end, a home-made protocol was applied to parametrize the Leg and to build the repeating unit of *A. baumannii* CPS. The conformational analysis of the CPS in the free state, showed a moderate flexibility of the ligand. Fluctuations were observed in the ω torsion angles (flipping between *tg* and *gt* conformations) around β-Gal-(1-6)-β-Gal and α-Leg5Ac7Ac-(2-6)-β-Gal glycosidic linkages.

Furthermore, an evaluation of the immunogenic activity of the CPS was conducted. During my stay at GSK, the CPS was randomly conjugated to the $\text{CRM}_{197}$ carrier protein. In particular, two samples, one bearing the Leg residue and the hydrolyzed CPS, were used. Subsequent analysis of the sera from immunized mice revealed that the Leg-conjugated epitope was the immunodominant epitope, and its removal not only reduced the antibody response but also altered the 3D structure of the CPS, resulting in the loss of many conformational epitopes. These results could pave the way for the development of glycoconjugate-based vaccines against *A. baumannii* infections worldwide.

Further analysis will be performed to investigate the molecular recognition of *A. baumanni* CPS by the host immune system. Notably, preliminary ELISA results indicated the interaction of *A. baumannii* CPS with two different human lectins: Fc-DC-SIGN and Fc-Siglec-7. Thus, NMR experiments and computational studies are currently in progress to enhance our understanding of how *A. baumannii* is recognized by the host immune system.

# SECTION III – EXPERIMENTAL SECTION

# Chapter 7 – Materials and methods

## 7.1 Bacterial cells growth and extractions

### 7.1.1 *E. coli* R1-type cells growth, extraction and purification (related to Chapter 3)

*E. coli* cells carrying R1 type core oligosaccharide were grown in 8 L of Luria Bertani (LB) medium at 37°C, under stirring (180 rpm), up to 0.9 $OD_{600nm}$. Cells were harvested by centrifugation at 15°C – 10000 g for 20 min and then, washed with two rounds of water, one round of ethanol, one round of acetone and one last round of ethyl ether, to remove contaminants from cell growth.

Freeze-dried cells were extracted using petroleum ether-chloroform-phenol (PCP) extraction (2:5:8, v/v/v),[129, 255] to isolate the LOS fraction, analysed through 12% SDS-PAGE, using silver nitrate staining.[256] The obtained LOS was then de-lipidated to remove ester- and amide-linked fatty acids. Firstly, it was treated with anhydrous hydrazine (1 ml for 20 mg of sample), stirred at 37°C for 90 min, cooled in an ice bath and some drops of romil acetone were poured to allow the precipitation. The precipitate was then centrifuged, washed twice with romil acetone and dried.[257] Subsequently, the O-deacylated sample was N-deacylated with a strong alkaline treatment (KOH 4 M, 120 °C). After desalting using a column of Sephadex G-10, the resulting sample was the core oligosaccharide fraction, namely OSR1.

### 7.1.2 *B. vulgatus* strain mpk extraction and purification (related to Chapter 4)

*B. vulgatus* strain mpk LPS was extracted, purified, and characterized in our group, as already described.[217]

### 7.1.3 *A. baumannii* CPS extraction and purification (related to Chapter 6)

Dried cells from an *A. baumannii* clinical isolate were provided by Prof. E. De Gregorio (Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II). Pellets were washed in water at 4°C in three consecutive centrifugation cycles of 3h, 1h and 45 min, to remove eventual contaminants derived from cell growth, then, PCP and hot phenol-water (1:1, v/v) extractions were performed in order to isolate CPS.[129] The extracted phases were subjected to enzymatic hydrolysis with DNAse, RNAse and Proteinase K and dialyzed against water to remove proteins, nucleic acids and phenol traces.[129] A 12% SDS-PAGE using silver nitrate staining, was run to control the sample and CPS was detected in the water phase. As still LOS contaminant was found, the CPS-containing water phase was subjected to an ultracentrifugation step at 45000 g, for 24h, at 4°C and then, another 12% SDS-PAGE was run to control the obtained supernatant.

In order to perform the structural characterization by NMR, a mild acid hydrolysis was performed with acetic acid 1% (100°C, 3 h) in order to separate the saccharidic portion from the lipidic ones. Then, a purification step with a biogel P6 SEC eluted in water, was done to further purify the CPS sample.

Moreover, to understand how Leg residue was linked to the CPS backbone, taking the advantage of its lability in the linkage, a chemical hydrolysis of the CPS was performed with acetic acid 6% (100°C, 2 h). The obtained supernatant was purified on a biogel P10 SEC eluted in water, obtaining two fractions: the CPS backbone and the Leg residue free in solution.

## 7.2 Chemical analysis

Monosaccharides' determination was performed through GC-MS analysis as described.[204] Sugar residues were identified as acetylated *O*-methyl glycosides (AMG) derivatives. After methanolysis (1.25M HCl/MeOH, 85°C, 16 h) and acetylation with acetic anhydride in pyridine (85°C, 30 min), samples were analyzed by GC-MS. Linkage analysis (PMAA) was carried out by methylation of complete saccharidic regions: the samples were methylated with $CH_3I$, hydrolyzed with 2 M TFA (100°C, 2 h), carbonyl-reduced with $NaBD_4$, acetylated with acetic anhydride and pyridine, and analyzed by GC-MS. All sugar derivatives were analyzed on an Agilent Technologies gas chromatograph 6850A equipped with a mass selective detector 5973N and a Zebron ZB-5 capillary column (Phenomenex, 30 m x 0.25 mmi.d., flow rate 1 $cm^3$/min, He as carrier gas).

## 7.3 Production and purification of the recombinant protein DC SIGN (related to Chapters 3 and 4)

The expression and purification of the C-type lectin DC-SIGN was carried out during my secondment in the lab of Prof. F. Fieschi (Institut de Biologie Structurale, Grenoble, France). The cDNA encoding DC-SIGN ECD monomer (amino acids 66–404) was cloned into a pET30b plasmid (Novagen) resulting in expression vector pET30b-DC-SIGN ECD, with subsequent transformation of *E. coli* BL21(DE3) competent cells,[258, 259] following the protocol suggested by the provider. First, the plasmid (10 µg) was incubated with cells for 30 min in ice and then permeabilized through the cell membrane using a heat shock at 42°C for 45 sec. After the transformation, 500 µL of LB medium (20 g/L) was added and cells were grown for 1h at 37°C, 180 rpm. Finally, 200 µL of cells were seeded in an Agar LB plate using Kanamycin as antibiotic (50 mg/L) and incubated at 37°C overnight. Antibiotic allowed the selection of the cells with the plasmid integrated. A single bacterial colony selected from the Agar plate was grown overnight in 250 mL of LB media, supplemented with Kanamycin, at 37°C, 180 rpm. Next day, 50 mL of preculture per Liter of culture with 1 mL of Kanamycin were inoculated, at 37°C, 180 rpm for 3h. When the $OD_{600}$ (optical density at 600 nm) was 0.6, the culture was induced with 1 mM of Isopropyl β- d-1-thiogalactopyranoside (IPTG), and the grown was prolonged for other 3h at 37°C,180 rpm. When the final $OD_{600}$ was 3-4, cells were harvested by centrifugation at 5000 RCF for 20 min, at 4°C. The pellet was collected and resuspended in buffer 150 mM NaCl, 25 mM TRIS pH 8, 4 mM $CaCl_2$ with Complete-EDTA-free protease inhibitor (Roche). Cells were lysed by sonication for 2 cycles of 12' (2 sec on, 10 sec off, Amp= 90%), and then, centrifugated at 4°C, 100000 g, for 30 min. After centrifugation, the supernatant was removed and pellet was washed by cycles of resuspension and ultracentrifugation, in three different buffer, to separate as much as possible soluble proteins from aggregated recombinant DC-SIGN: (i) 150 mM NaCl, 25 mM TRIS pH 8, 4 mM $CaCl_{2,}$ to remove soluble proteins, (ii) 150 mM NaCl, 25 mM TRIS pH 8, 4 mM $CaCl_{2,}$ to remove membrane proteins and (iii) 6 M Gdn/HCl, 150 mM NaCl, 25 mM TRIS pH 8, 4 mM $CaCl_2$ + 0.01% β-mercaptoethanol was used to solubilized the inclusion bodies containing DC-SIGN. DC-SIGN was then subjected to a refolding procedure, allowing its tetramerization, achieved by gradual removal of the denaturing agent (Gdn/HCl), through drop-by-drop dilution of solubilized inclusion bodies in buffer 1.25 M NaCl, 210 mM TRIS pH 8, 25 mM $CaCl_{2.}$ The sample was then dialyzed overnight against water. After a centrifugation at 4°C, 100000 g for 30 min, the supernatant was filtered through a 0.45 micron sterile filter membrane, ready for the purification steps.

Filtered DC-SIGN was loaded on a mannan-agarose affinity column (GE Life Sciences), previously equilibrated with buffer 150 mM NaCl, 25 mM Tris pH 8, 4 mM $CaCl_{2.}$ The protein was eluted in the same buffer without $CaCl_2$ but supplemented with 1 mM EDTA. During this step, it was possible to control the correct tetramerization of DC-SIGN, because the affinity for the Man is higher for DC-SIGN tetramer than monomer. Then, the eluted fraction was directly injected in a gel filtration Superose 12 using buffer 150 mM NaCl, 25 mM Tris pH 8, 4 mM $CaCl_{2.}$ The protein was collected and concentrated in vivaspin concentrators with a cut-off of 30 kDa, up to a minimum volume of around 2 mL, at a concentration of 10 mg/mL, measured by NANOdrop. 12% SDS-PAGE with Coomassie staining were repeatedly run during all the production steps to control the quality of the sample. Additionally, the identity of the final protein was assessed by MS, in Institut de Biologie Structurale.

## 7.4 NMR methods

NMR experiments were recorded on a Bruker AVANCE NEO 600-MHz equipped with a cryo probe. Data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in 50 mM deuterate phosphate buffer (NaCl 140 mM, $Na_2HPO_4$ 10 mM, KCl 3 mM, pH 7.4 and 2,2,3,3-d(4)-3-(trimethylsilyl)propionic acid, sodium salt (TSP, 10 uM) was used as internal reference to calibrate all the spectra. The chemical shifts of all the glycan compounds used in this thesis, were assigned by $^1$H, COSY, TOCSY, NOESY, ROESY and HSQC experiments.[131]

### 7.4.1 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 (7.5 ppm and 0 ppm). STD experiments were performed by using the sequence "stddiffesgp" with an excitation sculpting with gradient pulses (esgp) 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 epitope mapping of the ligands was obtained by the calculation of the ratio of the STD signals ($I_0 – I_{sat}$) and the 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.

Related to Chapter 5, all the NMR spectra were acquired in the same conditions for P110 and P40/P90, using protein concentrations of 20 µM ligand concentrations of 1mM (protein:ligand ratio of 1:50) and temperature of 298K.

### 7.4.2 Tr-NOESY/ROESY

Protein/ligand molar ratios varied from 1:10 to 1:20. Homonuclear 2D $^1$H-$^1$H NOESY/ROESY experiments were conducted 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 $^1$H-$^1$H distances were calculated using Equation 2.4 (cf. §2.4.1.1).

Regarding the construction of NOE build up curves (cf. Chapter II), 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 exponential Equation 2.2 (cf. §2.4.1.1).

## 7.5 Biophysical techniques

### 7.5.1 Fluorescence spectroscopy (related to Chapter 5)

Steady-state fluorescence experiments were performed on a Fluoromax-4 spectrofluorometer from Horiba Scientific (Edison, USA). All the measurements were acquired at a controlled temperature of 25°C, the excitation wavelength was selected at 280 nm and the emission spectra collected between 290 and 500 nm. The slits were set to 5 nm for both the excitation and emission monochromators. All the spectra were recorded after an equilibration time of 3 minutes in a 1 cm path-length quartz cuvette, under constant stirring.

Related to Chapter 5, a fixed concentration of both P110 and P40/P90 proteins was selected at 0.07 μM in 1,2 mL PBS buffer (pH 7.4) and titrated by adding small amounts of each ligand 3’-SLn, 6’-SLn, sTa-Thr in the range from 0 to 6.5 μM (1–100 μL of a ligand stock solution of 84 μM). In these conditions, it was possible to observe the quenching of the proteins' fluorescence in presence of the three different ligands; in particular, no ligands emission interference was observed.

Data analysis was performed using the software Origin 8.1. Specifically, the binding curves were obtained by plotting $F/F_0$ values versus ligand concentrations, where F and $F_0$ are fluorescence intensities in presence and in absence of the ligands, respectively. The binding constants ($K_b$) were determined by *non*-linear regression with One Site-Specific Binding model, as previously described (cf. §2.4.2).[115, 149, 260]

### 7.5.2 Surface Plasmon Resonance (related to Chapters 3 and 4)

All the experiments were performed on a Biacore T200, in the lab of Prof. F. Fieschi (Institut de Biologie Structurale, Grenoble, France), using functionalized CM3 sensor chips.

Related to Chapter 3, competition experiments were performed using flow-cells functionalized with mannosylated-BSA. Flow cells were activated as previously described.[261] Flow cell 1 was functionalized with final density of 2112 RU of BSA and used as a control surface, flow cell 2 and 3 were functionalized with final density of 1960 RU and 2067 RU of BSA-Manα1–3[Manα1–6]Man (BSA-Mannotriose, Dextra), all blocked with ethanolamine. The BSA-Mannotriose used to functionalize sensor chip harbors an average coupling ratio of 13 α1-3, α1-6 Mannotriose per BSA with 14 atoms spacer. The affinity of OSR1 was then evaluated through determination of its $IC_{50}$ using a DC-SIGN ECD binding inhibition assay. The ECD of DC-SIGN was injected onto the BSA-Mannotriose surface, at 20 μM alone or in presence of an increasing concentration of OSR1, from 5 μM to 5 mM. Injections were performed at 20 μL/min using 25 mM Tris-HCl, pH 8, 150 mM NaCl, 4 mM $CaCl_2$, and 0.05% of P20 surfactant as running buffer. Analysis has been performed using the steady-state interaction model. The $R_{eq}$ steady state binding responses of DC-SIGN ECD to BSA-Mannotriose surface were obtained from sensorgrams of each condition and converted to percent relative residual activity values (y) with respect to DC-SIGN alone. This relative residual DC-SIGN ECD activity (y) was plotted against OSR1 concentration and fitted using four-parameter logistic model (Equation 7.1); where $R_{hi}$ and $R_{lo}$ are maximum and minimum asymptotes of activity, $A_1$ is the inflection point and $A_2$ is a slope of the curve. Finally, $IC_{50}$ was calculated using the values of fitted parameters ($R_{hi}$, $R_{lo}$, $A_1$ and $A_2$) and Equation 7.2.

$$\boldsymbol{y = R_{hi} - \frac{R_{hi}-R_{lo}}{1+\left(\frac{Conc}{A_1}\right)^{A_2}}}$$ **Equation 7.1**

$$\boldsymbol{IC_{50} = A_1\left(\frac{R_{hi}-R_{lo}}{R_{hi}-50}\right)^{\frac{1}{A_2}}}$$ **Equation 7.2**

Related to Chapter 4, flow cell 1 was functionalized with BSA, blocked with ethanolamine, and subsequently used as a control surface. Flow cells 2 and 3 were treated with BSA-Mannotriose (Dextra) (60 µg/ml) in 10 mM NaOAc pH 4 to reach different binding densities and blocked with ethanolamine. The final densities on flow cells 2 and 3 were 2690 and 1715 RU, respectively. The use of these two-flow cell will allow redundancy in the measurements. The affinity of the different oligomers of Repeating unit (from 2-Mer to 16-Mer) for DC-SIGN extra-cellular domain (ECD) were evaluated via an established inhibition assay[262] in which DC-SIGN ECD was injected at 20 µM alone or in the presence of increasing concentration of oligomers (ranging from 20 mM to 40 µM of the corresponding oligomer obtained by serial dilution by a factor of two). Injections were performed at 5 µL/min using 25 mM Tris-HCl pH 8, 150 mM NaCl, 3.8 mM $CaCl_2$, 0,05% P20 surfactant as running buffer. The surface was regenerated by the injection of 50 mM EDTA. DC-SIGN ECD equilibrium binding responses ($R_{eq}$) for each sample were obtained from the reference surface corrected sensorgrams, 95 s after the start of the injection. The obtained $R_{eq}$ values were converted to DC-SIGN residual activity values (y, %) with respect to $R_{eq}$ of DC-SIGN alone, which was assigned a 100 % activity value. After plotting residual activity against corresponding compound concentration, the 4-parameter logistic model (Equation 7.1**Errore. L'origine riferimento non è stata trovata.**) was fitted to the plots, and finally the $IC_{50}$ values were calculated using Equation 7.2 (cf. §2.4.3).

## 7.6 Computational studies

### 7.6.1 Docking

Docking calculations of all the systems were performed using AutoDock 4.2.2 and analyzed with AutoDock Tools.[167] Ligands were prepared for the docking calculations using AutoDockTools, La Jolla, CA, setting all rotatable bonds free to move during the calculations except for the glycosidic bonds.

Different 3D grids were used for the systems, selecting them taking into account the volume necessary to cover the ligand and the key amino acids of protein's binding site. The grid point spacing thus centered to the ligand. 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 the Lamarckian Genetic algorithm were performed, with a population size of 100, and 250,000 energy evaluations. After docking, the 200 poses were clustered in groups with a root mean square deviation less than 2.0 Å. The clusters were ranked according to the lowest energy representative of each cluster.

### 7.6.2 Molecular Mechanics calculations

Molecular mechanics studies were carried out using Maestro software.[222] The adiabatic maps were built for each disaccharide connected by a glycosidic linkage, defined by the torsion angles $\phi$ (H1-C1-O-CX'), $\psi$ (C1-O-CX'-HX') and $\omega$ (O6-C6-C5-O5). MM3* force field included in MacroModel405[223] at a dielectric constant of 80 was used for the calculations. For each disaccharide, $\phi$, $\psi$ and $\omega$ 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.

### 7.6.3 Parametrization of *non*-canonical sugars

The uncommon sugars were parametrized by Dr F. Nieto-Fabregat, through a protocol optimized in our group. The following *non*-canonical sugars Gal*f* and Kdo (related to Chapter 4) and Leg5Ac7Ac (related to Chapter 6) were built with Gaussian

09,[263] performing the Restrained ElectroStatic Potential (RESP) charges calculation with a Hartree-Fock calculation and a 6–31G* basis set. VFFDT[264, 265] Antechamber, San Francisco, CA[266] and xLeap were combined to generate the .prep and .frcmod files.

### 7.6.4 Molecular dynamics simulations

The ligands were built using the glycam website (https://www.glycam.org)[202] builder utility; the torsion angles were chosen by following the values from the molecular mechanics calculations. For the protein preparation, missing hydrogen atoms were added, and the protonation state of ionisable groups and cap termini were computed using Maestro Protein Preparation Wizard.[267]

Molecular dynamics calculations of 100 ns on ligands alone and bound to the proteins were performed using AMBER 18 software package[168] in explicit waters using AMBER ff14SB,[171] Glycam06j-1,[170] and TIP3P force fields for the protein residues, the saccharide ligand, and the water solvent molecules, respectively. A glycam adapted force field was used for the different *non*-parametrized residues. The complexes were minimized using Sander tools. All systems were neutralized by adding Na+ ions and then hydrated with an octahedral box of TIP3P water of the proper size and the remote interactions were calculated using a cut-off of 10 or 15 Å, applying *tleap* module of the AMBER package. The MD calculations were performed by using the PMEMD.CUDA implementation within AMBER 18 package. At this point, an energy minimization process was performed to refine the initial structures. The calculations employed SHAKE for the C-H bonds and 1 fs of integration step. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. At first the systems were minimized holding the complexes, while a further minimization step was performed on the entire systems. Furthermore, the whole systems were slowly heated from 0 to 300 K using a weak restrain on the solute and then, equilibrated at 300 K using constant pressure and removing the restrains on the solute. Related to

Chapter 5, for the complexes of P110 and P40/P90 with sTa-Thr, the value of the peptide dihedral angle (O-CB-CA-N) was restrained approximately at 60 degrees. Related to Chapter 4, in the specific case of the tRha-LPS:DC-SIGN complex, a restrain was applied to the Rha-Glc torsion in order to avoid the population of the *non-exo* conformation.

Coordinates were obtained in order to acquire 10000 structures representing the progression of the dynamics. The trajectories were assessed using the ptraj module in AMBER 18 and the MD results were visualized using the VMD program.[268] A cluster analysis of the MD trajectory was applied based on ligand RMSD, employing the K-means algorithm integrated into the *ptraj* module. The representative structures of the most populated clusters were selected to depict the complexes interactions. Multiple representative poses were chosen for each MD simulation. The identification of hydrogen bonds was computed using the CPPTAJ module within AMBER 18, defining a hydrogen bond as forming between an acceptor heavy atom (A), a donor hydrogen atom (H), and a donor heavy atom (D). The distance cut-off was chosen at 3 Å, and the A-H-D angle cut-off was set to 135°. The frequency of protein-ligand bonds formed during the dynamics was reported with a 5 Å cut-off. Dihedral conformation analysis has been performed using a home-made script to depict the variation of the torsion during the MD simulation and obtain a histogram of the most populated values.

## 7.7 CORCEMA-ST (related to Chapter 4)

CORCEMA-ST protocol was employed, as previously described.[184] The coordinates for the complexes were chosen from the MD trajectory analyses. It was assumed, as observed by NMR, that the ligand's conformation remained constant between the free and bound states. The input parameters, such as ligand and protein concentrations, were determined experimentally. A saturation time of 2 seconds was set, and the $K_D$ were initially determined based on experimental data for the DC-SIGN/4-Mer and DC-SIGN/tRha-LPS complex, with further adjustments to achieve the best fit. A binding site $\text{cut}_{\text{off}}$ of 6 Å was applied. Fractional intensity changes for each ligand proton were calculated by computing the R matrix and spectral densities. These changes were then compared to the experimental STD effects using a NOE R factor, a normalized root-mean-square deviation value. In these calculations, only the STD values of the ligand's isolated signals were considered.

# Bibliography

1. Seyrek, E.; Decher, G., 7.09 - Layer-by-Layer Assembly of Multifunctional Hybrid Materials and Nanoscale Devices. In *Polymer Science: A Comprehensive Reference*, Matyjaszewski, K.; Möller, M., Eds. Elsevier: Amsterdam, **2012**; pp 159-185.
2. Thursby, E.; Juge, N., Introduction to the human gut microbiota. *The Biochemical journal* **2017,** *474* (11), 1823-1836.
3. Santajit, S.; Indrawattana, N., Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed research international* **2016,** *2016*, 2475067.
4. Doron, S.; Gorbach, S. L., Bacterial Infections: Overview. In *International Encyclopedia of Public Health*, **2008**; pp 273-282.
5. Ratner, D. M.; Adams, E. W.; Disney, M. D.; Seeberger, P. H., Tools for glycomics: mapping interactions of carbohydrates in biological systems. *Chembiochem : a European journal of chemical biology* **2004,** *5* (10), 1375-83.
6. Reitsma, S.; Slaaf, D. W.; Vink, H.; van Zandvoort, M. A.; oude Egbrink, M. G., The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Archiv : European journal of physiology* **2007,** *454* (3), 345-59.
7. Gagneux, P.; Hennet, T.; Varki, A., Biological Functions of Glycans. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor Laboratory Press Copyright © 2022 The Consortium of Glycobiology Editors, La Jolla, California; published by Cold Spring Harbor Laboratory Press; doi:10.1101/glycobiology.4e.7. All rights reserved.: Cold Spring Harbor (NY), **2022**; pp 79-92.
8. Mogensen, T. H., Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **2009,** *22* (2), 240-73, Table of Contents.
9. Khan, H.; Shi, B.; Tian, Y.; Wang, T.; Hussain, S.; Khan, F.; Khan, Z.; Ashfaq, B.; Ahmad, H.; Ahmad, T., Glycan regulation in cancer, nervous and immune system: A narrative review. *Biomedical Research and Therapy* **2019,** *6*, 3113-3120.
10. Nicolaou, K. C.; Mitchell, H. J., Adventures in Carbohydrate Chemistry: New Synthetic Technologies, Chemical Synthesis, Molecular Design, and Chemical Biology. *Angewandte Chemie (International ed. in English)* **2001,** *40* (9), 1576-1624.
11. Hart, G. W., Thematic minireview series on glycobiology and extracellular matrices: glycan functions pervade biology at all levels. *J Biol Chem* **2013,** *288* (10), 6903.
12. Fernández-Tejada, A.; Cañada, F. J.; Jiménez-Barbero, J., Glycans in Medicinal Chemistry: An Underexploited Resource. *ChemMedChem* **2015,** *10* (8), 1291-5.
13. Valverde, P.; Ardá, A.; Reichardt, N. C.; Jiménez-Barbero, J.; Gimeno, A., Glycans in drug discovery. *MedChemComm* **2019,** *10* (10), 1678-1691.

14. Yu, R. K.; Tsai, Y. T.; Ariga, T.; Yanagisawa, M., Structures, biosynthesis, and functions of gangliosides--an overview. *Journal of oleo science* **2011,** *60* (10), 537-44.
15. Komath, S. S.; Fujita, M.; Hart, G. W.; Ferguson, M. A. J.; Kinoshita, T., Glycosylphosphatidylinositol Anchors. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor Laboratory Press Copyright © 2022 The Consortium of Glycobiology Editors, La Jolla, California; published by Cold Spring Harbor Laboratory Press; doi:10.1101/glycobiology.4e.12. All rights reserved.: Cold Spring Harbor (NY), **2022**; pp 141-54.
16. Fisher, P.; Thomas-Oates, J.; Wood, A. J.; Ungar, D., The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. *Front. Cell Dev. Biol.* **2019,** *7.*
17. Stanley, P.; Moremen, K.; Lewis, N.; Taniguchi, N.; Aebi, M., N-Glycans. In *Essentials of Glycobiology*, 4th ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; al., e., Eds. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2022.
18. Talabnin, K.; Talabnin, C.; Ishihara, M.; Azadi, P., Increased expression of the high-mannose M6N2 and NeuAc3H3N3M3N2F tri-antennary N-glycans in cholangiocarcinoma. *Oncology letters* **2018,** *15* (1), 1030-1036.
19. Nakano, M.; Mishra, S. K.; Tokoro, Y.; Sato, K.; Nakajima, K.; Yamaguchi, Y.; Taniguchi, N.; Kizuka, Y., Bisecting GlcNAc Is a General Suppressor of Terminal Modification of N-glycan. *Molecular & cellular proteomics : MCP* **2019,** *18* (10), 2044-2057.
20. Ye, Z.; Marth, J. D., N-glycan branching requirement in neuronal and postnatal viability. *Glycobiology* **2004,** *14* (6), 547-58.
21. Brockhausen, I.; Wandall, H.; Hagen, K.; Stanley, P., O-GalNAc Glycans. In *Essentials of Glycobiology*, 4th ed.; Varki, A.; Cummings, R.; Esko, J.; al., e., Eds. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, **2022**.
22. Di Carluccio, C.; Forgione, R. E.; Molinaro, A.; Crocker, P.; Marchetti, R.; Silipo, A., Exploring the fascinating world of sialoglycans in the interplay with Siglecs. In *Carbohydrate Chemistry*, **2020**; pp 31-55.
23. Shivatare, S. S.; Shivatare, V. S.; Wong, C.-H., Glycoconjugates: Synthesis, Functional Studies, and Therapeutic Developments. *Chemical Reviews* **2022,** *122* (20), 15603-15671.
24. Cavalcante, T.; Medeiros, M. M.; Mule, S. N.; Palmisano, G.; Stolf, B. S., The Role of Sialic Acids in the Establishment of Infections by Pathogens, With Special Focus on Leishmania. *Front. Cell. Infect. Microbiol.* **2021,** *11.*
25. Varki, A.; Schnaar, R. L.; Schauer, R., Sialic Acids and Other Nonulosonic Acids. In *Essentials of Glycobiology*, 3rd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor Laboratory Press Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved.: Cold Spring Harbor (NY), **2015**; pp 179-95.

26. Ghosh, S., Sialic acid and biology of life: An introduction. *Sialic Acids Sialoglycoconjugates in the Biology of Life, Health Disease* **2020**, 1 - 61.

27. Yehuda, S.; Padler-Karavani, V., Glycosylated Biotherapeutics: Immunological Effects of N-Glycolylneuraminic Acid. *Front Immunol* **2020,** *11*, 21.

28. Samraj, A. N.; Pearce, O. M.; Läubli, H.; Crittenden, A. N.; Bergfeld, A. K.; Banda, K.; Gregg, C. J.; Bingman, A. E.; Secrest, P.; Diaz, S. L.; Varki, N. M.; Varki, A., A red meat-derived glycan promotes inflammation and cancer progression. *Proc Natl Acad Sci U S A* **2015,** *112* (2), 542-7.

29. Varki, A., Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **2007,** *446* (7139), 1023-9.

30. Angata, T.; Varki, A., Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem Rev* **2002,** *102* (2), 439-69.

31. Severi, E.; Hood, D. W.; Thomas, G. H., Sialic acid utilization by bacterial pathogens. *Microbiology (Reading, England)* **2007,** *153* (Pt 9), 2817-2822.

32. Coico, R., Gram staining. *Current protocols in microbiology* **2006,** (1), A. 3C. 1-A. 3C. 2.

33. Schleifer, K. H., Classification of Bacteria and Archaea: Past, present and future. *Systematic and Applied Microbiology* **2009,** *32* (8), 533-542.

34. Messner, P.; Schäffer, C.; Kosma, P., Bacterial cell-envelope glycoconjugates. *Advances in carbohydrate chemistry and biochemistry* **2013,** *69*, 209-72.

35. Schäffer, C.; Messner, P., The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. *Microbiology (Reading, England)* **2005,** *151* (Pt 3), 643-651.

36. Schleifer, K. H.; Kandler, O., Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews* **1972,** *36* (4), 407-477.

37. Whitfield, C.; Szymanski, C. M.; Lewis, A. L.; Aebi, M., Eubacteria. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor Laboratory Press Copyright © 2022 The Consortium of Glycobiology Editors, La Jolla, California; published by Cold Spring Harbor Laboratory Press; doi:10.1101/glycobiology.4e.21. All rights reserved.: Cold Spring Harbor (NY), **2022**; pp 279-96.

38. Swoboda, J. G.; Campbell, J.; Meredith, T. C.; Walker, S., Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem : a European journal of chemical biology* **2010,** *11* (1), 35-45.

39. van Dalen, R.; Peschel, A.; van Sorge, N. M., Wall Teichoic Acid in Staphylococcus aureus Host Interaction. *Trends in microbiology* **2020,** *28* (12), 985-998.

40. Neuhaus, F. C.; Baddiley, J., A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiology and molecular biology reviews : MMBR* **2003,** *67* (4), 686-723.

41. Weidenmaier, C.; McLoughlin, R. M.; Lee, J. C., The zwitterionic cell wall teichoic acid of Staphylococcus aureus provokes skin abscesses in mice by a novel CD4+ T-cell-dependent mechanism. *PLoS One* **2010,** *5* (10), e13227.

42. Poli, A.; Anzelmo, G.; Nicolaus, B., Bacterial exopolysaccharides from extreme marine habitats: production, characterization and biological activities. *Marine drugs* **2010,** *8* (6), 1779-802.

43. Willis, L. M.; Whitfield, C., Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydrate research* **2013,** *378*, 35-44.

44. Cress, B. F.; Englaender, J. A.; He, W.; Kasper, D.; Linhardt, R. J.; Koffas, M. A., Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS microbiology reviews* **2014,** *38* (4), 660-97.

45. Kröncke, K. D.; Golecki, J. R.; Jann, K., Further electron microscopic studies on the expression of Escherichia coli group II capsules. *Journal of bacteriology* **1990,** *172* (6), 3469-3472.

46. Giguère, D., Surface polysaccharides from Acinetobacter baumannii: Structures and syntheses. *Carbohydrate research* **2015,** *418*, 29-43.

47. Avery, O. T.; Heidelberger, M., Immunological relationships of cell constituents of *Pneumococcus*. *The Journal of experimental medicine* **1923,** *38* (1), 81-5.

48. Berti, F.; Adamo, R., Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification. *Chemical Society Reviews* **2018,** *47* (24), 9015-9025.

49. Thanawastien, A.; Cartee, R. T.; Griffin IV, T. J.; Killeen, K. P.; Mekalanos, J. J., Conjugate-like immunogens produced as protein capsular matrix vaccines. *Proceedings of the National Academy of Sciences* **2015,** *112* (10), E1143-E1151.

50. Jones, C. H.; Zhang, G.; Nayerhoda, R.; Beitelshees, M.; Hill, A.; Rostami, P.; Li, Y.; Davidson, B. A.; Knight III, P.; Pfeifer, B. A., Comprehensive vaccine design for commensal disease progression. *Science Advances* **2017,** *3* (10), e1701797.

51. Micoli, F.; Costantino, P.; Adamo, R., Potential targets for next generation antimicrobial glycoconjugate vaccines. *FEMS microbiology reviews* **2018,** *42* (3), 388-423.

52. Di Lorenzo, F.; Duda, K. A.; Lanzetta, R.; Silipo, A.; De Castro, C.; Molinaro, A., A Journey from Structure to Function of Bacterial Lipopolysaccharides. *Chemical Reviews* **2022,** *122* (20), 15767-15821.

53. Raetz, C. R.; Whitfield, C., Lipopolysaccharide endotoxins. *Annual review of biochemistry* **2002,** *71*, 635-700.

54. Molinaro, A.; Holst, O.; Di Lorenzo, F.; Callaghan, M.; Nurisso, A.; D'Errico, G.; Zamyatina, A.; Peri, F.; Berisio, R.; Jerala, R.; Jiménez-Barbero, J.; Silipo, A.; Martín-Santamaría, S., Chemistry of lipid A: at the heart of innate immunity. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2015,** *21* (2), 500-19.

55. Silipo, A.; Molinaro, A., The diversity of the core oligosaccharide in lipopolysaccharides. *Sub-cellular biochemistry* **2010,** *53*, 69-99.

56. Lerouge, I.; Vanderleyden, J., O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS microbiology reviews* **2002,** *26* (1), 17-47.

57. Chaplin, D. D., Overview of the immune response. *The Journal of allergy and clinical immunology* **2010,** *125* (2 Suppl 2), S3-23.

58. Manna, P. R.; Gray, Z. C.; Reddy, P. H., Healthy Immunity on Preventive Medicine for Combating COVID-19. *Nutrients* **2022,** *14* (5).

59. Moser, M.; Leo, O., Key concepts in immunology. *Vaccine* **2010,** *28*, C2-C13.

60. Akira, S.; Uematsu, S.; Takeuchi, O., Pathogen Recognition and Innate Immunity. *Cell* **2006,** *124* (4), 783-801.

61. O'Neill, L. A. J.; Golenbock, D.; Bowie, A. G., The history of Toll-like receptors — redefining innate immunity. *Nature Reviews Immunology* **2013,** *13* (6), 453-460.

62. Shi, Y.; Evans, J. E.; Rock, K. L., Molecular identification of a danger signal that alerts the immune system to dying cells. *Nature* **2003,** *425* (6957), 516-21.

63. Martinon, F.; Burns, K.; Tschopp, J., The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. *Molecular cell* **2002,** *10* (2), 417-26.

64. Stuart, L. M.; Ezekowitz, R. A., Phagocytosis: elegant complexity. *Immunity* **2005,** *22* (5), 539-50.

65. Luster, A. D.; Alon, R.; von Andrian, U. H., Immune cell migration in inflammation: present and future therapeutic targets. *Nature Immunology* **2005,** *6* (12), 1182-1190.

66. Bachmann, M. F.; Kopf, M.; Marsland, B. J., Chemokines: more than just road signs. *Nature reviews. Immunology* **2006,** *6* (2), 159-64.

67. Li, M.; Carpio, D. F.; Zheng, Y.; Bruzzo, P.; Singh, V.; Ouaaz, F.; Medzhitov, R. M.; Beg, A. A., An essential role of the NF-kappa B/Toll-like receptor pathway in induction of inflammatory and tissue-repair gene expression by necrotic cells. *Journal of immunology (Baltimore, Md. : 1950)* **2001,** *166* (12), 7128-35.

68. Bonilla, F. A.; Oettgen, H. C., Adaptive immunity. *The Journal of allergy and clinical immunology* **2010,** *125* (2 Suppl 2), S33-40.

69. Murphy, K.; Weaver, C., *Janeway's immunobiology*. Garland science: 2016.

70. Boyd, W. C.; Shapleigh, E., Diagnosis of subgroups of blood groups A and AB by use of plant agglutinins (lectins). *The Journal of Laboratory Clinical Medicine* **1954,** *44* (2), 235-237.

71. Lepenies, B.; Lang, R., Lectins and their ligands in shaping immune responses. *Frontiers in immunology* **2019,** *10*, 2379.

72. Rabinovich, G. A.; van Kooyk, Y.; Cobb, B. A., Glycobiology of immune responses. *Annals of the New York Academy of Sciences* **2012,** *1253*, 1-15.

73. Raposo, C. D.; Canelas, A. B.; Barros, M. T., Human Lectins, Their Carbohydrate Affinities and Where to Find Them. *Biomolecules* **2021,** *11* (2).

74. Gabius, H. J., Animal lectins. *European journal of biochemistry* **1997,** *243* (3), 543-76.

75. Zaheer ul, H.; Dalal, P.; Aronson, N. N.; Madura, J. D., Family 18 chitolectins: Comparison of MGP40 and HUMGP39. *Biochemical and Biophysical Research Communications* **2007,** *359* (2), 221-226.

76. Vasta, G. R.; Amzel, L. M.; Bianchet, M. A.; Cammarata, M.; Feng, C.; Saito, K., F-type lectins: a highly diversified family of fucose-binding proteins with a unique sequence motif and structural fold, involved in self/non-self-recognition. *Frontiers in immunology* **2017,** *8*, 1648.

77. Yoshida, Y., F-box proteins that contain sugar-binding domains. *Bioscience, biotechnology,biochemistry* **2007,** *71* (11), 2623-2631.

78. Matsushita, M., Ficolins: complement-activating lectins involved in innate immunity. *Journal of innate immunity* **2010,** *2* (1), 24-32.

79. Cummings, R. D. E., M. E.; Ramya, T. N. C.; Kato, K.; Rabinovich, G. A.; Surolia, A.,, L-type lectins. In *Essentials of Glycobiology*, 4th ed.; Cold Spring Harbor: New York: **2022**; pp 768-786.

80. Bauters, L.; Naalden, D.; Gheysen, G., The Distribution of Lectins across the Phylum Nematoda: A Genome-Wide Search. *International journal of molecular sciences* **2017,** *18* (1).

81. Cummings, R.; Schnaar, R.; Ozeki, Y., R-type lectins. In *Essentials of Glycobiology*, 4th ed.; Cold Spring Harbor: New York: 2022; pp 748-767.

82. Chen, L.; Li, J.; Yang, G., A comparative review of intelectins. *Scandinavian journal of immunology* **2020,** *92* (1), e12882.

83. Müller, W. E., *Molecular evolution: evidence for monophyly of Metazoa*. Springer Science & Business Media: 2012; Vol. 19.

84. Wang, B.; Boons, G.-J., *Carbohydrate recognition: biological problems, methods, and applications*. John Wiley & Sons: 2011.

85. Santos, A. F.; Da Silva, M.; Napoleão, T.; Paiva, P.; Correia, M. d. S.; Coelho, L., Lectins: Function, structure, biological properties and potential applications. *Current Topics in Peptide & Protein Research* **2014**.

86. Houser, J.; Kozmon, S.; Mishra, D.; Hammerová, Z.; Wimmerová, M.; Koča, J., The CH-π Interaction in Protein-Carbohydrate Binding: Bioinformatics and In Vitro Quantification. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2020,** *26* (47), 10769-10780.

87. Zelensky, A. N.; Gready, J. E., The C-type lectin-like domain superfamily. *The FEBS journal* **2005,** *272* (24), 6179-6217.

88. Brown, G. D.; Willment, J. A.; Whitehead, L., C-type lectins in immunity and homeostasis. *Nature reviews. Immunology* **2018,** *18* (6), 374-389.

89. Brown, J.; O'Callaghan, C. A.; Marshall, A. S.; Gilbert, R. J.; Siebold, C.; Gordon, S.; Brown, G. D.; Jones, E. Y., Structure of the fungal beta-glucan-binding immune receptor dectin-1: implications for function. *Protein science : a publication of the Protein Society* **2007,** *16* (6), 1042-52.

90. Cummings, R. D.; Chiffoleau, E.; van Kooyk, Y.; McEver, R. P., C-Type Lectins. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor

Laboratory Press Copyright © 2022 The Consortium of Glycobiology Editors, La Jolla, California; published by Cold Spring Harbor Laboratory Press; doi:10.1101/glycobiology.4e.34. All rights reserved.: Cold Spring Harbor (NY), **2022**; pp 455-74.

91. Moriuchi, H.; Unno, H.; Goda, S.; Tateno, H.; Hirabayashi, J.; Hatakeyama, T., Mannose-recognition mutant of the galactose/N-acetylgalactosamine-specific C-type lectin CEL-I engineered by site-directed mutagenesis. *Biochimica et Biophysica Acta -General Subjects* **2015,** *1850* (7), 1457-1465.

92. Alenton, R. R.; Koiwai, K.; Miyaguchi, K.; Kondo, H.; Hirono, I., Pathogen recognition of a novel C-type lectin from Marsupenaeus japonicus reveals the divergent sugar-binding specificity of QAP motif. *Scientific reports* **2017,** *7*, 45818.

93. Drickamer, K.; Taylor, M. E., Recent insights into structures and functions of C-type lectins in the immune system. *Current Opinion in Structural Biology* **2015,** *34*, 26-34.

94. van Kooyk, Y.; Ilarregui, J. M.; van Vliet, S. J., Novel insights into the immunomodulatory role of the dendritic cell and macrophage-expressed C-type lectin MGL. *Immunobiology* **2015,** *220* (2), 185-192.

95. Gringhuis, S. I.; Den Dunnen, J.; Litjens, M.; Van Der Vlist, M.; Geijtenbeek, T. B., Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. *Nature immunology* **2009,** *10* (10), 1081-1088.

96. Geijtenbeek, T. B.; Torensma, R.; van Vliet, S. J.; van Duijnhoven, G. C.; Adema, G. J.; van Kooyk, Y.; Figdor, C. G., Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **2000,** *100* (5), 575-85.

97. Mitchell, D. A.; Fadden, A. J.; Drickamer, K., A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR: subunit organization and binding to multivalent ligands. *J Biol Chem* **2001,** *276* (31), 28939-28945.

98. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D. A.; Alvarez, R.; Blixt, O.; Taylor, M. E.; Weis, W. I.; Drickamer, K., Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nature Structural & Molecular Biology* **2004,** *11* (7), 591-598.

99. Engering, A.; Geijtenbeek, T. B.; van Vliet, S. J.; Wijers, M.; van Liempt, E.; Demaurex, N.; Lanzavecchia, A.; Fransen, J.; Figdor, C. G.; Piguet, V., The dendritic cell-specific adhesion receptor DC-SIGN internalizes antigen for presentation to T cells. *The Journal of Immunology* **2002,** *168* (5), 2118-2126.

100. Michaelis, L.; Treß, M.; Löw, H. C.; Klees, J.; Klameth, C.; Lange, A.; Grießhammer, A.; Schäfer, A.; Menz, S.; Steimle, A.; Schulze-Osthoff, K.; Frick, J. S., Gut Commensal-Induced IκBζ Expression in Dendritic Cells Influences the Th17 Response. *Front Immunol* **2020,** *11*, 612336.

101. van Liempt, E.; Bank, C. M.; Mehta, P.; Garciá-Vallejo, J. J.; Kawar, Z. S.; Geyer, R.; Alvarez, R. A.; Cummings, R. D.; Kooyk, Y.; van Die, I., Specificity of

DC-SIGN for mannose- and fucose-containing glycans. *FEBS letters* **2006,** *580* (26), 6123-31.

102. Bloem, K.; García-Vallejo, J. J.; Vuist, I. M.; Cobb, B. A.; van Vliet, S. J.; van Kooyk, Y., Interaction of the Capsular Polysaccharide A from Bacteroides fragilis with DC-SIGN on Human Dendritic Cells is Necessary for Its Processing and Presentation to T Cells. *Front Immunol* **2013,** *4*, 103.

103. Zhang, P.; Snyder, S.; Feng, P.; Azadi, P.; Zhang, S.; Bulgheresi, S.; Sanderson, K. E.; He, J.; Klena, J.; Chen, T., Role of N-acetylglucosamine within core lipopolysaccharide of several species of gram-negative bacteria in targeting the DC-SIGN (CD209). *Journal of immunology (Baltimore, Md. : 1950)* **2006,** *177* (6), 4002-11.

104. Klena, J.; Zhang, P.; Schwartz, O.; Hull, S.; Chen, T., The core lipopolysaccharide of Escherichia coli is a ligand for the dendritic-cell-specific intercellular adhesion molecule nonintegrin CD209 receptor. *Journal of bacteriology* **2005,** *187* (5), 1710-5.

105. Lewis, A.; Kohler, J.; Aebi, M., Microbial Lectins: Hemagglutinins, Adhesins, and Toxins. In *Essentials of Glycobiology*, 4th ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press Copyright © 2009, The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), **2022.**

106. Dekaboruah, E.; Suryavanshi, M. V.; Chettri, D.; Verma, A. K., Human microbiome: an academic update on human body site specific surveillance and its possible role. *Archives of microbiology* **2020,** *202* (8), 2147-2167.

107. Stones, D. H.; Krachler, A. M., Against the tide: the role of bacterial adhesion in host colonization. *Biochemical Society transactions* **2016,** *44* (6), 1571-1580.

108. Lewis, A. L.; Szymanski, C. M.; Schnaar, R. L.; Aebi, M., Bacterial and Viral Infections. In *Essentials of Glycobiology*, 4th ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press: **2022.**

109. Werneburg, G. T.; Thanassi, D. G., Pili Assembled by the Chaperone/Usher Pathway in Escherichia coli and Salmonella. *EcoSal Plus* **2018,** *8* (1).

110. Servin, A. L., Pathogenesis of Afa/Dr diffusely adhering Escherichia coli. *Clin Microbiol Rev* **2005,** *18* (2), 264-92.

111. Kerr, J. R.; Matthews, R. C., Bordetella pertussis Infection: Pathogenesis, Diagnosis, Management, and the Role of Protective Immunity. *European Journal of Clinical Microbiology and Infectious Diseases* **2000,** *19* (2), 77-88.

112. Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; De Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A., Burkholderia cenocepacia lectin A binding to heptoses from the bacterial lipopolysaccharide. *Glycobiology* **2012,** *22* (10), 1387-98.

113. Werdan, K.; Dietz, S.; Löffler, B.; Niemann, S.; Bushnaq, H.; Silber, R.-E.; Peters, G.; Müller-Werdan, U., Mechanisms of infective endocarditis: pathogen–host interaction and risk states. *Nature Reviews Cardiology* **2014,** *11* (1), 35-50.

114. Di Carluccio, C.; Forgione, R. E.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B. A.; Marchetti, R.;

Silipo, A., Molecular recognition of sialoglycans by streptococcal Siglec-like adhesins: toward the shape of specific inhibitors. *RSC chemical biology* **2021,** *2* (6), 1618-1630.

115. Forgione, R. E.; Di Carluccio, C.; Guzmán-Caldentey, J.; Gaglione, R.; Battista, F.; Chiodo, F.; Manabe, Y.; Arciello, A.; Del Vecchio, P.; Fukase, K.; Molinaro, A.; Martin-Santamaria, S.; Crocker, P. R.; Marchetti, R.; Silipo, A., Unveiling molecular recognition of sialoglycans by human siglec-10. *iScience* **2020,** *23* (6).

116. Talkington, D. F.; Waites, K. B.; Schwartz, S. B.; Besser, R. E., Emerging from obscurity: understanding pulmonary and extrapulmonary syndromes, pathogenesis, and epidemiology of human Mycoplasma pneumoniae infections. *Emerg. Infect. Dis.* **2001**, 57-84.

117. Roberts, D. D.; Olson, L. D.; Barile, M. F.; Ginsburg, V.; Krivan, H. C., Sialic Acid-Dependent Adhesion of Mycoplasma pneumoniae to Purified Glycoproteins. *J. Biol. Chem.* **1989,** *264* (16), 9289-9293.

118. Waites, K. B.; Talkington, D. F., Mycoplasma pneumoniae and Its Role as a Human Pathogen. *Clin. Microbiol. Rev.* **2004,** *17* (4), 697-728.

119. Atkinson, T. P.; Balish, M. F.; Waites, K. B., Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections. *FEMS microbiology reviews* **2008,** *32* (6), 956-73.

120. Yueyue, W.; Feichen, X.; Yixuan, X.; Lu, L.; Yiwen, C.; Xiaoxing, Y., Pathogenicity and virulence of Mycoplasma genitalium: Unraveling Ariadne's Thread. *Virulence* **2022,** *13* (1), 1161-1183.

121. McGowin, C. L.; Anderson-Smits, C., Mycoplasma genitalium: an emerging cause of sexually transmitted disease in women. *PLoS pathogens* **2011,** *7* (5), e1001324.

122. Ona, S.; Molina, R. L.; Diouf, K., Mycoplasma genitalium: An Overlooked Sexually Transmitted Pathogen in Women? *Infectious diseases in obstetrics and gynecology* **2016,** *2016*, 4513089.

123. Nakane, D.; Kenri, T.; Matsuo, L.; Miyata, M., Systematic Structural Analyses of Attachment Organelle in Mycoplasma pneumoniae. *PLoS Pathog.* **2015,** *11* (12), e1005299.

124. Vizarraga, D.; Torres-Puig, S.; Aparicio, D.; Pich, O. Q., The Sialoglycan Binding Adhesins of Mycoplasma genitalium and Mycoplasma pneumoniae. *Trends in microbiology* **2021,** *29* (6), 477-481.

125. Aparicio, D.; Torres-Puig, S.; Ratera, M.; Querol, E.; Piñol, J.; Pich, O. Q.; Fita, I., Mycoplasma genitalium adhesin P110 binds sialic-acid human receptors. *Nat. Commun.* **2018,** *9* (1), 4471.

126. Vizarraga, D.; Kawamoto, A.; Matsumoto, U.; Illanes, R.; Pérez-Luque, R.; Martín, J.; Mazzolini, R.; Bierge, P.; Pich, O. Q.; Espasa, M.; Sanfeliu, I.; Esperalba, J.; Fernández-Huerta, M.; Scheffer, M. P.; Pinyol, J.; Frangakis, A. S.; Lluch-Senar, M.; Mori, S.; Shibayama, K.; Kenri, T.; Kato, T.; Namba, K.; Fita, I.; Miyata, M.; Aparicio, D., Immunodominant proteins P1 and P40/P90 from human pathogen Mycoplasma pneumoniae. *Nat. Commun.* **2020,** *11* (1).

127. Aparicio, D.; Scheffer, M. P.; Marcos-Silva, M.; Vizarraga, D.; Sprankel, L.; Ratera, M.; Weber, M. S.; Seybert, A.; Torres-Puig, S.; Gonzalez-Gonzalez, L.; Reitz, J.; Querol, E.; Piñol, J.; Pich, O. Q.; Fita, I.; Frangakis, A. S., Structure and mechanism of the Nap adhesion complex from the human pathogen Mycoplasma genitalium. *Nat. Commun.* **2020,** *11* (1), 2877.

128. Chernova, O. A.; Medvedeva, E. S.; Mouzykantov, A. A.; Baranova, N. B.; Chernov, V. M., Mycoplasmas and Their Antibiotic Resistance: The Problems and Prospects in Controlling Infections. *Acta naturae* **2016,** *8* (2), 24-34.

129. De Castro, C.; Parrilli, M.; Holst, O.; Molinaro, A., Chapter Five - Microbe-Associated Molecular Patterns in Innate Immunity: Extraction and Chemical Analysis of Gram-Negative Bacterial Lipopolysaccharides. In *Methods in Enzymology*, Fukuda, M., Ed. Academic Press: **2010**; Vol. 480, pp 89-115.

130. De Castro, C. Gas Chromatography - Mass Spectrometry for Glycosciences. https://www.glycopedia.eu/.

131. Speciale, I.; Notaro, A.; Garcia-Vello, P.; Di Lorenzo, F.; Armiento, S.; Molinaro, A.; Marchetti, R.; Silipo, A.; De Castro, C., Liquid-state NMR spectroscopy for complex carbohydrate structural analysis: A hitchhiker's guide. *Carbohydr. Polym.* **2022,** *277*, 118885.

132. Palmer, I.; Wingfield, P. T., Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Current protocols in protein science* **2004,** *Chapter 6*, 6.3.1-6.3.18.

133. Furukawa, A.; Konuma, T.; Yanaka, S.; Sugase, K., Quantitative analysis of protein–ligand interactions by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2016,** *96*, 47-57.

134. Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K., Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2018,** *19* (8), 895-906.

135. Marchetti, R.; Perez, S.; Arda, A.; Imberty, A.; Jimenez-Barbero, J.; Silipo, A.; Molinaro, A., “Rules of engagement” of protein–glycoconjugate interactions: a molecular view achievable by using NMR spectroscopy and molecular modeling. *ChemistryOpen* **2016,** *5* (4), 274-296.

136. Meyer, B.; Peters, T., NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition* **2003,** *42* (8), 864-890.

137. Neuhaus, D.; Williamson, M. P. In *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, **1989**.

138. Di Carluccio, C.; Forgione, M. C.; Martini, S.; Berti, F.; Molinaro, A.; Marchetti, R.; Silipo, A., Investigation of protein-ligand complexes by ligand-based NMR methods. *Carbohydrate research* **2021,** *503*, 108313.

139. Angulo, J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Krishna, N. R.; Peters, H.; Palcic, M. M.; Peters, T., Blood group B galactosyltransferase: insights into substrate binding from NMR experiments. *Journal of the American Chemical Society* **2006,** *128* (41), 13529-13538.

140. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angewandte Chemie (International ed. in English)* **1999,** *38* (12), 1784-1788.

141. Mayer, M.; Meyer, B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* **2001,** *123* (25), 6108-17.

142. Angulo, J.; Enríquez-Navas, P. M.; Nieto, P. M., Ligand–receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry–A European Journal* **2010,** *16* (26), 7803-7812.

143. Szczepina, M. G.; Zheng, R. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M., STD-NMR studies of two acceptor substrates of GlfT2, a galactofuranosyltransferase from Mycobacterium tuberculosis: epitope mapping studies. *Bioorganic medicinal chemistry* **2010,** *18* (14), 5123-5128.

144. Alderson, T. R.; Kay, L. E., NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* **2021,** *184* (3), 577-595.

145. Giuntini, S.; Balducci, E.; Cerofolini, L.; Ravera, E.; Fragai, M.; Berti, F.; Luchinat, C., Characterization of the Conjugation Pattern in Large Polysaccharide-Protein Conjugates by NMR Spectroscopy. *Angewandte Chemie (International ed. in English)* **2017,** *56* (47), 14997-15001.

146. Ardá, A.; Canales, A.; Cañada, F. J.; Jiménez-Barbero, J., Carbohydrate–Protein Interactions: A 3D View by NMR. In *Carbohydrates in Drug Design and Discovery*, Jimenez-Barbero, J.; Canada, F. J.; Martin-Santamaria, S., Eds. The Royal Society of Chemistry: **2015**; pp 1-20.

147. Williamson, M. P., Chemical Shift Perturbation. In *Modern Magnetic Resonance*, Webb, G. A., Ed. Springer International Publishing: Cham, **2018**; pp 995-1012.

148. Fernández, C.; Wider, G., TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **2003,** *13* (5), 570-80.

149. Ribeiro, M. M.; Franquelim, H. G.; Castanho, M. A.; Veiga, A. S., Molecular interaction studies of peptides using steady-state fluorescence intensity. Static (de) quenching revisited. *J. Pept. Sci.* **2008,** *14* (4), 401-406.

150. Boguszewska, K.; Szewczuk, M.; Urbaniak, S.; Karwowski, B., Review: immunoassays in DNA damage and instability detection. *Cellular and Molecular Life Sciences* **2019,** *76*.

151. Sapay, N.; Nurisso, A.; Imberty, A., Simulation of carbohydrates, from molecular docking to dynamics in water. *Methods in molecular biology (Clifton, N.J.)* **2013,** *924*, 469-83.

152. Woods, R. J.; Tessier, M. B., Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr Opin Struct Biol* **2010,** *20* (5), 575-83.

153. Pérez, S., Computational modeling of protein-carbohydrate interactions: Current trends and future challenges. *Advances in carbohydrate chemistry and biochemistry* **2023,** *83*, 133-149.

154. Grant, O. C.; Woods, R. J., Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. *Curr Opin Struct Biol* **2014,** *28*, 47-55.

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

156. Pagadala, N. S.; Syed, K.; Tuszynski, J., Software for molecular docking: a review. *Biophysical reviews* **2017,** *9* (2), 91-102.

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

158. Crampon, K.; Giorkallos, A.; Deldossi, M.; Baud, S.; Steffenel, L. A., Machine-learning methods for ligand-protein molecular docking. *Drug discovery today* **2022,** *27* (1), 151-164.

159. Dias, R.; De Azevedo Jr, W., Molecular Docking Algorithms. *Current drug targets* **2009,** *9*, 1040-7.

160. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery* **2015,** *10* (5), 449-61.

161. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000,** *295* (2), 337-56.

162. Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E., Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein engineering* **1995,** *8* (7), 677-91.

163. Böhm, H. J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of computer-aided molecular design* **1994,** *8* (3), 243-56.

164. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design* **1997,** *11* (5), 425-45.

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

166. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.,* **1998,** *19* (14), 1639-1662.

167. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking

with selective receptor flexibility. *Journal of computational chemistry* **2009,** *30* (16), 2785-91.

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

169. Dickson, C. J.; Walker, R. C.; Gould, I. R., Lipid21: Complex Lipid Membrane Simulations with AMBER. *Journal of chemical theory and computation* **2022,** *18* (3), 1726-1736.

170. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry* **2008,** *29* (4), 622-55.

171. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation* **2015,** *11* (8), 3696-713.

172. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E., 3rd, Assessing the Current State of Amber Force Field Modifications for DNA. *Journal of chemical theory and computation* **2016,** *12* (8), 4114-27.

173. Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E., III; Jurečka, P., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of chemical theory and computation* **2011,** *7* (9), 2886-2902.

174. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

175. He, X.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J., A fast and high-quality charge model for the next generation general AMBER force field. *J Chem Phys* **2020,** *153* (11), 114502.

176. Patel, D. S.; Pendrill, R.; Mallajosyula, S. S.; Widmalm, G.; MacKerell, A. D., Jr., Conformational properties of α- or β-(1→6)-linked oligosaccharides: Hamiltonian replica exchange MD simulations and NMR experiments. *The journal of physical chemistry. B* **2014,** *118* (11), 2851-71.

177. French, A. D., Comparisons of rigid and relaxed conformational maps for cellobiose and maltose. *Carbohydrate Res.* **1989,** *188*, 206-211.

178. Allinger, N.; Rahman, M.; Lii, J. H., A Molecular Mechanics Force Field (MM3) for Alcohols and Ethers. *J. Am. Chem. Soc.* **1990,** *112*, 8293–8307.

179. Mackerell, A. D., Jr., Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry* **2004,** *25* (13), 1584-604.

180. Perez, S.; Fadda, E.; Makshakova, O., *Computational Modeling in Glycoscience*. 2nd ed.; **2020.**

181. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.

182. Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D., Jr., Molecular mechanics. *Current pharmaceutical design* **2014,** *20* (20), 3281-92.

183. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D., Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of Modern Physics* **1992,** *64* (4), 1045-1097.

184. Jayalakshmi, V.; Krishna, N. R., Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2002,** *155* (1), 106-18.

185. Angulo, J.; Nieto, P. M., STD-NMR: application to transient interactions between biomolecules-a quantitative approach. *European biophysics journal : EBJ* **2011,** *40* (12), 1357-69.

186. Bäckhed, F.; Ley, R. E.; Sonnenburg, J. L.; Peterson, D. A.; Gordon, J. I., Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)* **2005,** *307* (5717), 1915-20.

187. Maalej, M.; Forgione, R. E.; Marchetti, R.; Bulteau, F.; Thépaut, M.; Lanzetta, R.; Laguri, C.; Simorre, J. P.; Fieschi, F.; Molinaro, A.; Silipo, A., Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of Escherichia coli Lipooligosaccharide. *Chembiochem : a European journal of chemical biology* **2019,** *20* (14), 1778-1782.

188. Kusumoto, S.; Fukase, K.; Shiba, T., Key structures of bacterial peptidoglycan and lipopolysaccharide triggering the innate immune system of higher animals: chemical synthesis and functional studies. *Proceedings of the Japan Academy. Series B, Physical and biological sciences* **2010,** *86* (4), 322-37.

189. Steimle, A.; Autenrieth, I. B.; Frick, J.-S., Structure and function: Lipid A modifications in commensals and pathogens. *International Journal of Medical Microbiology* **2016,** *306* (5), 290-301.

190. Mnich, M. E.; Van Dalen, R.; Van Sorge, N. M., C-type lectin receptors in host defense against bacterial pathogens. *Front. cell. infect. microbiol.* **2020,** *10*, 309.

191. van Kooyk, Y.; Geijtenbeek, T. B., DC-SIGN: escape mechanism for pathogens. *Nature reviews. Immunology* **2003,** *3* (9), 697-709.

192. Zhang, P.; Schwartz, O.; Pantelic, M.; Li, G.; Knazze, Q.; Nobile, C.; Radovich, M.; He, J.; Hong, S. C.; Klena, J.; Chen, T., DC-SIGN (CD209) recognition of Neisseria gonorrhoeae is circumvented by lipooligosaccharide variation. *Journal of leukocyte biology* **2006,** *79* (4), 731-8.

193. Iyori, M.; Ohtani, M.; Hasebe, A.; Totsuka, Y.; Shibata, K. I., A role of the Ca2+ binding site of DC-SIGN in the phagocytosis of E. coli. *Biochem Biophys Res Commun* **2008,** *377* (2), 367-372.

194. Bergman, M. P.; Engering, A.; Smits, H. H.; van Vliet, S. J.; van Bodegraven, A. A.; Wirth, H. P.; Kapsenberg, M. L.; Vandenbroucke-Grauls, C. M.; van Kooyk, Y.; Appelmelk, B. J., Helicobacter pylori modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. *The Journal of experimental medicine* **2004,** *200* (8), 979-90.

195. Zhang, P.; Skurnik, M.; Zhang, S. S.; Schwartz, O.; Kalyanasundaram, R.; Bulgheresi, S.; He, J. J.; Klena, J. D.; Hinnebusch, B. J.; Chen, T., Human dendritic cell-specific intercellular adhesion molecule-grabbing nonintegrin (CD209) is a receptor for Yersinia pestis that promotes phagocytosis by dendritic cells. *Infection and immunity* **2008,** *76* (5), 2070-9.

196. Chiffoleau, E., C-Type Lectin-Like Receptors As Emerging Orchestrators of Sterile Inflammation Represent Potential Therapeutic Targets. *Front Immunol* **2018,** *9*, 227.

197. Allocati, N.; Masulli, M.; Alexeyev, M. F.; Di Ilio, C., Escherichia coli in Europe: an overview. *International journal of environmental research and public health* **2013,** *10* (12), 6235-54.

198. Leclercq, S. O.; Branger, M.; Smith, D. G. E.; Germon, P., Lipopolysaccharide core type diversity in the Escherichia coli species in association with phylogeny, virulence gene repertoire and distribution of type VI secretion systems. *Microbial genomics* **2021,** *7* (9).

199. Amor, K.; Heinrichs, D. E.; Frirdich, E.; Ziebell, K.; Johnson, R. P.; Whitfield, C., Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. *Infection and immunity* **2000,** *68* (3), 1116-24.

200. Halary, F.; Amara, A.; Lortat-Jacob, H.; Messerle, M.; Delaunay, T.; Houlès, C.; Fieschi, F.; Arenzana-Seisdedos, F.; Moreau, J. F.; Déchanet-Merville, J., Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. *Immunity* **2002,** *17* (5), 653-64.

201. Porkolab, V.; Pifferi, C.; Sutkeviciute, I.; Ordanini, S.; Taouai, M.; Thépaut, M.; Vivès, C.; Benazza, M.; Bernardi, A.; Renaudet, O.; Fieschi, F., Development of C-type lectin-oriented surfaces for high avidity glycoconjugates: towards mimicking multivalent interactions on the cell surface. *Organic & Biomolecular Chemistry* **2020,** *18* (25), 4763-4772.

202. Group, W. GLYCAM Web. http://glycam.org.

203. Angulo, J.; Díaz, I.; Reina, J. J.; Tabarani, G.; Fieschi, F.; Rojo, J.; Nieto, P. M., Saturation transfer difference (STD) NMR spectroscopy characterization of dual binding mode of a mannose disaccharide to DC-SIGN. *Chembiochem : a European journal of chemical biology* **2008,** *9* (14), 2225-7.

204. Guzzi, C.; Alfarano, P.; Sutkeviciute, I.; Sattin, S.; Ribeiro-Viana, R.; Fieschi, F.; Bernardi, A.; Weiser, J.; Rojo, J.; Angulo, J.; Nieto, P. M., Detection and quantitative analysis of two independent binding modes of a small ligand

responsible for DC-SIGN clustering. *Organic & Biomolecular Chemistry* **2016,** *14* (1), 335-344.

205. Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P. H.; Weis, W. I., Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J Biol Chem* **2007,** *282* (6), 4202-9.

206. Sutkeviciute, I.; Thépaut, M.; Sattin, S.; Berzi, A.; McGeagh, J.; Grudinin, S.; Weiser, J.; Le Roy, A.; Reina, J. J.; Rojo, J.; Clerici, M.; Bernardi, A.; Ebel, C.; Fieschi, F., Unique DC-SIGN clustering activity of a small glycomimetic: A lesson for ligand design. *ACS chemical biology* **2014,** *9* (6), 1377-85.

207. Delaunay, C.; Pollastri, S.; Thépaut, M.; Cavazzoli, G.; Belvisi, L.; Bouchikri, C.; Labiod, N.; Gimeno, A.; Franconetti, A.; Jiménez-Barbero, J.; Ardá, A.; Delgado, R.; Bernardi, A.; Fieschi, F., Development of a highly selective glycomimetic ligand for L-SIGN: a new tool against SARS-CoV-2 and Ebola viruses. *ChemRxiv.* **2023**.

208. Thépaut, M.; Luczkowiak, J.; Vivès, C.; Labiod, N.; Bally, I.; Lasala, F.; Grimoire, Y.; Fenel, D.; Sattin, S.; Thielens, N.; Schoehn, G.; Bernardi, A.; Delgado, R.; Fieschi, F., DC/L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist. *PLoS pathogens* **2021,** *17* (5), e1009576.

209. de Vos, W. M.; Tilg, H.; Van Hul, M.; Cani, P. D., Gut microbiome and health: mechanistic insights. *Gut* **2022,** *71* (5), 1020-1032.

210. Takeuchi, O.; Akira, S., Pattern recognition receptors and inflammation. *Cell* **2010,** *140* (6), 805-20.

211. Steimle, A.; Michaelis, L.; Di Lorenzo, F.; Kliem, T.; Münzner, T.; Maerz, J. K.; Schäfer, A.; Lange, A.; Parusel, R.; Gronbach, K.; Fuchs, K.; Silipo, A.; Öz, H. H.; Pichler, B. J.; Autenrieth, I. B.; Molinaro, A.; Frick, J. S., Weak Agonistic LPS Restores Intestinal Immune Homeostasis. *Molecular therapy : the journal of the American Society of Gene Therapy* **2019,** *27* (11), 1974-1991.

212. Mohr, A. E.; Crawford, M.; Jasbi, P.; Fessler, S.; Sweazea, K. L., Lipopolysaccharide and the gut microbiota: considering structural variation. *FEBS letters* **2022,** *596* (7), 849-875.

213. Wexler, H. M., Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **2007,** *20* (4), 593-621.

214. Wang, C.; Xiao, Y.; Yu, L.; Tian, F.; Zhao, J.; Zhang, H.; Chen, W.; Zhai, Q., Protective effects of different Bacteroides vulgatus strains against lipopolysaccharide-induced acute intestinal injury, and their underlying functional genes. *Journal of advanced research* **2022,** *36*, 27-37.

215. Li, S.; Wang, C.; Zhang, C.; Luo, Y.; Cheng, Q.; Yu, L.; Sun, Z., Evaluation of the Effects of Different Bacteroides vulgatus Strains against DSS-Induced Colitis. *Journal of immunology research* **2021,** *2021*, 9117805.

216. Wexler, A. G.; Goodman, A. L., An insider's perspective: Bacteroides as a window into the microbiome. *Nature microbiology* **2017,** *2*, 17026.

217. Di Lorenzo, F.; Pither, M. D.; Martufi, M.; Scarinci, I.; Guzmán-Caldentey, J.; Łakomiec, E.; Jachymek, W.; Bruijns, S. C. M.; Santamaría, S. M.; Frick, J.-S.;

van Kooyk, Y.; Chiodo, F.; Silipo, A.; Bernardini, M. L.; Molinaro, A., Pairing Bacteroides vulgatus LPS Structure with Its Immunomodulatory Effects on Human Cellular Models. *ACS Cent Sci* **2020,** *6* (9), 1602-1616.

218. Singh, S.; Almuhanna, Y.; Alshahrani, M. Y.; Lowman, D. W.; Rice, P. J.; Gell, C.; Ma, Z.; Graves, B.; Jackson, D.; Lee, K.; Juarez, R.; Koranteng, J.; Muntaka, S.; Daniel, A. M.; da Silva, A. C.; Hussain, F.; Yilmaz, G.; Mastrotto, F.; Irie, Y.; Williams, P.; Williams, D. L.; Cámara, M.; Martinez-Pomares, L., Carbohydrates from Pseudomonas aeruginosa biofilms interact with immune C-type lectins and interfere with their receptor function. *NPJ biofilms and microbiomes* **2021,** *7* (1), 87.

219. Rahimi, N., C-type Lectin CD209L/L-SIGN and CD209/DC-SIGN: Cell Adhesion Molecules Turned to Pathogen Recognition Receptors. *Biology* **2020,** *10* (1).

220. Garcia-Vallejo, Juan J.; van Kooyk, Y., DC-SIGN: The Strange Case of Dr. Jekyll and Mr. Hyde. *Immunity* **2015,** *42* (6), 983-985.

221. Zhu, Q.; Shen, Z.; Chiodo, F.; Nicolardi, S.; Molinaro, A.; Silipo, A.; Yu, B., Chemical synthesis of glycans up to a 128-mer relevant to the O-antigen of Bacteroides vulgatus. *Nat Commun* **2020,** *11* (1), 4142.

222. *Schrödinger Release 2022-3*, Maestro, Schrödinger, LLC: New York, **2021.**

223. *Schrödinger Release 2022-3*, MacroModel, Schrödinger, LLC: New York, **2021.**

224. Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I., Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science (New York, N.Y.)* **2001,** *294* (5549), 2163-6.

225. Cramer, J.; Lakkaichi, A.; Aliu, B.; Jakob, R. P.; Klein, S.; Cattaneo, I.; Jiang, X.; Rabbani, S.; Schwardt, O.; Zimmer, G.; Ciancaglini, M.; Abreu Mota, T.; Maier, T.; Ernst, B., Sweet Drugs for Bad Bugs: A Glycomimetic Strategy against the DC-SIGN-Mediated Dissemination of SARS-CoV-2. *J Am Chem Soc* **2021,** *143* (42), 17465-17478.

226. Sattin, S.; Daghetti, A.; Thépaut, M.; Berzi, A.; Sánchez-Navarro, M.; Tabarani, G.; Rojo, J.; Fieschi, F.; Clerici, M.; Bernardi, A., Inhibition of DC-SIGN-mediated HIV infection by a linear trimannoside mimic in a tetravalent presentation. *ACS chemical biology* **2010,** *5* (3), 301-12.

227. WoodsGroup *GLYCAM Web*, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, 2005-2022.

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

Wu, X.; Xiao, L.; York, M.; Kollman, P. A. *AMBER 2018*, University of California, San Francisco, **2018.**

229. Valverde, P.; Delgado, S.; Martínez, J. D.; Vendeville, J. B.; Malassis, J.; Linclau, B.; Reichardt, N. C.; Cañada, F. J.; Jiménez-Barbero, J.; Ardá, A., Molecular Insights into DC-SIGN Binding to Self-Antigens: The Interaction with the Blood Group A/B Antigens. *ACS chemical biology* **2019,** *14* (7), 1660-1671.
230. Mnich, M. E.; van Dalen, R.; van Sorge, N. M., C-Type Lectin Receptors in Host Defense Against Bacterial Pathogens. *Frontiers in cellular and infection microbiology* **2020,** *10*, 309.
231. McCarville, J. L.; Caminero, A.; Verdu, E. F., Novel perspectives on therapeutic modulation of the gut microbiota. *Therapeutic Advances in Gastroenterology* **2016,** *9* (4), 580-593.
232. Castro-Bravo, N.; Wells, J. M.; Margolles, A.; Ruas-Madiedo, P., Interactions of Surface Exopolysaccharides From Bifidobacterium and Lactobacillus Within the Intestinal Environment. *Front Microbiol* **2018,** *9*, 2426.
233. Oerlemans, M. M. P.; Akkerman, R.; Ferrari, M.; Walvoort, M. T. C.; de Vos, P., Benefits of bacteria-derived exopolysaccharides on gastrointestinal microbiota, immunity and health. *Journal of Functional Foods* **2021,** *76*, 104289.
234. Pizarro-Cerdá, J.; Cossart, P., Bacterial Adhesion and Entry into Host Cells. *Cell* **2006,** *124* (4), 715-727.
235. Qin, L.; Chen, Y.; You, X., Subversion of the Immune Response by Human Pathogenic Mycoplasmas. *Front. Microbiol.* **2019,** *10.*
236. Kashyap, S.; Sarkar, M., Mycoplasma pneumonia: Clinical features and management. *Lung India : official organ of Indian Chest Society* **2010,** *27* (2), 75-85.
237. Smullin, C. P.; Green, H.; Peters, R.; Nyemba, D.; Qayiya, Y.; Myer, L.; Klausner, J.; Joseph Davey, D., Prevalence and incidence of Mycoplasma genitalium in a cohort of HIV-infected and HIV-uninfected pregnant women in Cape Town, South Africa. *Sexually transmitted infections* **2020,** *96* (7), 501-508.
238. Bothner-By, A. A.; Gassend, R., Binding of small molecules to proteins. *Annals of the New York Academy of Sciences* **1973,** *222*, 668-76.
239. Clore, G. M.; Gronenborn, A. M., Theory of the time dependent transferred nuclear Overhauser effect: Applications to structural analysis of ligand-protein complexes in solution. *Journal of Magnetic Resonance (1969)* **1983,** *53* (3), 423-442.
240. Poppe, L.; Brown, G. S.; Philo, J. S.; Nikrad, P. V.; Shah, B. H., Conformation of sLex Tetrasaccharide, Free in Solution and Bound to E-, P-, and L-Selectin. *J. Am. Chem. Soc.* **1997,** *119* (7), 1727-1736.
241. Behren, S.; Westerlind, U., Novel Approaches To Design Glycan-Based Antibacterial Inhibitors. *European Journal of Organic Chemistry* **2023,** *26* (1), 1-13.
242. Fournier, P. E.; Richet, H.; Weinstein, R. A., The Epidemiology and Control of Acinetobacter baumannii in Health Care Facilities. *Clinical Infectious Diseases* **2006,** *42* (5), 692-699.
243. Denissen, J.; Reyneke, B.; Waso-Reyneke, M.; Havenga, B.; Barnard, T.; Khan, S.; Khan, W., Prevalence of ESKAPE pathogens in the environment: Antibiotic

resistance status, community-acquired infection and risk to human health. *International Journal of Hygiene and Environmental Health* **2022,** *244*, 114006.

244. Kenyon, J. J.; Hall, R. M., Variation in the Complex Carbohydrate Biosynthesis Loci of Acinetobacter baumannii Genomes. *PLOS ONE* **2013,** *8* (4), e62160.

245. Rappuoli, R.; Aderem, A., A 2020 vision for vaccines against HIV, tuberculosis and malaria. *Nature* **2011,** *473* (7348), 463-9.

246. Costantino, P.; Rappuoli, R.; Berti, F., The design of semi-synthetic and synthetic glycoconjugate vaccines. *Expert opinion on drug discovery* **2011,** *6* (10), 1045-66.

247. Peltola, H.; Mäkelä, H.; Käyhty, H.; Jousimies, H.; Herva, E.; Hällström, K.; Sivonen, A.; Renkonen, O. V.; Pettay, O.; Karanko, V.; Ahvonen, P.; Sarna, S., Clinical efficacy of meningococcus group A capsular polysaccharide vaccine in children three months to five years of age. *The New England journal of medicine* **1977,** *297* (13), 686-91.

248. Soysal, A.; Gönüllü, E.; Yıldız, I.; Aydemir, G.; Tunç, T.; Fırat, Y.; Erdamar, B.; Karaböcüoğlu, M., Impact of the 13-valent pneumococcal conjugate vaccine on the incidences of acute otitis media, recurrent otitis media and tympanostomy tube insertion in children after its implementation into the national immunization program in Turkey. *Human vaccines & immunotherapeutics* **2020,** *16* (2), 445-451.

249. Arunachalam, P. S.; Walls, A. C.; Golden, N.; Atyeo, C.; Fischinger, S.; Li, C.; Aye, P.; Navarro, M. J.; Lai, L.; Edara, V. V.; Röltgen, K.; Rogers, K.; Shirreff, L.; Ferrell, D. E.; Wrenn, S.; Pettie, D.; Kraft, J. C.; Miranda, M. C.; Kepl, E.; Sydeman, C.; Brunette, N.; Murphy, M.; Fiala, B.; Carter, L.; White, A. G.; Trisal, M.; Hsieh, C. L.; Russell-Lodrigue, K.; Monjure, C.; Dufour, J.; Spencer, S.; Doyle-Meyers, L.; Bohm, R. P.; Maness, N. J.; Roy, C.; Plante, J. A.; Plante, K. S.; Zhu, A.; Gorman, M. J.; Shin, S.; Shen, X.; Fontenot, J.; Gupta, S.; O'Hagan, D. T.; Van Der Most, R.; Rappuoli, R.; Coffman, R. L.; Novack, D.; McLellan, J. S.; Subramaniam, S.; Montefiori, D.; Boyd, S. D.; Flynn, J. L.; Alter, G.; Villinger, F.; Kleanthous, H.; Rappaport, J.; Suthar, M. S.; King, N. P.; Veesler, D.; Pulendran, B., Adjuvanting a subunit COVID-19 vaccine to induce protective immunity. *Nature* **2021,** *594* (7862), 253-258.

250. Micoli, F.; Bagnoli, F.; Rappuoli, R.; Serruto, D., The role of vaccines in combatting antimicrobial resistance. *Nature reviews. Microbiology* **2021,** *19* (5), 287-302.

251. Mishra, R. P. N.; Yadav, R. S. P.; Jones, C.; Nocadello, S.; Minasov, G.; Shuvalova, L. A.; Anderson, W. F.; Goel, A., Structural and immunological characterization of E. coli derived recombinant CRM(197) protein used as carrier in conjugate vaccines. *Bioscience reports* **2018,** *38* (5).

252. Malito, E.; Bursulaya, B.; Chen, C.; Surdo, P. L.; Picchianti, M.; Balducci, E.; Biancucci, M.; Brock, A.; Berti, F.; Bottomley, M. J.; Nissum, M.; Costantino, P.; Rappuoli, R.; Spraggon, G., Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. *Proc Natl Acad Sci U S A* **2012,** *109* (14), 5229-5234.

253. Shafer, D. E.; Toll, B.; Schuman, R. F.; Nelson, B. L.; Mond, J. J.; Lees, A., Activation of soluble polysaccharides with 1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP) for use in protein-polysaccharide conjugate vaccines and immunological reagents. II. Selective crosslinking of proteins to CDAP-activated polysaccharides. *Vaccine* **2000,** *18* (13), 1273-81.

254. Rauter, A. P.; Christensen, B. E.; Somsák, L.; Kosma, P.; Adamo, R., *Recent Trends in Carbohydrate Chemistry: Synthesis and Biomedical Applications of Glycans and Glycoconjugates*. Elsevier: **2020.**

255. Leone, S.; Silipo, A.; E, L. N.; Lanzetta, R.; Parrilli, M.; Molinaro, A., Molecular structure of endotoxins from Gram-negative marine bacteria: an update. *Marine drugs* **2007,** *5* (3), 85-112.

256. Kittelberger, R.; Hilbink, F., Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. *Journal of biochemical and biophysical methods* **1993,** *26* (1), 81-6.

257. Holst, O., Deacylation of lipopolysaccharides and isolation of oligosaccharide phosphates. *Methods in molecular biology (Clifton, N.J.)* **2000,** *145*, 345-53.

258. Reina, J. J.; Sattin, S.; Invernizzi, D.; Mari, S.; Martínez-Prats, L.; Tabarani, G.; Fieschi, F.; Delgado, R.; Nieto, P. M.; Rojo, J.; Bernardi, A., 1,2-Mannobioside mimic: synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity. *ChemMedChem* **2007,** *2* (7), 1030-6.

259. Tabarani, G.; Thépaut, M.; Stroebel, D.; Ebel, C.; Vivès, C.; Vachette, P.; Durand, D.; Fieschi, F., DC-SIGN neck domain is a pH-sensor controlling oligomerization: SAXS and hydrodynamic studies of extracellular domain. *J Biol Chem* **2009,** *284* (32), 21229-40.

260. Oliva, R.; Battista, F.; Cozzolino, S.; Notomista, E.; Winter, R.; Del Vecchio, P.; Petraccone, L., Encapsulating properties of sulfobutylether-β-cyclodextrin toward a thrombin-derived antimicrobial peptide. *J. Therm. Anal. Calorim.* **2019,** *138*, 3249-3256.

261. Luczkowiak, J.; Sattin, S.; Sutkevičiūtė, I.; Reina, J. J.; Sánchez-Navarro, M.; Thépaut, M.; Martínez-Prats, L.; Daghetti, A.; Fieschi, F.; Delgado, R.; Bernardi, A.; Rojo, J., Pseudosaccharide functionalized dendrimers as potent inhibitors of DC-SIGN dependent Ebola pseudotyped viral infection. *Bioconjugate chemistry* **2011,** *22* (7), 1354-65.

262. Andreini, M.; Doknic, D.; Sutkeviciute, I.; Reina, J. J.; Duan, J.; Chabrol, E.; Thepaut, M.; Moroni, E.; Doro, F.; Belvisi, L.; Weiser, J.; Rojo, J.; Fieschi, F.; Bernardi, A., Second generation of fucose-based DC-SIGN ligands : affinity improvement and specificity versus Langerin. *Organic & Biomolecular Chemistry* **2011,** *9* (16), 5778-5786.

263. Frisch, M. J. J. h. w. g. c., gaussian09. **2009**.

264. Kurt, B.; Temel, H., Parameterization of boronates using VFFDT and paramfit for molecular dynamics simulation. *Molecules* **2020,** *25* (9), 2196.

265. Zheng, S.; Tang, Q.; He, J.; Du, S.; Xu, S.; Wang, C.; Xu, Y.; Lin, F., VFFDT: a new software for preparing AMBER force field parameters for metal-

containing molecular systems. *Journal of chemical informationmodeling* **2016,** *56* (4), 811-818.

266. Roe, D. R.; Cheatham III; E, T., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theorycomputation* **2013,** *9* (7), 3084-3095.

267. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **2013,** *27* (3), 221-34.

268. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of molecular graphics* **1996,** *14* (1), 33-8, 27-8.

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”

UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Università
e della Ricerca

PON
RICERCA
E INNOVAZIONE
2014 - 2020

This PhD project is a collaboration between “University of Naples Federico II” and “GlaxoSmithKline Biologicals SA”.

GSK