University of Naples Federico II
Polytechnic and Basic Sciences School
Department of Chemical Sciences

UNIONE EUROPEA
Fondo Sociale Europeo

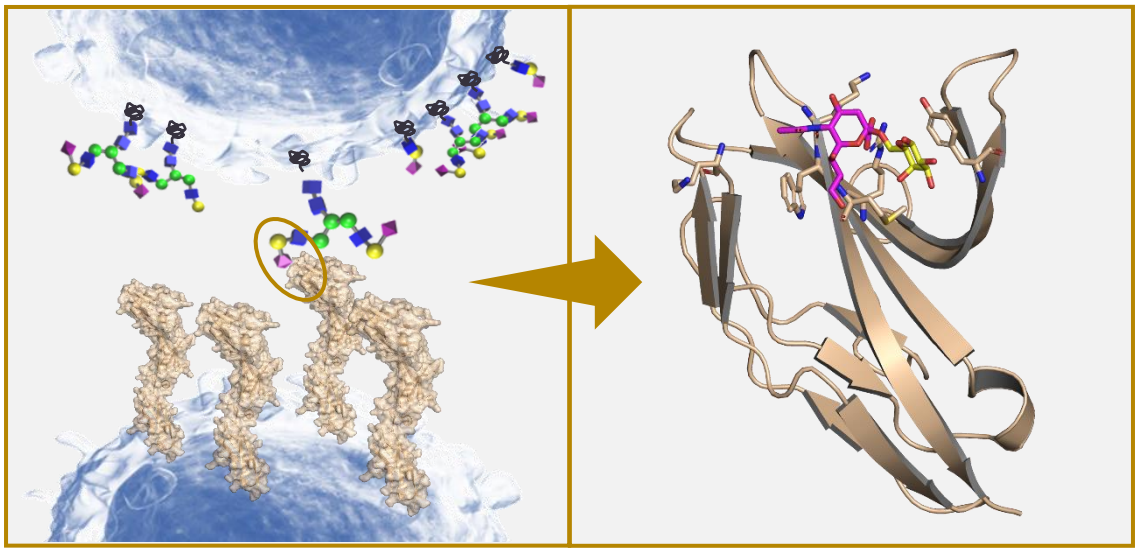
Ministero dell'Istruzione dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE

neapōlis

Ph.D. in Chemical Sciences

Study of the interaction between sialic acid binding immunoglobulin- type lectins (Siglecs) and sialylated glycans for the development of a new generation of immunomodulators



Rosa Ester Forgione

Advisors:
prof. Alba Silipo
prof. Sonsoles Martìn-Santamarìa
prof. Marco Fragai

Examiner:
prof. Maria Monti

XXXIII Cycle 2018 – 2021
Coordinator: prof. Angela Lombardi

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

To my beloved father

TABLE OF CONTENTS

List of Abbreviations i
Sugar Abbreviations v
ABSTRACT vi
SECTION I-INTRODUCTION 9
Chapter 1: The importance of glycocode based cross- talk 10
1.1 The sugar code 10
1.2 Biological roles of glycans 14
1.2.1 Structural and modulatory roles 15
1.2.2 Molecular mimicry of host glycans 17
1.2.3 Recognition by glycan binding receptors 19
1.3 Siglecs: a family of lectins for the specific recognition of sialic acid 20
1.3.1 The binding properties of Siglecs 25
1.3.2 Recognition between Siglecs and sialoglycans 27
1.3.3 Siglecs' roles in the immune system 30
Chapter 2- Tools to study receptor- glycans interactions 34
2.1 Investigation of glycan-receptors complexes 34
2.2 NMR methods for the study of protein- ligand interactions 35
2.2.1 NMR characterization of free state oligo and polysaccharides 35
2.2.2 NMR: the ligand-based approach 38
2.2.3 NOE and transferred NOE 42
2.2.4 Saturation transfer difference NMR 46
2.3 Computational techniques for carbohydrates modelling 51
2.3.1 Docking 52
2.3.2 Molecular Mechanics and Molecular Dynamics Simulations 55
2.3.3 Homology modelling 59
2.3.4 Complex validation: CORCEMA-ST 60
2.5 Objectives 62
SECTION II-RESULTS AND DISCUSSION 64
Chapter 3-Structural insights on the recognition of sialoglycans by Siglec-2 65
3.1 Introduction 65

3.2 Characterization of the dynamic interactions between complex N-glycans and human CD22 69
3.2.1 Affinity of h- CD22 toward different sialoglycans: SPR and Alpha Screen 69
3.2.2 Molecular recognition of sialotrisaccharide 1 73
3.2.3 Molecular recognition of complex-type N-glycans by h-CD22 82
3.3 Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22 89
3.3.1 Binding specificity of m- and h- CD22 toward Neu5Ac and Neu5Gc ligands 90
3.3.2 STD NMR analysis of Neu5Ac and Neu5Gc containing glycans by human and murine CD22 91
3.3.3 Molecular overview of the interaction of glycolylated glycans with murine and human CD22 101
3.3.4 Comparing acetylated and glycolylated glycans in the interaction with murine and human CD22 111
3.4 Molecular recognition of a sialic acid analog by CD22 113
3.5 Discussion 123
Chapter 4- Characterization of Siglec-10/ Sialoglycans interactions by means of biophysical, computational and spectroscopic techniques 127
4.1 Introduction 127
4.2 Unveiling molecular recognition of Sialoglycans and Siglec-10 132
4.2.1 Molecular interaction between Siglec-10 and trisaccharide 1 134
4.2.2 3D-View of Siglec10 in interaction with trisaccharide 1 139
4.2.3 Molecular interaction between Siglec-10 and trisaccharide 2 148
4.2.4 3D-View of Siglec10 in interaction with trisaccharide 2 152
4.3 Molecular interaction between Siglec-10 and complex-type N-glycans 156
4.4 Recombinant expression of Siglec-10 in Escherichia coli 158
4.5 Discussion 165
Chapter 5- Molecular basis for the recognition of Fusobacterium nucleatum lipopolysaccharide by Siglec-7 169
5.1 Introduction 169
5.1.1 Siglec-7 169

5.1.2 F. nucleatum 174
5.3 STD-NMR and fluorescence studies of F. Nucleatum LPS by Siglec-7 176
5.4 Molecular modelling of F. nucleatum O-chain and Siglec-7 182
5.4.1 F. nucleatum OPS hexasaccharide /Siglec-7 modelling 182
5.4.2 F. nucleatum OPS dodecasaccharide /Siglec-7 modelling 191
5.4 Discussion 196
SECTION III-OTHER PROJECTS 199
Chapter 6- Human Macrophage Galactose-Type Lectin recognizes the Outer Core of E. coli lipooligosaccharide 200
6.1 Introduction 200
6.1.1 Bacterial LPS 200
6.1.2 MGL- a major member of C- type lectins 202
6.2 Molecular recognition of MGL 203
6.4 Discussion 217
Chapter 7-Structural basis for Glycan-receptor binding by mumps virus hemagglutinin-neuraminidase 219
7.1 Introduction 219
7.1.1 Mumps Virus 219
7.2 Molecular basis of sialoglycan binding to MuV-HN from Hoshino strain 224
7.2.1 Kinetic analysis of the hydrolysis of sialoglycans by MuV-HN 225
7.2.2 Molecular recognition of sialoglycans by MuV-HN protein 230
7.3 Molecular basis of sialoglycan binding to MuV-HN from SBL-strain 249
7.4 Discussion 267
Chapter 8-Concluding remarks 273
SECTION IV- EXPERIMENTAL SECTION 277
Chapter 9 Experimental section 278
9.1 Materials 278
9.1.1 Recombinant expression of Siglecs 278
9.1.2 Recombinant expression of other proteins 281
9.1.3 Extraction of the LPS from F. nucleatum strain 10953 (Section II-Chapter 5) 283

9.1.5 Extraction of the LOS from E. coli OSR1(Section III-Chapter 6) 283
9.2 NMR methods 284
9.2.1 Free ligand spectroscopic characterization 284
9.2.2 Bound ligand spectroscopic characterization 285
9.3 Molecular mechanics and molecular dynamics simulation 288
9.3.1 Molecular mechanics simulations 288
9.3.2 Molecular dynamics simulations 289
9.3.3 Homology modelling 291
9.3.4 Docking 292
9.3.5 CORCEMA-ST 293
9.4 Ligand binding assays 294
9.4.1 Surface Plasmon Resonance (Section II-Chapter 3) 294
9.4.2 AlphaScreen (Section II-Chapter III) 295
9.4.3 Binding analysis by ELISA (Section II-Chapter 4) 295
9.4.4 Binding analysis by fluorescence spectroscopy (Section II- Chapters 3,4,5 and Section III- Chapter 7) 296
10 References 298
AKNOWLEDGMENTS 334

List of Abbreviations

PG- PROTEOGLYCANS

GAGs- GLYCOSAMINOGLYCANS

Hp-HEPARIN

CS- CHONDROITIN SULFATE

ER- ENDOPLASMIC RETICULUM

AGPS- ARABINOGALACTANPROTEINS

LPS- LIPOPOLISACCHARIDES

SNFG- SYMBOL NOMENCLATURE FOR GLYCANS

EM-EXTRACELLULAR MATRIX

NCAM- NATURAL CELL ADHESION MOLECULE

MAMPs- MOLECULAR ASSOCIATED MOLECULAR PATTERNS

PAMPs- PATHOGEN ASSOCIATED MOLECULAR PATTERNS

PRRs- PATTERN RECOGNITION RECEPTORS

TLRs- TOLL-LIKE RECEPTORS

CTLs- C-TYPE LECTINS

NLRs- NOD LIKE RECEPTORS

DAMPs- DAMAGE ASSOCIATED MOLECULAR PATTERNS

CDAMPs- CELL DEATH ASSOCIATED MOLECULAR PATTERNS

SAMPs- SELF ASSOCIATED MOLECULAR PATTERNS

GBPs -GLYCAN BINDING PROTEINS

SIGLECs- SIALIC ACID BINDIN IMMUNOGLOBULIN LIKE LECTINS

CMAH-CYTIDINE MONOPHOSPHO N-ACETYL NEURAMINIC ACID HYDOXYLASE

IG- IMMUNOGLOBULIN

ITIMs- IMMUNORECEPTOR TYROSIN-BASED INHIBITION MOTIFS

ITAMs- IMMUNORECEPTOR TYROSIN-BASED ACTIVATION MOTIFS

MAG-MYELIN ASSOCIATED GLYCOPROTEIN

CRD- CARBOHYDRATE RECOGNITION DOMAIN

BCR- B CELL RECEPTOR

NK- NATURAL KILLER CELLS

DCs- DENDRITIC CELLS

GBS- GROUP B *STREPTOCOCCUS*

HIV-1- HUMAN IMMUNODEFICIENCY VIRUS

3D- TRIDIMENSIONAL

ITC- ISOTHERMAL TITRATION CALORIMETRY

SPR- SURFACE PLASMON RESONANCE

ELISA- ENZYME-LINKED IMMUNOSORBENT ASSAY

NMR- NUCLEAR MAGNETIC RESONANCE

CRYO-EM- CRYOGENIC ELECTRON MICROSCOPY

NOESY- NUCLEAR OVERHAUSER SPECTROSCOPY

NOE- NUCLEAR OVERHAUSER EFFECT

ROESY-ROTATING FRAME NUCLEAR OVERHAUSER EFFECT

COSY- CORRELATION SPECTROSCOPY

TOCSY- TOTAL CORRELATION SPECTROSCOPY

HSQC- HETERONUCLEAR SINGLE-QUANTUM CORRELATION SPEC-TROSCOPY

HMBC-HETERONUCLEAR MULTIPLE BOND CORRELATION SPECTROSCOPY

DOSY- DIFFUSION ORDERED SPECTROSCOPY

STD NMR- SATURATION TRANSFER DIFFERENCE

Tr-NOESY- TRANSFERRED NOESY

MM- MOLECULAR MECHANICS

MD-MOLECULAR DYNAMICS

MC- MONTE CARLO

EA- EVOLUTIONARY ALGORITHMS

GA- GENETIC ALGORITHMS

LS- LOCAL SEARCH

FF- FORCE FIELD

KB- KNOWLEDGE BASED

QM- QUANTUM MECHANICS

GB- GENERALIZED BORN

PB-POISSON BOLTZMANN

PBC- PERIODIC BOUNDARY CONDITIONS

PME- PARTICLE MESH EWALD

PDB- PROTEIN DATA BANK

NCBI- NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

BLAST- BASIC LOCAL ALIGNMENT SEARCH TOOL

CORCEMA-ST-COMPLETE RELAXATION AND CONFORMATIONAL EXCHANGE MATRIX ANALYSIS OF SATURATION TRANSFER

MGL-MACROPHAGE GALACTOSE TYPE LECTIN

MuV-HN- MUMPS VIRUS HEMAGGLUTININ NEURAMINIDASE

PAA- POLYACRYLAMIDE

RMSF- ROOT MEAN SQUARE FLUCTUATION

RMSD- ROOT MEAN SQUARE DEVIATION

IL-10- INTERLEUKINE-10

ORF- OPEN READING FRAME

TNFα- TUMOR NECROSIS FACTOR ALPHA

MD-2- MYELOID DIFFERENTIATION FACTOR 2

OPS- O-POLISACCHARIDE

LOS- LIPOOLIGOSACCHARIDE

ECD- EXTRACELLULAR DOMAIN

MR- MANNOSE RECEPTOR

WHO- WORLD HEALTH ORGANIZATION

MUV- MUMPS VIRUS

Sugar Abbreviations

GlcNAc- N-acetyl glucosamine

Fuc- Fucose

Glu-Glucose

Gal-Galactose

Sia-Sialic acid

Neu5Ac- N-acetyl neuraminic acid

Neu5Gc- N-glycolyl neuraminic acid

FuNAcN-2-acetamido-4-amino-2,4,6-trideoxy-a-galactopyranose

Kdo- 3-deoxy-D-manno-oct-2-ulopyranosonic acid

GalNAc-N- acetyl galactosamine

6'SLN- 6' sialyl N-acetyl lactosamine

3'SLN-3' sialyl N- acetyl lactosamine

ABSTRACT

Glycans and complementary glycan-binding proteins represent essential components in the control of both innate and adaptive immunity. Sialic acids are a family of sugars found on the terminal end of mammalian glycoconjugates; they able to act as marker of self in the immune system, as such residues are absent in most microbes.

Sialic acid-binding immunoglobulin-like lectins, or Siglecs, are cell surface receptors that recognize sialic acids and are known to modulate immune responses, influencing almost every cell in the hematopoietic system. Siglecs are involved in events like cell adhesion and signalling, inhibition or regulation of the immune cell activation, all mediated by the interaction with sialylated ligands. Sialic acid-Siglec interactions have been associated with a broad spectrum of diseases, stretching from autoimmunity to neurodegeneration and cancer. Thus, strategies for a rational modulation of the interactions between Siglecs and sialylated glycans in pathophysiological processes exhibit a great therapeutic potential.

In this context, the present thesis project aimed at the study of the interaction between Siglecs and their cognate sialic acid containing ligands, to disclose the key recognition events underlining host immune suppression or activation. To this end, a multidisciplinary approach combining advanced technologies as ligand-based NMR techniques, including STD NMR and tr-NOESY, biophysical binding assays and computational methodologies, such as homology modelling docking and MD simulations, was applied to provide an atomistic depiction of the interaction interfaces between various sialoglycans and their receptors.

The described strategy has been employed to characterize the binding features of several receptors of the Siglecs family, namely CD22/Siglec-2, Siglec-10 and Siglec-7. CD22 is a B-cell surface inhibitory protein capable of selectively $\alpha(2,6)$

linked sialylated glycans, thus dampening autoimmune responses against self-antigens. The characterization of complex-type N-glycans by CD22 allowed to describe the conformational behavior of the flexible ligands; the formation of CD22 homo-oligomers on the B-cell surface was also addressed. Furthermore, it was provided a global vision of how the most diffuse neuraminic acid forms of sialylated N-glycans are accommodated in the binding pocket of CD22. Moreover, the elucidation of the binding epitope of a synthetic sialo-mimetic upon CD22 interaction afforded new hints for the design and synthesis of high-affinity ligands of therapeutic relevance against B-cell derived malignancies.

Then, the Siglec-10, an inhibitory receptor that recognize α(2,3) and α(2,6) -linked sialoglycans was studied, thus providing the first insights of the molecular mechanisms regulating the interaction between Siglec-10 and naturally occurring sialoglycans.

After that, Siglec-7, a well-established inhibitory receptor that is primarily located on natural killer where it acts as inhibitor of cancer cells cytotoxicity *via* sialylated ligands binding, has been characterized in the interplay with the oncogenic pathogen *F. nucleatum*. Indeed, the presence of sialylated lipopolysaccharides on certain *F. nucleatum* strains, hinted that it may have a significant role at the immune interface. The interaction between Siglec-7 and the O-polysaccharide chain from the LPS of *F. nucleatum* 10953 strain has been depicted, thus delineating a structural binding model that might contribute to explain the etiological role of *F. nucleatum* in carcinogenesis.

A similar approach was employed to other sialoglycan- related systems, i. e. to dissect the mechanism of sialic acid recognition and hydrolysis by mumps virus hemagglutinin neuraminidase, a viral glycoprotein that plays key roles in virus entry and infection; and to assess the binding of the human macrophage galactose-type lectin in the interplay with lipooligosaccharide of *E. coli* strain R1.

In conclusion, the structural and functional characterization of Siglec- sialylated glycans interaction have allowed the analysis, at a molecular level, of the crucial features of the receptors/glycans interfaces, highlighting the molecular determinants involved in recognition and binding events, that will aid for the development or optimization of molecules for the therapeutic targeting of Siglecs.

SECTION I-INTRODUCTION

Chapter 1: The importance of glycocode based cross- talk

1.1 The sugar code

Glycans are nowadays recognized as major actors in molecular recognition processes fundamental for development, maintenance, and survival of all living organisms. Their degree of diversity and heterogeneity underlies multivarious biological functions ranging from structural, metabolic, and regulatory purposes.[1,2] Among the four building blocks of life (proteins, nucleic acids, lipids and glycans), glycans represent the most abundant and information rich macromolecules; these features are ascribable to their wide structural diversification that relies on several factors, including:

i) The chemical variability of the constituent monosaccharide units, in terms of configurational and constitutional isomerism and ring size, that give rise to many possible configurations and conformations.

ii) The various regio- and stereo- chemistries by which the glycosidic bonds can be arranged (α or β glycosidic linkages to multiples hydroxyls of the other monosaccharide). This also results in the possibility to form either linear or branched structures.

iii) The micro-heterogeneity, as glycans may present several modifications, such as methylation, esterification, phosphorylation, deoxygenation, acetylation.[3]

iv) The flexibility of the glycosidic linkages, that confers rather distinct conformational and dynamic properties.[4]

Furthermore, another important aspect that adds extra complexity is the possibility for glycans to exist attached to proteins, peptides or lipids. As a matter of fact, one

of the most frequent post-translational modifications in proteins is the glycosylation, i.e. the process in which a defined saccharide portion is attached either to asparagine or to serine/ threonine protein residues, resulting in the formation of *N*- and *O*-glycans, respectively (Figure 1.1, a).[5] Proteoglycans (PG), are native proteins linked to glycosaminoglycans (GAGs) i.e. long and highly sulfated heteropolysaccharide chains such as heparin (Hp), chondroitin sulfate (CS) and related compounds.[6] On the other hand, glycolipids arise when a glycan moiety is linked to a lipid anchor, generally to ceramide (*glycosphingolipids*) or glycerol (*glyceroglycolipids*) chains (Figure 1.1, b).[7,8]

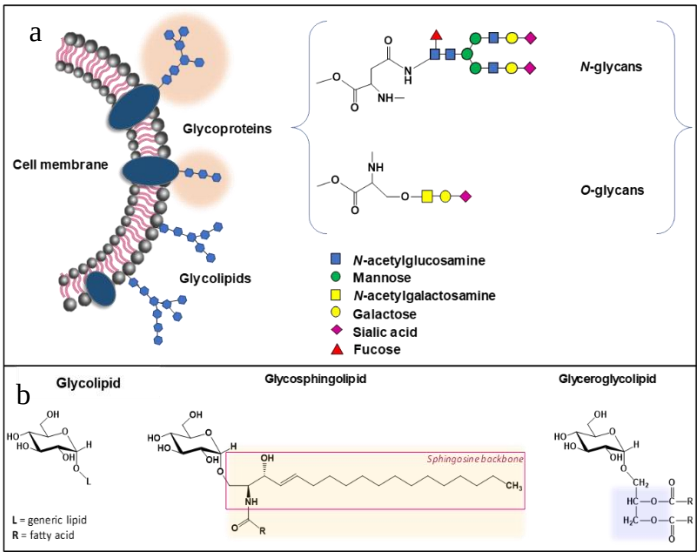


Figure 1.1 The heterogeneity of glycan structures. a) O- and N-linked glycans **b)** Glycolipids. Figure from ref. 9.

Glycans and glycoconjugates are the major components of the glycocalyx, a rich and diverse sugar layer that surrounds the surface of eukaryotic and prokaryotic cells (Figure 1.2).[10] The composition and structure of the glycocalyx is different among cell types and species;[11]the glycan profile of the whole cell arises from

the ordered and combined action of a set of glycosyltransferases and glycosidases throughout the endoplasmic reticulum (ER) and the Golgi apparatus of a given biological system.[12,13] The *glycome* uniqueness sets the molecular basis for recognition systems between different species; eukaryotic species across kingdoms are characterized by specific glyco-structures, such as *β-glucans* in fungi and *arabinogalactanproteins* (AGPS) in plants[14] (Figure 1.2, b and c). Noteworthy, peculiar glycan structures i. e. lipopolisaccharides (LPS), peptidoglycans and capsular polysaccharides, are exclusively expressed by bacteria and differ from those found in eukaryotes (Figure 1.2, b).[15,16]

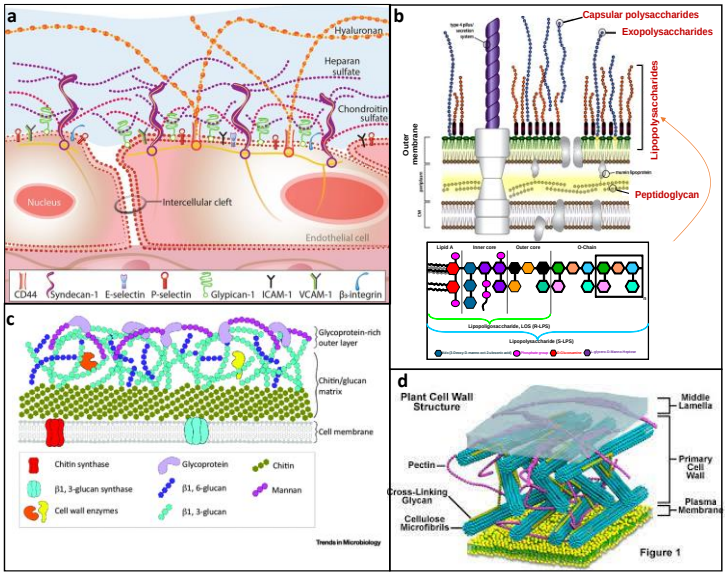


Figure 1.2 Structure the cell envelope of different organisms a) The endothelial glycocalyx (Figure rom ref. 17) **b)** Gram negative envelope (Figure from ref. 18) **c)** fungal (Figure from ref. 19) and **d)** plant cell wall (Figure from ref. 20).

The mammalian glycome encompasses a well-preserved repertoire of glycans that differs from phylogenetically different entities such as invertebrates, lower eukaryotes, and prokaryotes[21] (Figure 1.3). It comprises a plethora of structurally

diverse glycoforms (*N*-glycans/*O*-glycans, glycolipids) which derive from the assembly of nine monosaccharides building blocks through conserved biosynthetic routes across the secretory pathway; that also control the production and function of each structure type.[15]

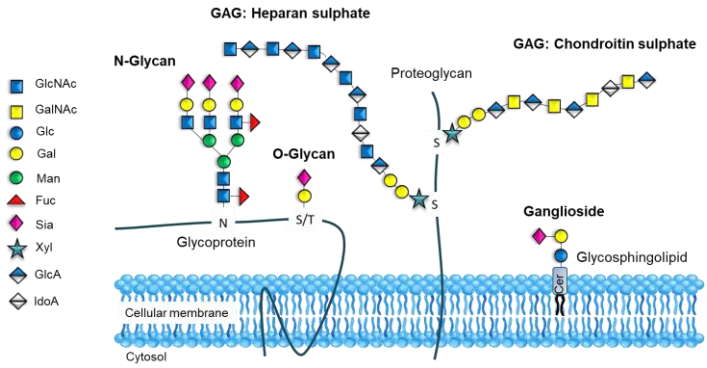


Figure 1.3 The mammalian glycome. Grafic representation of the major classes of structures found in the mammalian glycome (Figure adapted from ref. 22), sugars are indicated by Symbol Nomenclature for Glycans (SNFG) nomenclature.[23]

Being the foremost point of contact with the external environment, the cellular glycocalyx is the key biological interface responsible for the *intra*- and *inter*- species exchange of information and, by all means, represents a physical barrier against outside incursions.[24,25] The glycoconjugates spatial arrangement across the glycocalyx determines where and to which extent their activity takes place, whereas the structure of the glycans itself account for the affinity and selectivity in the recognition events.[26] Indeed, the above discussed glycans structural features permit, through the promotion or inhibition of *intra*- and *inter*- molecular binding processes, the control of homotypic and heterotypic interactions with other molecules.[15] In the next section an overview of the fundamental roles played by glycans alone and in the interplay with their specific receptors will be discussed.

1.2 Biological roles of glycans

It is very well established that glycans exhibit functional diversity which is essential to the growth, differentiation, and communication within complex organisms as well as for their crosstalk with other organisms in the external environment.[27]

The multivarious functions performed by glycans comprise i) structural, organizational and protective roles, ii) attachment points for microorganisms to infect, iii) defense or lure, iv) modulation of protein functions in a glycosylation-dependent manner (Figure 1.4).[28] Even though bacterial glycans are significantly involved in virulence and pathogenicity of a plethora of microbes, it has emerged that many gut commensal bacteria bacterial proteins are glycosylated. Moreover, the importance in fundamental processes key for gut homeostasis as provision of nutrients and serving as adhesion sites thus contributing to commensal colonization has been pointed out.[29,30] Here, classification in three broad categories based on structural and modulatory roles, molecular mimicry and recognition by glycan binding receptors is considered, as proposed by Varki and Gagneux.[25]

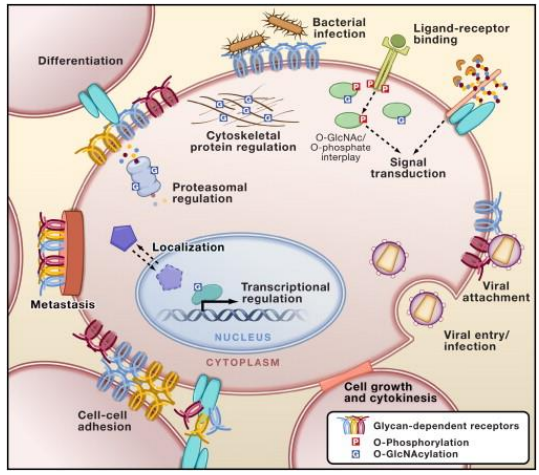


Figure 1.4 Schematic depiction of the main function of glycans (Figure from ref. 25).

1.2.1 Structural and modulatory roles

Concerning the structural and modulatory roles, glycans are critical elements to protein folding and stability as well as for the preservation of tissues supramolecular organization and physiochemical properties.[31]

Extracellular matrix (EM) proteoglycans support to endure high pressure through the formation of high viscosity complexes with hyaluronan[32] . Moreover, due to the hydrophilic character, peptidoglycans at the EM retain water molecules and maintain osmotic pressure.[33] Along with the structural functions, PGs perform important signalling and transport tasks within the matrix, through the binding of positively charged ions and signal molecules.[1]

The importance of the N-glycosylation of proteins for the maintenance of their functionality has been widely demonstrated[34,35] Glycosylation is essential for quality control of native secretory proteins as specific glycoforms reflect their folding status. Indeed, certain *N*-glycan motifs act as guides to proper folding, serving as binders for chaperones and agents aiding the achievement of the proper functional conformation.[36] On other hand, peculiar oligosaccharide structures are markers of a misfolded state and act as signals for protein ubiquitination.[37] Protein glycosylation could also perform a defensive role, as in the case of the heavily *O*-glycosylated mucins that represents the main components of mucus. Rich in negative charges for the presence of huge clusters of *O*-linked sialic acid glycans, they bind a large quantity of water molecules and form a mucus membrane that physically shield the epithelial surface cells from adverse environmental factors such as pathogens attachment.[1] Mucins contain GalNAc *O*-glycan moieties that function to protect the peptide portion from degradation by proteases.[38]

Moreover, mucin glycosylation is associated to the occurrence intestinal inflammatory diseases. Indeed, aberrant expressions of mucin glycans induced by

gut microbiome alterations for example as effect of poor alimentary habits affects the occurrence and progression of gut inflammatory diseases. However, further studies are needed to establish which dietary components affects gut mucin glycosylation.[39]

Glycans modulatory roles are related to the regulation of protein- protein interactions and comprehend cell-adhesion, intra- and inter-cellular molecular trafficking and mediation of cell–matrix /cell–cell interactions and also tissue development.[1,15]These roles rely on the glycosylation ability to induce functional variety to an otherwise fixed set of primary receptor–ligand interactions. Nevertheless, this tuning capability significantly shapes the resulting biological effects. A remarkable example is the differential modulation of the natural cell adhesion molecule (NCAM), a receptor involved in the initiation of myoblast fusion in mammalians brain.[40] The adhesion neuronal cell binding process is modulated by definite glycosylation of NCAM muscle specific domain. More in detail, NCAM-mediated homophilic adhesion, that is the attachment to the same protein on the adjacent cell, is promoted by mucin- like *N*-acetylglucosamine (GlcNAc) O-glycosylation whereas it is inhibited by the polysialylation.[41]

Different types of glycans, precisely *O*-fucose (Fuc), *O*-GlcNAc and *O*-glucose (Glc) glycosylations are key regulators of distinct aspects of the complex Notch signalling pathway thereby participating to embryonal growth as well as the preservation of tissues homeostasis. For example, *O*-Glc glycosylation of Notch receptors has been proved to indirectly promote trafficking to the cell surface by causing a conformational change on the glycosylated receptors that allowed for the subsequent cleavage by specific protases. Conversely, many studies demonstrated that *O*-Fuc glycosylation on Notch affected the strength of signalling through a direct interaction with Notch ligands. [42]

1.2.2 Molecular mimicry of host glycans

The bacterial glycome spreads across the entire cellular system, from cytoplasm to cell envelope and is involved in colonization, symbiosis, immune evasion, adaptation, and biofilm formation processes, also accounting for the microorganism pathogenicity.

Bacterial monosaccharides display larger complexity with respect to mammals, comprising over 700 different monosaccharide building blocks, and such variability often stems from evolutionary selective pressures (Figure 1.5).[15,43] The huge variability of bacterial glycome is facilitated by the plasticity of the sugar building blocks that can quickly undergo to modifications, without necessarily needing genetic mutations, even on the sole basis of environmental changes.[44,45]

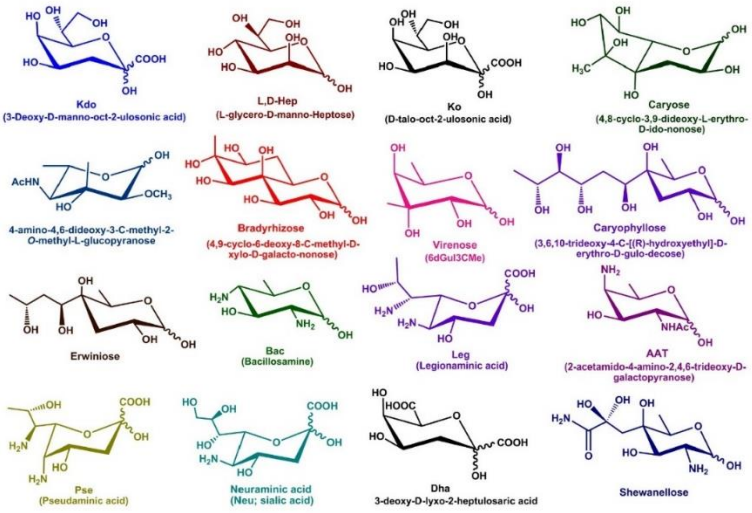


Figure 1.5 Representative examples of peculiar bacterial monosaccharides (Figure from ref. 15).

The innate host immune system detects conserved glycan structures, such as bacterial glycoconjugates that are known as microbe-associated molecular patterns (MAMPs),[46] or, likewise, as pathogens molecular patterns (PAMPs),

through Pattern Recognition Receptors (PRRs).[47] The PRRs are indeed responsible of the rapid and immediate pathogens recognition and subsequent activation the innate non-specific response. Toll-like receptors (TLRs),[48,49] C-type lectins (CTLs)[50] and NOD-like receptors (NLRs) belong to PRRs category.[51,52]

Nonetheless, glycans released from damaged and dying cells, as hyaluronan, are danger-associated molecular patterns (DAMPs)[53,54] and cell death-associated molecular patterns (CDAMPs), and thus can be recognized by PRRs, thereby participating to inflammatory diseases, apoptosis, autophagy and necrosis.[55]

The interaction of PRRs with PAMPs, DAMPs and CDAMPs works through a similar pathway by which a specific recognition event, such as bacterial entry, triggers the activation of intracellular signals that consequently promotes the synthesis of proinflammatory molecules thus stimulating adaptive immunity.[47]

On the other hand, certain host inhibitory receptors can also sense precise endogenous glycans, thus acting as self-associated molecular patterns (SAMPs) when is necessary to limit excessive inflammatory responses or to prevent of autoimmune reactions, or even to maintain a non -inflammatory state.[56]

As reaction to the above -described host-microbe interplay, many pathogens have developed several ways to mimic host glycans to escape immune surveillance or to establish tolerance, like in the case of microbial symbionts.[1] Molecular mimicry often involves sialic acid, usually found on the terminal moiety of vertebrates glycans (see next sections for further details) and mostly occur by direct or indirect acquisition of host glycans (as in the case of detrimental bacteria like Streptococcus pneumoniae and beneficial like the ones of the gut microbiota[57,58], convergent evolution toward host glycans biosynthetic routes (i.e. *Campylobacter jejuni, P. aeruginosa,* group B Streptococcus, *Haemophilus influenzae*) and lateral gene transfer *(Trichomonas vaginalis)*.[1, 59]

1.2.3 Recognition by glycan binding receptors

The modulation of the molecular interactions between glycans and their receptors is a key element for the regulation of multiple cellular mechanisms which are relevant to both health and disease.[1,60,61 ,62]

Glycan-binding proteins (GBPs) are responsible for the specific recognition of glycan epitopes. A variety of proteins such as lectins, toxins, microbial adhesins, antibodies and enzymes are included in this group, although GPBs can be further distinguished in extrinsic and intrinsic receptors depending on the capability to recognize glycans from same or from different organism.[1]Each GPB is characterized by a distinct glycan recognition domain that confers the selectivity for glycan binding and which is associated to other domains that translate the binding event to the functional biological response.[60]

Intrinsic GPBs are involved in direct interactions between cells and binding of glycans within the same cell. Various GPBs such as Galectins, Selectins and Siglecs (Sialic acid–binding immunoglobulin-like lectins) cover the surface of the immune system cells (Figure 1.6) thus performing crucial roles in immunity; accounting for inflammation and immunomodulation processes by prompting intracellular signalling and/or binding to cell surface targets to trigger a specific response.[1, 63]

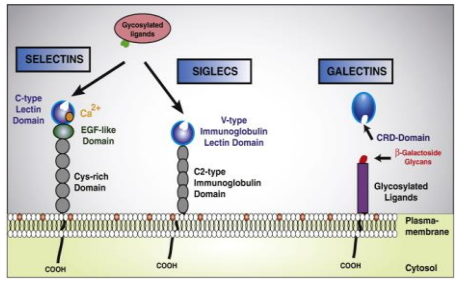


Figure 1.6. Schematic representation of different families of lectins found within the immune system From the left: Selectins (Sia recognition), Siglec (Sia recognition), Galectins (β-Gal recognition). Figure from ref. 64

Some significant examples are i) fine tuning of adaptive immunity by Galectin-3 via crosslink by the binding of *N*-linked galactose (Gal) and GlcNAc motif on the T cell receptor[65] ii) engagement of leukocyte surface glycans by endothelial E- and P-Selectins to initiate the inflammation response[66] iii) clearence of sialylated pathogens by macrophage Siglec-1(Sialoadhesin).[67,68]

As previously mentioned, many host glycans are specifically targeted by extrinsic GPBs from viruses, bacteria and parasites, thus serving as receptors by microbes to attach, invade and colonize host cells. Many viruses can bind host glycans, exploiting them for adherence and invasion. A notable example is the interplay of Influenza A virus hemagglutinin and neuraminidase proteins that, through the binding and cleavage of host cell sialylated glycans, are accountable for viral attachment and entry, respectively.[69] Likewise, several bacteria are capable to express glycan binding adhesion factors on their pili structures that support host cell attachment and colonization. *Helicobacter pylori* can indeed adhere to sialylated glycans produced in chronic inflammation (Sialyl Lewis^B and Sialyl Lewis^X) upon the expression of specific receptors, namely BabA, SabA, and LabA.[70,71]

1.3 Siglecs: a family of lectins for the specific recognition of sialic acid

Sialic acids (Sias), or neuraminic acids, are a peculiar family of nonulosonic sugars that are usually attached at the terminal moieties of *N*-glycans, *O*-glycans and glycosphingolipids.[72] Sialic acids are widely diverse, encompassing over 50 natural derivatives bearing different kind of modifications of the sugar skeletons at various positions (Figure 1.7).[73,74] The two major forms sialic acids in mammals are the *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) (Figure 1.7).[75,76,77]

The latter is biosynthesized by the enzymatic addition of a hydroxyl group to *N*-acetyl moiety at 5'-position of Neu5Ac, catalyzed by the hydroxylase/monooxigenase enzyme, namely cytidine monophospho-*N*-acetyl neuraminic acid hydroxylase (CMAH).[78] However, in contrast to mouse or great apes like chimpanzee,[79] the specific loss of Neu5Gc expression in humans is ascribed to a inactivating genomic mutation in CMAH that leads to the gene loss in hominin lineage (Figure 1.7). Despite the inability to produce Neu5Gc, regardless of the CMAH loss, Neu5Gc can be exogenously introduced from specific dietary sources, i.e. from red meat and cow's milk, and then metabolically incorporated following the Neu5Ac biochemical pathway.[78] Different studies reported the presence of Neu5Gc particularly on fetal tissues and on tumours, including melanoma, retinoblastoma, colon cancer and breast cancer.[80,81] Low levels of Neu5Gc were also found on the surfaces of human secretory epithelia and small- and large-blood vessels endothelia.[82] Consequently, Neu5Gc can be found on human cell surfaces being a pioneering example of "xeno-autoantigen"; anti-Neu5Gc antibodies detected in humans contribute to establish "xenosialitis", chronical inflammation state in Neu5Gc enriched tissues that can significantly impact cancer progression, increasing tumor-associated inflammation.[83]

As Siglecs are considered effective glyco-immuno checkpoints that could be employed for cancer immunotherapy,[84] understanding the basis of Neu5Gc-Siglec interaction appear of fundamental importance for the potential impact on the prevention and treatment of different diseases. Indeed, changes in the Neu5Gc/Neu5Ac ratio can potentially modulate binding and signalling properties of Siglecs receptors.

Due to their strategical exposition, structural diversity and ubiquitous distribution, sialic acids act as modulators of essential biological functions,[85,86,87]

controlling and balancing a wide array of essential immune responses upon the interaction with intrinsic and extrinsic GBPs.

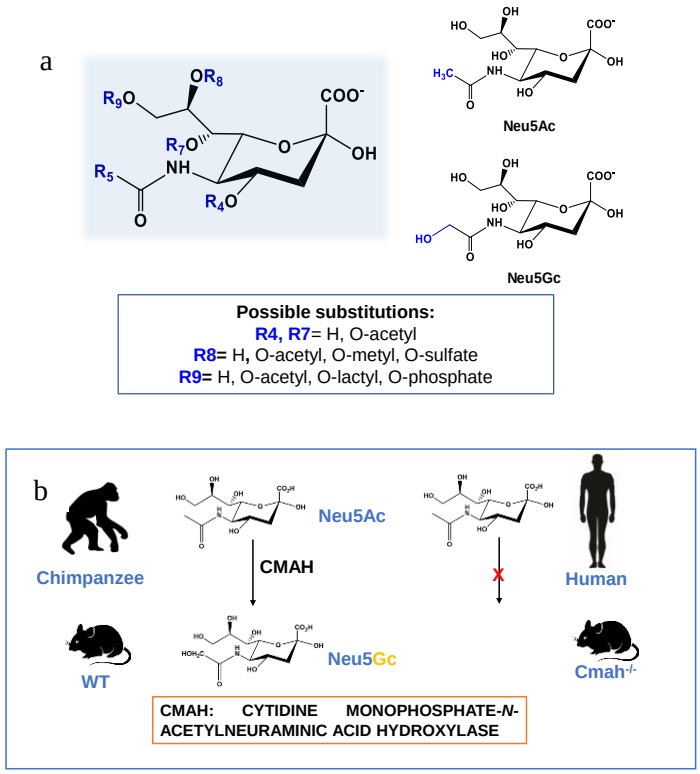


Figure 1.7 a) Sialic acid family. The nine-carbon backbone of sialic acid can be modified by several substituent groups (R) in different positions. On the right, the most diffuse sialic acids in mammals, N-acetylneuraminic acid and N-glycolylneuraminic acid. R4-R9: hydrogens, O-acetyl groups and, in some cases, O-methyl, O-sulfate, O-lactyl and O-phosphate groups. **b)** Loss of Neu5Gc in humans.

Siglecs are GBSs that specifically bind sialic acid on glycoconjugates and primarily present on the surface of innate and adaptative immune system cells. Siglecs share high sequence similarity, broadly similar domain organization and sialic acid-binding properties but possess diverse cell type distribution, giving rise to an elaborate and, to some degree, redundant pattern of expression within the

hematopoietic cells.[88,89,90] As Siglecs possess the characteristic immunoglobulin (Ig)- like topology,[89,90] they belong to the immunoglobulin superfamily. Moreover, they represent the prototype of the I-type lectins family as well as the major members,[89]that include structurally related receptors that share the capability to recognize Sia's on sialoglycans through specific binding domains.[88,91]

Nowadays, 15 human (chronological numeration) and 9 mouse Siglecs have been discovered (Figure 1.8).[92,93] Human Siglecs are categorized considering the sequence homology and conservation across orthologs[94] in "evolutionary conserved" and "CD33/Siglec-3 related" sub-categories. In particular, Siglecs -1, -2, -4 and -15, which displays low sequence similarity but high degree of gene preservation in mammalian species, belong to the evolutionary conserved group.[95] On the other hand, Siglecs with high homology (roughly above 50%) to Siglec-3 and among each other are clustered into the CD33 related group that include Siglecs 5 to 16, besides Siglec-3.[88,96]

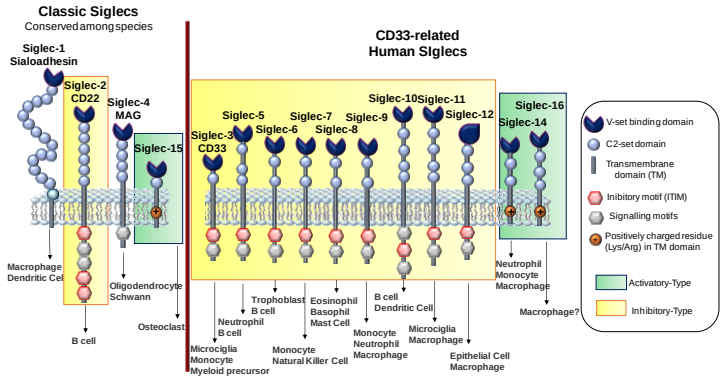


Figure 1.8 Human Siglecs. Human Siglecs are numbered chronologically. The Siglecs family consists of a number of highly evolutionary conserved proteins, including Sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG (Siglec-4) and Siglec-15, and a group of rapidly evolving Siglecs designated as CD33 (Siglec-3)-related Siglecs.[30] Human Siglecs are typically composed of a number of extracellular immunoglobulin domains, a transmembrane region, and most of them possess a number of intracellular motifs. Inhibitory-type and

activatory-type Siglecs are noted. The cellular distribution of each Siglec is indicated below by arrows.

From a structural viewpoint, Siglecs are constituted by extracellular N-terminal V-set Ig-like domain responsible for binding sialylated ligands, that is linked by a disulphide bridge to a variable number of C2-set (Ig)- like domains, (1 to 16 units, varying across different Siglecs).[63,88,93] The extracellular portion is followed by single pass transmembrane domain connecting to a variable number of cytosolic tails, that strongly influences Siglecs biological functions.

Indeed, concerning the cytoplasmic tails, most of Siglecs feature one or more tyrosine-based signalling motifs, mainly immunoreceptor tyrosine-based inhibition motifs (ITIMs),[97] that confer the ability to inhibit immune cell stimulation throughout binding and activation of SH-2 domain- containing protein tyrosine phosphatases as SHP-1 and SHP-2 (Figure 1.9).[90,98,99,100]

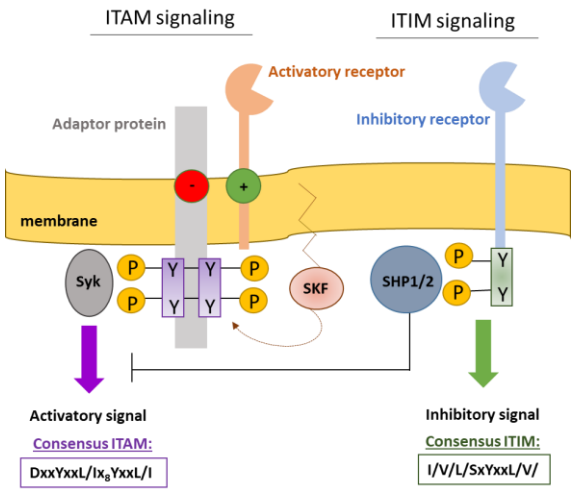


Figure 1.9 ITAM-/ITIM signalling. Left side: Activatory Siglecs associate with ITAM containing adaptor proteins such as DAP12 through interactions between charged amino acids within the proteins transmembrane regions. Then, Src kinase enzyme family (SKF) phosphorylate tyrosine residues of ITAMs. Phosphotyrosine residues bind to Syk protein kinases that mediate cellular activation via downstream cascades. Right side: Upon ligand

binding, inhibitory receptors recruit SHP1 and SHP2 which can in turn terminate intracellular signals emanating from ITAM receptors.[101]

Conversely, few Siglecs, namely Siglec-14, Siglec-15 and Siglec-16, display activating signalling properties owing to the co- association with immunoreceptor tyrosine-based activation motifs (ITAMs) containing adapter proteins such as DAP12.[102] The co- association is due to the presence of a positively charged residue in the Siglec transmembrane region which interacts with negatively charged residue in the adapter properties. This event initiates the ITAM tyrosine signal leading to the recruitment and activation of the Syk family tyrosine kinases, consequently triggering immune response (Figure 1.9).[88,90,93]

Furthermore, a third kind of Siglecs, particularly Siglec-1 and Siglec-4 or myelin associated glycoprotein (MAG), lack signalling motifs and possess neutral transmembrane domains; thus, they do not have signalling properties and only perform roles involving Sia recognition such as adherence.[91,103,104]

Siglecs and sialic acid interactions regulates the immune signalling thereby contributing to fundamental mechanisms in cell signalling, as well as cell-cell and host-cell interactions. In the next paragraph the molecular recognition of Siglecs will be discussed in depth.

1.3.1 The binding properties of Siglecs

Siglecs exhibits different preference for binding sialylated structures found on mammalian cells, ascribable to the type of sugar residues and to the linkage between Sia unit and the adjoining sugar moiety.[88,105]

The Siglec binding specificity depends on the properties of the N-terminal V-set carbohydrate recognition domain (CRD) encompassing the sialic acid binding region, in which the Sia and adjacent residues are accommodated (Figure 1.10).[106]

The V-set domain is composed by nine β strands (named A-G) arranged with a characteristic topology. The Sia binding region is located specifically on the F strand, where an arginine residue in conserved position represents the main determinant for Sia binding by Siglecs, interacting with the carboxylate group of Sia thus forming a strong salt bridge (Figure 1.10).[94] Another remarkable structural feature is the presence of an intra-sheet disulfide bond between the B and E β-strands, which increases separation between the β-sheets and allows a greater exposure of two aromatic residues in invariant positions (but of different nature among Siglecs) on the A and G strands, which establish interactions with the protons of the lateral glycerol chain and N-acetyl group of Sia.[63,93]

Nonetheless, Siglec recognition region is also defined by the CC' loop that outlines the F-G strands of the binding site. Indeed, the loop displays a variable sequence among Siglecs and plays a significant role into the specificity of recognition of sialylated ligands, particularly for the recognition of longer glycan chains.[63] Notably, it has been demonstrated that the CC' loop exhibits a conformational change when interacting with sialylated ligands, as demonstrated for the recognition GT1b ganglioside and Siglec-7.[107]

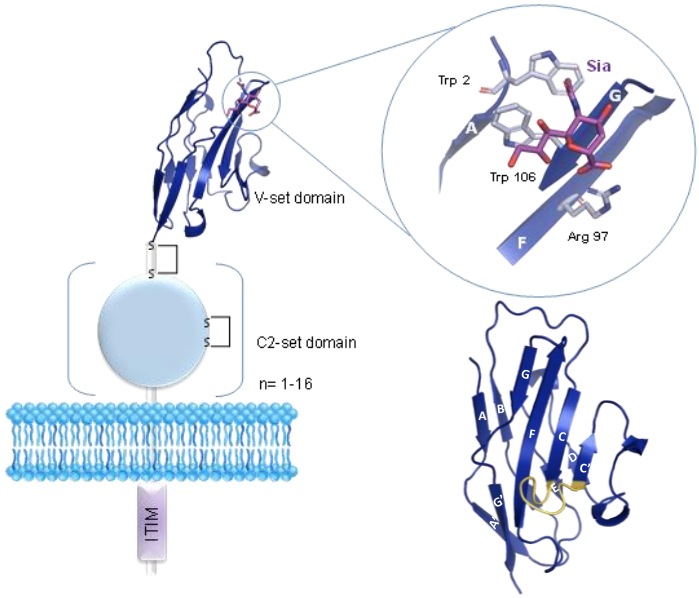


Figure 1.10 Structural features of Siglecs V-set domain. Details of Sialic acid binding by Sialoadhesin/Siglec-1 (PDB ID: 1QFO). The conserved arginine (Arg97) located of the F strand is the critical determinant of Sia recognition by Siglecs. The conserved intra-sheet disulphide in Sialoadhesin widens the Ig β-sandwich resulting in the exposure of two tryptophan residues (Trp2 and Trp106) on the A and G strands which establishes hydrophobic contacts with the N-acetyl and glycerol side chains of Sia. The variable CC' loop is highlighted in yellow.

1.3.2 Recognition between Siglecs and sialoglycans

Siglecs recognize sialylated glycans exposed on the surface of the same cell (*cis* interactions) or of different cells (*trans* interactions), as depicted in Figure 1.11.[108]

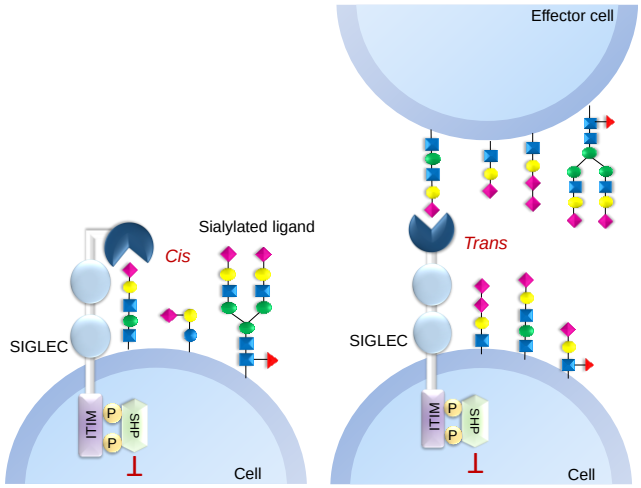


Figure 1.11 Siglecs recognition mechanisms (left) *Cis* interactions with sialoglycans exposed on the same cell surface. (right) *Trans* interactions with sialylated ligands on different cells.

A dynamic equilibrium between the two mechanism occurs as Siglecs can be engaged both by *cis* and *trans* ligands, depending on ligand affinity and availability. Due to the high local concentration of sialylated glycans on immune cell surfaces, Siglecs' binding sites are generally masked by *cis* interactions, thus outcompeting the *trans* interactions.[93] Siglec-1 represents an exception, as it possesses an rather long and extended topology that hamper the establishment of *cis* interactions, as sialic acid-binding site is too far from the cellular membrane.[109] Furthermore, the action of membrane sialidases, cellular activation phenomena or plasma membrane domains reorganization, may ease the interaction with *trans* ligands as they "unmask" Siglecs.[94,110] Noteworthy, "self" endogenous ligands (*cis* interactions) may not harness the *trans* on adjacent cells because of the relatively low affinity to high avidity multivalent sialylated probes[111]or to higher-density *trans* ligands on sialylated microbes. As result, they could produce

stronger inhibitory signals with respect to those (mainly in *cis*) delivered by self-ligands, thereby converting from *cis* to *trans* interactions.[111]

Overall, the binding occurring between Siglecs and *cis*- ligands has a significant impact in the regulation of the cellular functions, hampering non-specific cell-cell interactions that may stimulate unneeded signalling.[96]This kind of cis signals produced by inhibitory Siglecs may promote or exclude the association with an activating receptor. A notable case is the association of CD22 (Siglec-2) to the B cell receptor (BCR) which causes the inhibition of the calcium dependent signalling pathway.[112] In resting state the interaction between sialylated glycans on CD22 produce clusters that sequester the receptor far from the BCR. When a specific antigen binds to BCR, CD22 oligomers are brought in its proximity and the subsequent association triggers the negative regulation of the BCR (Figure 1.12).[113] Conversely, it was observed that Siglec-G, murine ortholog of Siglec-10, perform its inhibitory action through the direct binding to the sialylated BCR domain upon the occurrence of antigen ligation.[113]

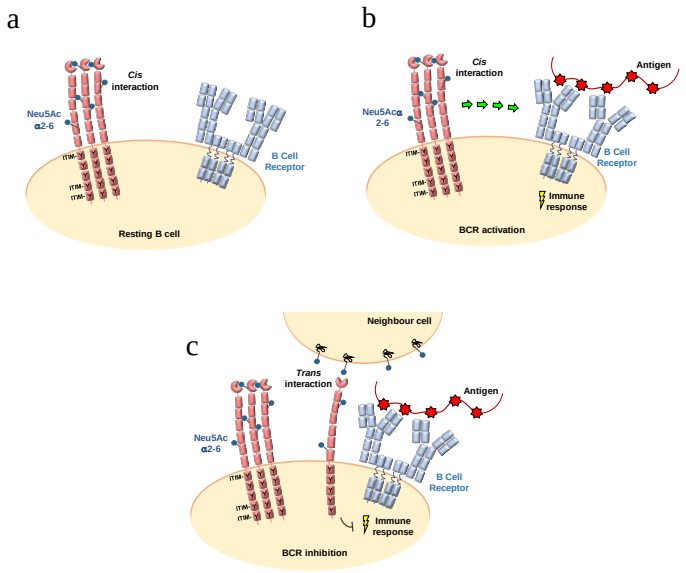


Figure 1.12 Immune response mechanism mediated by CD22 on B cells. **a)** formation of CD22 homo-oligomers on resting B cell upon cis interactions. **b)** CD22 recruitment upon BCR antigen ligation **c)** inhibition of immune response.

1.3.3 Siglecs' roles in the immune system

The function that each Siglec plays in the immune system depends on its localization, relative population, signalling properties, as well as on the binding specificity. Due to their remarkable regulatory properties, Siglec are associated to inflammatory, autoimmune, allergic and neurodegerative diseases that are consequence of aberrant interactions involving its associated sialylated ligands (Figure 1.13).[90]

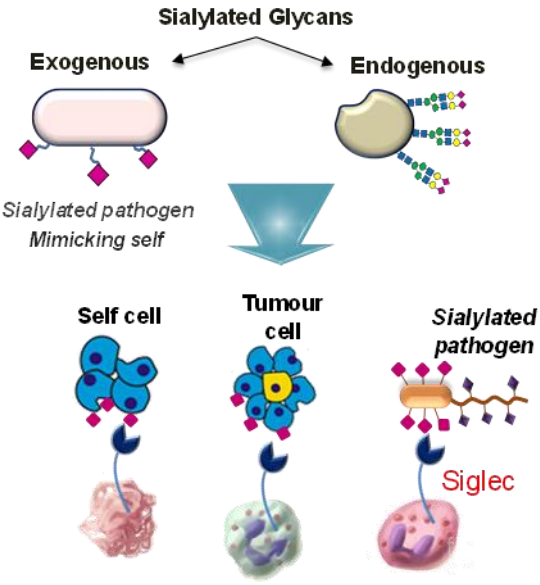


Figure 1.13 Schematic representation the interaction of Siglecs with exogenous and endogenous ligands.

Siglecs are essential actors in the "self" and "non self" discrimination for the ability to recognize specific sialylated glycoforms.[111] Indeed, Sia represent a marker of "self", namely SAMP (see above), and the interplay of inhibitory

Siglecs with sialylated glycans modulates the consequent down-regulation of cell signalling. Within the adaptive immune system, Siglecs contribute to the modulation of T-cell activation and polarization as well as balance B cells tolerance to antigens.[88,93,114] Concerning the innate immune system, Siglecs are modulators of natural killer (NKs) cells cytotoxicity and of programmed cell death in neutrophils and eosinophils.[99]

Siglecs are also involved in host- pathogen interplay, as dangerous human pathogens have evolved to exploit inhibitory receptors to elude pathogens the host immune response for the immune evasion.[93] As explained above, some pathogens can shield their envelope with SAMPs-like glycans, usually mimicking Sia, which are mistakenly recognized as "self" by host receptors, thus eluding immune response.[64] Many pathogenic bacteria including GBS (Group B *Streptococcus*), *Campylobacter jejuni*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Neisseria meningitides* and *gonhorreae* are coated with Sia containing structures synthesized or captured from the host[67,115] and can be potentially recognized as self-ligands by inhibitory Siglecs .[67,111,116] Noteworthy, *Streptococcus pneumoniae* can alter Siglec signalling pathway by producing sialidases that remove Siglec *cis* ligands[117]. Other microbes cover their cell envelope components with sialic acid analogues that are engaged by Siglecs. *C. jejuni* is a Gram-negative bacterium, spread worldwide, which causes bacterial gastroenteritis as well as the Guillain-Barré syndrome, an ensemble of post-infectious autoimmune neuropathies. *C. jejuni* activates the anti-inflammatory axis mediated by Siglec-10 on DC, through the recognition of sialic acid derivatives (pseudaminic acids) present on the pathogen flagellum and by this means stimulating the production of anti-inflammatory cytokines.[118]

Siglecs are also implicated in the development of viral infections. As a notable instance, the human immunodeficiency virus (HIV-1) contains sialylated

gangliosides on the viral envelope surface glycoproteins such as gp120, which are mainly engaged by Siglec-1 and Siglec-7 expressed on the surface of NKs, macrophages and DCs. Such interactions provoke or increase viral entry and are also important for viral proliferation and consequent illness aggravation.[119]

Within this frame, Siglec-dependent recognition of microbial and/or viral sialylated structures results in altered immune responses to the advantage of the pathogen, promoting successful host colonization. [120] Conversely, the presence of specific sialylated structures on microbes' surfaces may induce a protective response which prevents an excessive inflammation state in the host, as in the case of the recognition and clearance of sialylated pathogens by Siglec-1.[93]

It is nowadays recognized that aberrant glycosylation is related to cancer progression and metastatic behaviour for several cancerogenic cells. Consequently, Siglec-sialoglycan axis is by all means considered a glycoimmune checkpoint exploitable to enhance an antitumor immune response.[121] For example, the overexpression of α2,6 or α2,3 liked sialic acids represents the hallmark of some types of cancers[122] and it has been considered a potential target for the development of novel therapeutic approaches against cancer diseases.[123] The induction of antitumour immunity by the regulation of Siglec signalling using antibodies or glycomimetics has already been proven as a viable strategy to slow the growth of some tumours.[73,124]For instance, CD22 immune checkpoint inhibition by means of high affinity glycan ligands showed promising results for the treatment of various B cell related malignancies.[125]Other innovative approaches, such as functional targeting by means of CD22 agonists/antagonists or cell-based therapies, have been employed in preclinical and clinical trials.[125]

Interestingly, Siglec-7 and Siglec-9 ligands are overexpressed on certain carcinogenic cells and the *cis* engagement of their receptors[126,127] could directly modulate

the cytotoxicity of human NK cells and T cells activity; therefore, a strategic blockade of these interactions may possibly increase tumor cells death.

It was recently demonstrated that the interaction between Siglec-10, highly expressed by tumour associated macrophages, and the heat stable antigen CD24 is crucial for anti-tumor immunity regulation. Indeed, the suppression of CD24-Siglec-10 binding causes the phagocytosis of many CD24-expressing human tumors, thus representing a promising target for treatment of ovarian and breast cancer particularly.[128]

Being implicated in the inhibition of antigen-specific T cell responses, Siglec-15 represents a promising target for cancer immunotherapy. Indeed, the role of Siglec-15 as immune suppressor was recently highlighted, demonstrating its upregulation in tumor cells and tumor-infiltrating myeloid cells.[129]

On the contrary, sialic acid-Siglec interactions could be manipulated to provoke tolerance in an over-reactive immune system as in the case of allergic reactions and auto-immune diseases[68,130] by targeting inhibitory Siglecs in mononuclear phagocytes and eosinophils as an alternative to the current medical care. For example, Siglec-8, that is expressed on eosinophils and mast cells and is involved in allergic reactions, represents an appealing target to treat allergic diseases such as asthma, food intolerances and esophagus inflammation.[115,131]

Thus, the specificity and restricted cell-type expression make Siglecs ideal candidates for the design of immunotherapeutic agents against a broad spectrum of diseases. For these reasons, there is increasing interest in understanding the mechanisms of recognition and interaction of Siglecs with target substrates for the rational development of novel mimetics able to modulate the key Siglecs' signalling routes.

Chapter 2- Tools to study receptor- glycans interactions

2.1 Investigation of glycan-receptors complexes

As discussed in Chapter 1, many functions of glycoconjugates that are key in health and disease are mediated by binding proteins able to decode the information content of the glycome through the recognition of their cognate glycans[132,133,134,135]

Therefore, the explanation of the molecular basis underlying glycans recognition by GPBs, including Siglecs, represents a fundamental task toward the full comprehension of their biological functions. The atomic level knowledge of the structural details that drive the recognition and binding processes is fundamental to guide the synthesis and the optimization of glycomimetics exploiting these biologically relevant targets with diagnostic and therapeutic purposes.

The interactions between glycans and their receptors are characterized by rather low binding affinity.[136] Nonetheless, such relatively weak binding is enhanced by multivalent cell-surface interactions at the glycocalyx surface, where glycans reach millimolar concentrations.[137]The low affinity and the high degree of complexity of glycans complicates the detection and analysis of these systems *in-vitro*, therefore, the combination of a wide range of complementary techniques embodies the best strategy to effectively depict such interactions. Methods for the evaluation of the interaction affinity and/or thermodynamics including isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), enzyme-linked immunosorbent assays (ELISA), fluorescence-based assays, are usually implemented with structural biology data to provide the structure-activity relationships. Techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) and the emerging cryogenic electron microscopy (cryo-EM), allows for the

characterization of the three-dimensional structure and physicochemical properties of glycans receptor complexes.[15] However, many challenges are associated to the preparation of appropriate systems to analyze, such as difficult crystallization procedures and problems related to obtain isotopically labelled proteins. Computational techniques, namely homology modelling, Docking and MD simulations, may be employed to partially overcome these limitations and gather molecular insights of glycan- based interactions. The present chapter aims to describe the principal techniques that have been used for the thesis scopes, thus putting into focus ligand-based NMR techniques and molecular modelling methods.

2.2 NMR methods for the study of protein- ligand interactions

Nuclear Magnetic Resonance spectroscopy is an effective tool for the characterization of the structure, conformation, geometry and dynamic of carbohydrate ligands.[138,139,140]The following paragraphs will not get into the NMR fundamentals, but it will be focused on the main methodologies that allows to dissect glycans structure and behavior in solution as well as upon receptor biding.

2.2.1 NMR characterization of free state oligo and polysaccharides

The structural characterization of a glycan ligand underlies most of NMR interaction studies and is achieved through the combination of homo- and heteronuclear mono/ multi-dimensional NMR experiments. Indeed, a set of complementary experiments is necessary to assign all the spin systems, establish the nature and the location of non-carbohydrate substituents and the connectivity between monosaccharide units. The analysis of ^{1}H- NMR spectrum provides in first place the monosaccharide composition, obtained by the evaluation of NMR chemical

shifts (δ) and coupling constants (*J*). The characteristic ^{1}H and ^{13}C resonances of carbohydrates are enlisted in Table 2.1.

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

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

The integration of the signals corresponding to anomeric proton resonances, which are found in the shift range 5.8-4.4 ppm, delivers an initial evaluation of the number of different monosaccharide residues, in turn further confirmed by the analysis of the anomeric carbon resonances in $^{1}H^{13}C$-HSQC spectra. Moreover, the distinctive values of coupling constants $^{1}J_{C1,H1}$ and $^{3}J_{H1,H2}$ discriminates between the anomeric orientation of the sugar units. Indeed, for pyranose rings with *gluco* or *galacto* configuration (H-2 axial), a $^{3}J_{H1,H2}$ above 8 Hz is indicative of a β- configuration at the anomeric carbon and below 3 Hz of an α- configured anomer. Conversely, *manno* configured sugars (H-2 equatorial) are characterized by $^{3}J_{H1,H2}$ below 3 Hz. The anomeric configuration can be also established by the

magnitude of $^{1}J_{C1,H1}$, as a value under 165 Hz denotes the β-anomer whereas beyond 170 Hz indicates the α-anomeric configuration.

Several 2D spectra are required for the unambiguous identification of the sugar sequence, included non-carbohydrate substituents and conformational features. In detail, homonuclear through-bond experiments, such as COSY (Correlation Spectroscopy) and TOCSY (Total Correlation Spectroscopy) spectra, measure the correlations of protons within the same monosaccharide unit and do not provide details about the monosaccharide connection. Nuclear Overhauser Spectroscopy (NOESY) is a homonuclear through-space experiment which identifies spins undergoing cross-relaxation by measuring the cross-relaxation rates; consequently, the magnetization is transferred to proton that are spatially close. For intermediate sized molecules, as disaccharides and trisaccharides, the NOE intensity may be close to zero, and ROESY (Rotating frame Overhauser Effect Spectroscopy) experiments, in which homonuclear Nuclear Overhauser effects (NOEs) are measured under spin-locked conditions, can be used in place of NOESY. NOE correlations are very useful to confirm relative configuration of the sugar units and the configuration of the anomeric center. For example, in β-configured sugars like Glc, Gal, Man, H-1 gives strong *intra*- residue NOE correlation with H-3 and H-5, while in α-configured sugar units only correlates with H-2. Furthermore, by means of NOE spectra, the correlation between the anomeric proton and the protons of the adjacent sugar unit can be detected. Such *inter*-residual contacts are essential to obtain the sugar sequence in oligo/polisaccharides.

Further structural determinants can be deduced using heteronuclear 2D experiments as Heteronuclear Single-Quantum Correlation Spectroscopy (HSQC), which displays the correlations between carbons and the directly linked protons, and Heteronuclear Multiple Bond Correlation (HMBC) that gives signals between ^{13}C and ^{1}H nuclei that are correlated through several bonds. The

correlations that can be deduced from the analysis of 2D experiments are showed in Figure 2.1

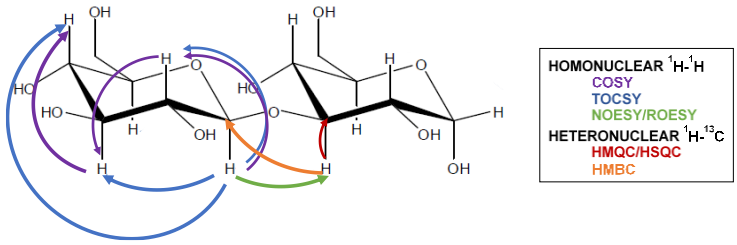


Figure 2.1 Representation of the correlations observable by different 2D NMR experiments.

Indeed, the presence of a non-carbohydrate substituent, such as methyl, acetyl or phosphate residue, often affects proton and carbon resonances where the group is located. For instence, in a HSQC spectrum the typical down-field shift of 4–10 ppm of the resonances of the carbons at the anomeric and linked position (glycosylation shift) can be observed. Also, the acylation sites can be easily locate analyzing the downfield shift of ring proton signals (acylation shift).[141]

2.2.2 NMR: the ligand-based approach

The past decade has witnessed the emergence of powerful and versatile NMR techniques for the study of proteins-glycoconjugates recognition events.[142] Indeed, multiple NMR-based methods are suitable for detection and characterization of binding processes between a small substrate and its large receptor. These experiments allow for the identification and characterization of the binding either by looking at the resonance signals of the ligand or of the protein, accurately depicting ligand-receptor complexes. Strong emphasis has been placed on ligand-based methods such as diffusion ordered spectroscopy (DOSY), and NOE based techniques, such as saturation transfer difference (STD NMR), NOE pumping,

waterLOGSY.[143] Indeed, such techniques are versatile due to their ability to examine weak binding interactions and for rapid screening the binding affinities of ligands with proteins at atomic resolution, thus allowing for complex characterization and for the determination of binding epitopes and bioactive conformations.[144]The attractiveness of these methods is also due to the wide applicability, few constraints regarding the size of the target protein, no need the isotopic labeling of none of the binding partners as well as low receptor amount required.[144] They are particularly useful in the medium– low affinity range; therefore, they found wide applicability within the study of transient protein glycan interactions.[145]

In general, ligand-based experiments are based on the variation of defined NMR properties parameters (chemical shift, scalar coupling, dipolar coupling, relaxation rate) of the bound state ligand with respect to its free state, which are observable for interacting systems in the fast exchange regime.

For a better comprehension of the fast exchange regime, it is worth to remark that the simplest binding process between a ligand and a protein is a dynamic exchange between free and bound state following a second order kinetic. Such equilibrium is defined by K_{on} and being the *on* (association) and *off* (dissociation) rate constants and where the dissociation constant K_D is the ratio K_{off}/K_{on}[146] (Figure 2.2).

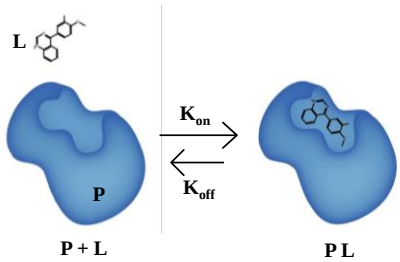


Figure 2.2 Scheme of a generic protein ligand binding process.

The binding process is a dynamic chemical exchange between free and bound state in which different conditions are possible depending on k_{off} values. Fast exchange processes are characterized by a K_D around 10^{-8} M, whereas strong exchange by K_D about 10^{-5}. From an NMR viewpoint, a chemical exchange can be defined as a process of nuclei exchanges between two or more environments as in the free to bound state transition, upon which a change of certain NMR parameters occurs, thus forming the basis of the NMR screening experiments (Figure 2.3). On chemical shift time scale (ω), the exchange can be fast (K_{off} » ω), intermediate (K_{off} ~ ω), or slow (K_{off} « ω), where $\Delta\omega = 2\pi\Delta\nu$, represents the chemical shift difference between the free and bound states, measured in Hz.[140]

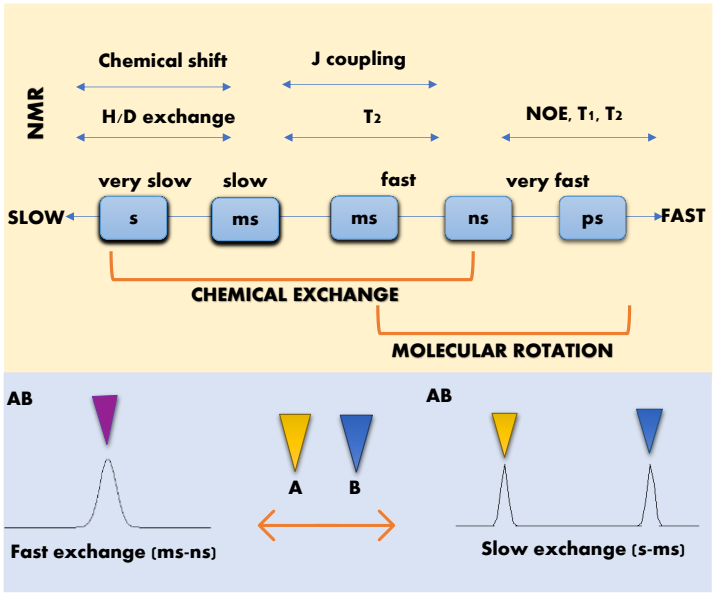


Figure 2.3 Schematic illustration of fast exchange vs slow exchange NMR timescales.

Considering the ligand in the fast exchange regime, the exchange is fast compared to chemical shift timescale, hence, the experimental signal corresponds to the average of the chemical shifts of ligand signals in the two states. Conversely, when the interaction between the ligand and the protein is strong, thus implying a

slow exchange, the ligand is no more detectable upon ligand-based approach experiments because the ligand behaves as part of the macromolecule. In general, most of ligand/ receptor interactions are assumed to be in fast exchange for two main reasons:

i) The fast exchange rates of the usually examined binding mechanisms are distinguished by $K_D \geq 100$ μM and the slow exchange rate K_{off} values range is $1000 < K_{off} < 100\ 000\ s^{-1}$

ii) Ligand-based NMR screening methods are mainly based on ^{1}H-resonancens and, consequently, the exchange constant exceeds the differences of intrinsic ^{1}H parameters, including relaxation rates and rotating frame precession frequencies.

For fast exchange systems in a ligand- based experiment, the observed NMR response of a ligand corresponds average of the NMR parameters of the free and bound states, weighted for the respective mole fractions:

$$M_{obs} = X_L M_L + XP_L MP_L \qquad \text{[Equation 2.1]}$$

where M_{obs} is any NMR parameters that characterize the equilibrium system; X_L and XP_L are the mole fractions of the free and bound ligand; and M_L and MP_L are the NMR property of the ligand in the free and bound state, respectively. Correspondingly, the individual populations of the ligands depend upon the K_D value and therefore on the protein ligand ratio. Hence, the resonances of the ligand in a ligand-receptor mixture can be observed if only the ligand is in excess with respect to the target.

In the free state, both receptor and ligand preserve their intrinsic NMR parameters. When interacting to each other, the mutual binding affinity drives an exchange process that modulates the NMR parameters of both molecules. As result, the

small ligand adopts characteristics of the much larger receptor, and the receptor interacting surface is in turn perturbed by the interplay with the substrate. In fast exchange, the frequency of free and bound state is sampled before exchange so solved and integrated signals may be expected for each status.[147]

The ligand-based approach NMR experiments that are treated in this thesis work, as mentioned above, are the tr-NOESY and the STD NMR, and their principles will be discussed in the next paragraphs.

2.2.3 NOE and transferred NOE

As discussed previously, NOE effects are extremely helpful in the determination of three -dimensional structure of molecules in solution for the relationship between NOE and nuclei spatial vicinity.[144]These effects are due to homonuclear dipolar interactions that decrease very fast with *inter*-nuclear separation, so they are detectable between nuclei below 5-6 Å distances. Furthermore, the intensity and the sign of NOE contacts is function molecular tumbling rates in solution, defined by the parameter $\omega_0\tau_c$., where τ_c is the correlation time and ω_0 is the Larmor precession frequency. As showed in Figure 2.4, small molecules possessing fast random rotation in solution show positive NOE contacts of maximum values 0.5, whereas molecules with a slower thumbling display negative NOE with a maximum value of -1.

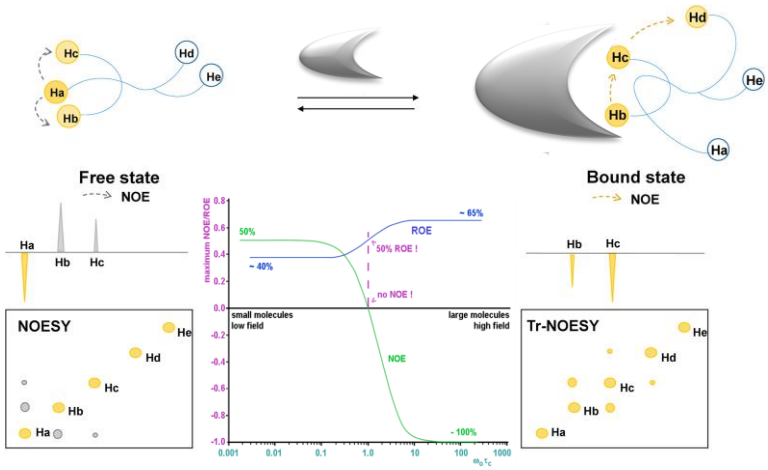


Figure 2.4 NOESY and tr-NOESY. In the NOESY spectrum (*left*) protons cross peaks correlated with Ha have changed sign respect to the diagonal. In the tr-NOESY (*right*) the same cross peaks have a negative sign. ω_0= Larmor precession frequency τc= correlation time. NOE and ROE curves are shown as functions of $\omega_0\tau_c$. Low molecular weight molecules with a low τ_c have a faster tumbling in solution and the NOE is positive (0.5 at maximum). High molecular weight molecules with a high τ_c have a slow tumbling in solution and the NOE is negative. When $\omega_0\tau_c = 1.115$, the NOE is zero. Figure adapted from ref.[148]

The entity of the correlation in solution depends on the magnitude of the correlation time and is mainly affected by the molecular weight of the molecule. The larger a molecule, the slower its re-orientation, therefore the longer its τ_c and vice versa. In ideal conditions, low molecular weight molecules show positive NOEs, while larger molecules display negative NOEs. However, the correlation time is also influenced by other parameters, such as temperature (the higher the temperature, the shorter τ_c), solvent viscosity (the more viscous, the longer τ_c) as well as aggregation in solution. For precise $\omega_0\tau_c$ value the NOE sign become zero, so the signals are not detectable although the nuclei are dipolar coupled. As mentioned above, ROE experiment allows to overcome this inconvenience. Indeed,

ROE signals display lower intensity respect to that of NOE ones, but their sign is always positive because of the different relationship to $\omega_0\tau_c$ (Figure 2.4).
To carry out a quantitative interpretation of steady state NOEs, the transient experiment must be performed. The transient NOEs are detected by 2D NOESY experiments recorded applying a sequence of high-frequency pulses which generate a non-equilibrium state and then returns to equilibrium by relaxation during the mixing time (T_m) (Figure 2.5).

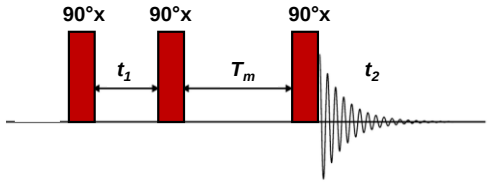


Figure 2.5 The NOESY experiment pulse sequence.

Transferred NOESY experiments result from the application of the above-described pulse sequence to a protein-ligand system in dynamic exchange, in which a ligand is present in excess. These experiments provide key information regarding the ligands' binding mode in the solution and the conformational features of the bound ligand (bioactive conformation). In general, *inter*- and *intra*-molecular transferred NOEs (tr-NOEs) are distinguished; *intra*-molecular tr-NOEs are observed within the ligand moieties and are relevant for the analysis of bound-ligand conformations. On the other hand, *inter*-molecular tr-NOEs occur between a ligand and a receptor protein, and therefore allow for the determination of the orientation of bound ligands in receptor binding pockets although their examination requires isotope-edited/isotope-edited experiments.[149]

The observation of *intra*-molecular tr-NOEs relies on different tumbling times of free and bound molecules. The ligand, freely exchanging between the bound and free states, retains the NMR properties of the protein receptor upon binding and because of its slower relaxation in the free state, the spectrum contains

information relative to its bound state. In the fast exchange regime, the bound ligand assumes the NOE behavior of the large receptor and the ligand NOEs consequently exhibit a change of sign passing from the free to the bound state (Figure 2.4).

However, in many interacting systems, including receptor-glycans binding, the ligands are quite large molecules and thus exhibit negative NOEs also in free state. In these cases, the discrimination between tr-NOEs originating from the bound state and NOEs of the ligand in solution may result from the determination of the build-up rates, that is the time required to achieve maximum NOE intensity. In detail, for a ligand bound to the receptor this time is in the range of 50 to 100 ms, whereas for the free ligand it is four- to ten-times longer; thus, the maximum enhancement for tr- NOEs is observed at significantly shorter mixing times compared to the free state ligand.

Notably, the construction build-up curves allow for the determination of *intra*-molecular proton-proton distances upon integration of the NOEs at different mixing times, the build-up curves are obtained from fitting using the double exponential function (Equation 2.2):

$$f = a(e^{-ct})\left(1 - e^{-bt}\right) \qquad \text{[Equation 2.2]}$$

where *f* is the cross-peaks integral a, b and c are adjustable parameters, *t* is the mixing time. The initial slope is determined from the first derivative at t = 0. Then the application of the isolated spin pair approximation leads to the corresponding distances values (Equation 2.3):

$$r_{ij} = r_{ref}\sqrt[6]{\frac{r_{ref}}{\sigma_{ij}}} \qquad \text{[Equation 2.3]}$$

where σ is the cross-relaxation rate, r_{ref} is the reference distance, r_{ij} is the unknown distance to calculate and σ_{ij} is the cross-relaxation time relative to the desired distances.

In conclusion, the tr-NOE allows fast screening of possible ligands respect to a specific target receptor and, at the same time, the knowledge of the recognized conformation of the ligand bound to the receptor, which has considerable implication for a rational structure-based drug design.

2.2.4 Saturation transfer difference NMR

STD NMR is one of the most widespread ligand-based NMR methods for detection and characterization of transient receptor- ligand interactions in solution. The technique depends on the NOE effect and allows for the determination of the binding epitope of a ligand by detecting the regions in closer contact to the receptor protein. The experiment is performed on a sample containing the receptor and an excess of ligand and relies on the transfer of saturation from receptor protons to the bound ligand that occurs exploiting the network of *inter*-molecular ^{1}H-^{1}H cross-relaxation pathways. Notably, in the experimental conditions the receptor resonances are often not detectable as the protein concentration in solution required for the STD experiment is in the micromolar range, and/ or can be easily removed prior to detection with R_2 relaxation filtering.

The STD experiment consists in the subtraction of two 1D ^{1}H-NMR spectra, the **off-resonance spectrum**, collected without protein and ligand saturation, and the **on-resonance spectrum,** acquired selectively saturating the receptor. (Figure 2.6).[150]

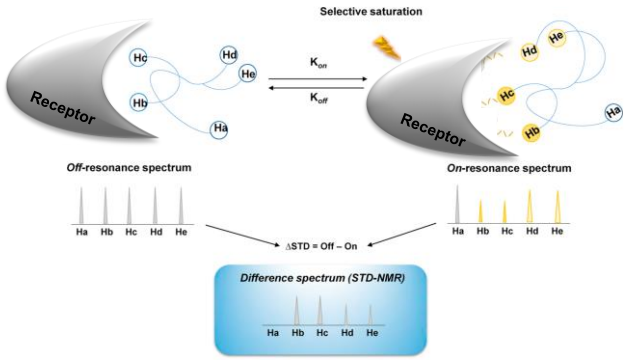


Figure 2.6 Representation of STD NMR spectroscopy. In the on-resonance spectrum frequency selective irradiation (r.f.) causes saturation of the receptor protein. The off-resonance experiment (left) produces spectra without saturation. The STD spectrum is produced by the difference between the on and off resonance spectra containing only the resonances of binding compounds. Figure adapted from ref.17

In detail, the off-resonance experiment represents the reference spectrum, recorded under conditions of thermal equilibrium and without saturation of the larger receptor. The on-resonance spectrum is acquired by selectively saturating the receptor through the application of a train of shaped pulses with a suitable **irradiation frequency** for a defined period called **saturation time**. The frequency selective irradiation (r.f.) is achieved by using shaped pulses of a value that lies into a region of the spectrum containing the receptor resonances but no signals relative to the ligand. The STD pulse scheme, by which both experiments are recorded, is depicted in Figure 2.7.

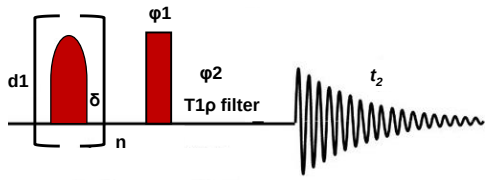


Figure 2.7 The ^{1}H STD-NMR pulse sequence.

Usually, the duration of saturation time (for on and off resonance) is 1-3 seconds and the rf pulse consists of a cascade of Gaussian shaped pulses of 50 ms length separated by an interpulse delay (δ) of ~1 ms. These parameters can be tuned to reach the desired selectivity and to avoid side-band irradiation. Usually, the on-resonance irradiation frequency values are set around -1 ppm, or, alternatively, the at the aromatic proton region, but only if no ligand resonances are present in proximity.

The receptor saturation occurs via spin diffusion and is transferred to the binding ligand via intermolecular ^{1}H-^{1}H cross relaxation at the ligand-receptor interface as they are close in space). In the fast exchange regime, the bound ligand dissociates back in solution and thus the perturbation of ligand polarization in the bound state is transferred to the free state, where saturation accumulates during the saturation time of the experiment. This process of quick transfer of the bound ligand magnetization into ligand molecules in solution, results in detection of the saturated binder signals in the spectrum. The resulting signals in the on-resonance spectrum possess intensity I_{SAT}, whereas those in the off resonance display the equilibrium value I_0. The STD response (η), in the same form as a traditional steady state NOE, is obtained by the subtraction of the two above intensities and display only resonances of the compound bound to the receptor, that constitutes the observable of the experiment.

$$\eta = \frac{(I_0 - I_{SAT})}{I_0} \qquad \text{[Equation 2.4]}$$

Because signals are reduced in the corresponding off-resonance experiment (I_0), subtraction of both experiments (I_0-I_{sat}) leads to positive difference signals for the molecule(s) affected by intermolecular magnetization transfer and, hence, to the identification of binding molecules. For non-binding molecules in the mixture, intensities in both spectra remain the same and their difference spectrum cancels

out. Thus, the difference spectrum exclusively contains signals of binding ligand(s) while nullifying any signal of non-binding compounds.[150]

Noteworthy, the ligand protons do not receive the same amount of saturation from the receptor, since the transfer of magnetization depends on the inverse sixth power of intermolecular ^{1}H-^{1}H distance in the bound state, meaning that ligand protons in closer contact to the binding site of the protein receive higher degree saturation, therefore exhibiting stronger STD signals. This enables for the evaluation of the ligand epitope mapping, i. e. the portions of the ligand that are more implicated in the recognition of a given receptor. The most common approach to identify the group epitope mapping relies on the comparison the STD response for different protons within a ligand through the normalization of all the measured STD signals against the most intense one, which is arbitrarily assumed to be 100%. The resulting STD percentages constitute a qualitative indication of the ligand part in close contact with the receptor for molecular interaction and of the ligand moieties far from the receptor binding site.

Within the ideal scenario, the epitope mapping should not depend on the chosen saturation time; however, ligand protons may possess significantly different R_1 relaxation rate that causes artifacts in the STD experiments as protons with slower R_1 relaxation efficiently build up saturation in solution and display higher STD relative intensity, thus implying an overestimation of their proximity to the receptor, and vice versa. To overcome this limitation, herein allowing the quantitative analysis of ligand epitopes, STD build-up curves can be employed.[151] Indeed, the acquisition of STD spectra at different saturation time (usually from 1 s to 5 s) and the consequent determination the initial slope (STDfit) for each ligand proton represents a good way remove all these artifacts, since at zero saturation time no increase of saturated ligand arises, as well as the intramolecular spin diffusion at the bound state is minimum. The slope of the curve at zero saturation

time is obtained from fitting the STD NMR intensities data using the following monoexponential equation (Equation 2.5):

$$STD = STD_{max}(1 - e^{(k_{sat}t)}) \quad \text{[Equation 2.5]}$$

where STD stand for the STD signal intensity of a given proton at a saturation time t, STDmax represents the asymptotic maximum of the curve, and ksat is the value of observed saturation rate constant, which measures the velocity of STD build up.

The epitope map of the ligand is obtained by normalizing all the values of different ligand protons to the largest STDfit, giving STD epitopes fit, that represents the STD intensity dependent only on the vicinity of the ligand protons to the receptor.

It is worth to remark that STD effects depends largely on the off rate; the K_D range suitable STD method has been estimated to be 10^{-8} – 10^{-3} M. STD NMR method are indeed well suited for the study of protein – carbohydrate complexes which exhibit excellent properties as the usual K_D values lies in the optimal range. For too weak interactions, the probability of the ligand to reside in the receptor pocket is very low, the population of the bound state decreases, leading to a reduction and eventually to the disappearance of the STD signals. Conversely, if the binding is very tight the receptor - ligand lifetime is too large, and the saturation transfer is not efficient enough, hence, weak or zero STD signals are displayed in the spectrum. To conclude, STD NMR is a solid and versatile technique which gives essential information, at a molecular level, on receptor – ligand interactions.

2.3 Computational techniques for carbohydrates modelling

Molecular modelling has proven to be very useful for the elucidation of the structural and conformational features of carbohydrates molecules. Furthermore, the combination of molecular modelling with experimental techniques such as solution state NMR spectroscopy methods is extremely valuable to determine the conformational and dynamic properties of free and bound saccharide structures.[15,152,153]

In general, docking techniques predict the possible binding poses, namely the bound conformations of small molecule ligands, at a specific binding site of a given macromolecule, and delivers a first estimation of the binding energy of each model. Molecular mechanics (MM) and molecular dynamics (MD) allows for the assessment of the type and strength of the interactions between the ligand and its biological target. These methods are also used to estimate the bioactive conformation of ligand and to observe the conformational changes that may occur to both interacting partners upon binding. When the tridimensional (3D) structure of the target has not been elucidated structurally, it can be predicted by homology modelling, using templates that share significant sequence similarity.

Molecular modelling of carbohydrates can be performed at different levels of complexity, the most common protocols are based on the application of molecular mechanics and molecular dynamics simulations using *ad-hoc* force fields (Figure 2.8)[154,155] Indeed, the research in this matter has been very active in recent years, yielding to the development of several force fields able to accurately reproduce the unique dynamic and electronic properties of oligo/polysaccharides.[156,157,158] The selection of the appropriate computational methods depends on the properties under investigation, and also on the initial available experimental

information. In the next sections, the molecular modelling tools employed during the thesis will be overviewed.

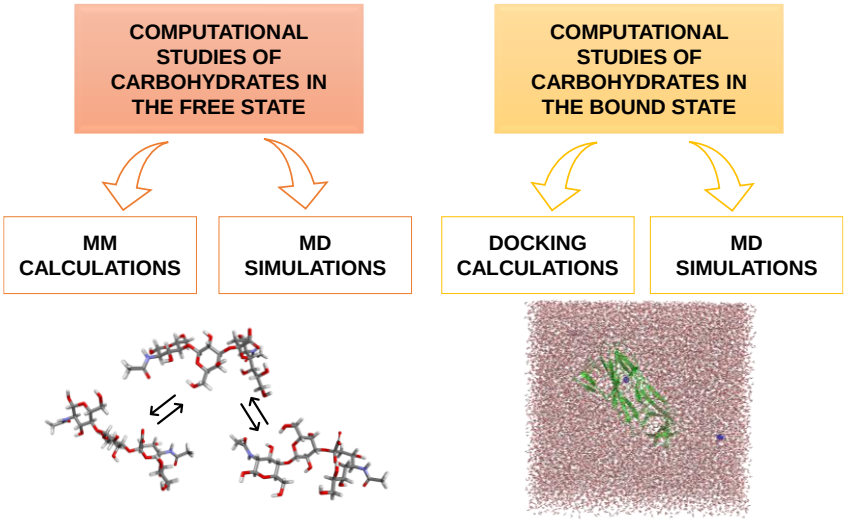


Figure 2.8 Computational methodologies commonly employed for the study of carbohydrates and their interaction with receptors.

2.3.1 Docking

Molecular docking is a computational method that, starting from the individual structures of a given ligand and a macromolecule target, provides the possible orientations of the ligand into the target binding site to form a stable complex. Docking is a useful tool to assess the bioactive conformations, identify putative binding sites within a given macromolecular target, pinpoint essential ligand-receptor interactions as well as to quickly screen libraries of potential ligands. Several docking programs are available in the context of receptor ligand recognition and drug design, including AutoDock[159], AutoDock Vina[160] and GLIDE[161]among others. Nowadays, the increased interest for the characterization

of protein-protein complexes, which are difficult to detect experimentally, led to the development of many specific protein- protein docking servers.[162]
Computational docking protocols require the availability of the 3D structure of the biological target that can be obtained by means of X-ray crystallography, NMR spectroscopy, cryogenic electron microscopy or homology modelling. Knowledge about the receptor binding site is preferable, otherwise it may be assumed by comparison of other related target proteins or through complementary bioinformatic analyses.

Even if docking programs/servers function differently, they share the overall workflow that consist in two main phases: a conformational search for the prediction of possible conformations of the ligand; followed by the energy scoring of the resulting binding poses, so the predicted ligand-receptor complexes are scored and ranked by the application of a scoring function that evaluates the binding energy. The main differences among the existing docking programs rely on the type of computational search algorithm and the nature of the scoring function applied to rank the docked poses.[163,164]

Concerning the conformational search step, the number of ligands rotable bonds is the main factor influencing the accuracy and the computational cost. Because the increased number of rotable bonds raises exponentially the computational cost, each program has a limit to its maximum number. The most used search algorithms applied for the conformational search are shape matching, systematic search and stochastic algorithms.
Stochastic algorithms such as Monte Carlo (MC) methods are Evolutionary Algorithms (EA) are the best in term of time-efficiency. They are based on the induction of random ligand changes (translation, rotations, orientations) which can be accepted or rejected depending upon a defined probability criterion. In MC methods Boltzmann probability function is applied for the optimization, whereas for

EA the selection is based on the genetics and evolutionary processes in biological systems. Amongst the EA class, the popular Genetic Algorithm (GA) applies an analogy with evolutionary biology to establish the fitness of the generated conformations, by which the favorable ones will ‘’ survive and will reproduce’’ (upon crossover and mutation), whereas the worse ones will be discarded.[165] This method can be further improved by the introduction of a Local Search method (LS), giving rise to the hybrid Lamarckian genetic algorithm with enhanced performance.[166]

Regarding the energy scoring step, the scoring and ranking of the predicted binding modes may be mainly performed by means of three types of scoring functions: force field scoring functions (ff), empirical scoring functions and knowledge based (kb) scoring functions. The ff function applies a classic force field to compute individual interaction terms such as van der Waals and electrostatic energies as well as stretching, bending, and torsional energies.[167] The empirical function summing up a set of weighted empirical energy terms, including hydrogen bond and hydrophobic interactions for the evaluation of the overall binding free energy.[168] The knowledge -based function expresses the binding energy as the sum of distance-dependent statistical potentials between the ligand and the target.[169]

For the scopes of this thesis AutoDock4[170] molecular docking program was mainly used. AutoDock works through an automated procedure, by which the grid maps of interaction energies are pre-calculated. These maps are then employed during the docking calculation to estimate the total energy of binding between the ligand and the macromolecule thus significantly reducing the time required for the docking calculations.

In AutoDock, the conformational search is usually performed by Lamarckian genetic algorithm and the docked binding poses are scored and ranked

by a semi-empirical function. According to such scoring function the binding free energy is calculated as the difference between the potential energy of ligand and the protein in the complex (bound state) and the potential energy of the ligand and protein in a "free" or unbound state (Equation 2.6):

$$\Delta G = \left(V_{bound}^{L-L} - V_{ubound}^{L-L}\right) + \left(V_{bound}^{P-P} - V_{ubound}^{P-P}\right) + \left(V_{bound}^{P-L} - V_{ubound}^{P-L} + \Delta S_{conf}\right) \quad \text{[Equation 2.6]}$$

Each potential energy contribution is obtained by the sum of van der Waals, hydrogen bonds, electrostatic and solvation terms. The loss of entropy (ΔS_{conf}) associated to the binding is also taken into account. This contribution is strictly related to the number of rotatable bonds (N_{tors}) in the ligand (Equation 2.7):

$$\Delta S_{conf} = W_{conf} N_{tors} \qquad \text{[Equation 2.7]}$$

2.3.2 Molecular Mechanics and Molecular Dynamics Simulations

The particle's behavior at the atomic level is accurately described by the laws of quantum chemistry even though this level of theory, considering the molecular orbitals and occupying electrons for each atom constituting the molecule, requires high computational costs especially for big size macromolecules. Nonetheless, Molecular Mechanics or classical mechanics describes molecules as a set of bonded atoms whose interactions can be modelled using standard classical physics combining mechanical tools (springs, tensors, rotators) and electrostatics. The use of classical mechanics to model molecular systems is called molecular mechanics. In this classical description, the atoms are represented as charged spheres, whose size is usually proportional to the van der Waals radius of the given atom. These spheres are linked by covalent bonds and by non-bonded interactions, comprising van der Waals and electrostatics. The atoms and the nuclei positions

are considered to have a static electron distribution and are the linkage is treated by elastic (harmonic) forces, according to the Hook's law.[171] In this context, the atomic interactions can be modelled with simple parameterized functions measuring the energetics of the system, -called force fields, that are typically obtained either from *ab initio* or semi- empirical quantum mechanics (QM) calculations or by fitting of experimental data. A set of equations describes the potential energy and forces relative to the particle coordinates, as well as a collection of parameters required to the equation solution. The potential energy function consists into two main terms: bonded potential energy and non-bonded. The bonded potential energy defines the covalent contributions (bond stretching, angle bending, and dihedral and improper torsions) to the total energy, and the non-bonded one defines the repulsive and van der Waals interactions (12-6 Lennard-Jones potential) and the Coulombic interactions, as described by the equation 2.8.

$$V_{(r)} = \sum_{bonds} K_l(l - l_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{nonbond} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{rij^6} + \frac{q_j q_i}{4\pi\varepsilon_0 r_{ij}}\right] \text{[Equation 2.8]}$$

The first term of the equation describes bond stretching in which K_l is the force constant relative to the displacement l from equilibrium positions l_0. The second term describe the bond angle vibrations ($k\theta$ is the force constant, θ the instantaneous angle and θ_0 the equilibrium angle). The third term describes the torsions, i.e. the dihedral angle potential in which *Vn* is the dihedral constant, *n* the periodicity parameter, ω the instantaneous dihedral angle and γ the phase term. The fourth term represents non-bonded or "through-space" interactions between atom pairs, that is decomposed into Lennard-Jones and Coulomb interactions, *qi* and *qj* are the respective atomic charges *i* and *j*, *rij* the distance between any two atoms and *Aij* and *Bij* parameters for the repulsive and attractive components of the Lennard-Jones potential.

MD simulations are a very important tool for a complete comprehension of the conformational dynamics of the system. They allow for the analysis of molecules and complexes at different timescale and in solution, and they are widely used to study the molecular properties as well as dynamic events such as receptor/ligand interactions. In MD simulations Newton's second law of motion is applied predict how atoms of a given molecular system will move over time.[172] An MD simulation is set up by assigning initial velocities and positions to all atoms in a biomolecular system. The forces acting on each atom are then calculated, and consequently the Newton's laws of motion can give the direction, thus predicting the spatial position of each atom as a function of time. The *iter* is repeated for defined time intervals, the forces evaluated each step are used to establish the position and velocity of each atom in the subsequent step. This produces a trajectory that describes the atomic-level configuration of the system at every point over the desired time period.[154]

A classical MD simulation consists in several steps. The first step is an energy minimization, in which, starting from higher initial energy state, the search of a minimum in the energy landscape of a system is performed. The most common method applied in MD is the steepest gradient, that execute the geometry optimization until reaching the local minimum.[173] This method depends on the starting geometry and is suitable for quicker optimization. During this phase, the solvent environment is very important. Three different methods can be applied to treat it. In *in vacuo* simulation, a distance-dependent solvent is used. If the solvent is considered implicit, two different models can be applied, Generalized Born (GB) and Poisson-Boltzmann (PB). With explicit water, the simulation of complex takes place into a box of a defined shape and dimension where solvent molecules surround the simulated molecules, periodic boundary conditions (PBC) may be applied to avoid surface artefacts. Different water models have been

proposed, but the most employed is the TIP3P model.[174]To calculate the infinite electrostatic interactions, particle mesh Ewald summation (PME) is used in which the summation into short- and long-range parts is split.[175]

The following step is the heating phase that aims to remove the unfavorable contacts, as steric clashes and donor/donor or acceptor/acceptor bonds, between solvent and solute by enhancing the atoms speed through a defined scheme of temperature increase. The resulting velocities are calculated by means of standard temperature-dependent Maxwell-Boltzmann distribution. The next phase is the equilibration that consists in the system relaxation under controlled energy, temperature, pression and volume conditions. Last, the production step produces the final trajectory.

Prior to MD simulations, the choice of the correct force field is a crucial step. The choice relies on the nature of the molecules to simulate as highly accurate force fields for specific systems has been developed. AMBER package,[176] the molecular dynamic simulation program here used, contains force field optimized for lipids (to date Lipid14),[177]for sugars (GLYCAM06),[156] for proteins, nucleic acids and water molecules (ff14SB)[178] and for organic molecules (GAFF).[179]

In the context of receptor carbohydrate interactions, MM and MD simulations can be applied to get insights into the conformational behavior in solution of glycan molecules, alone and in complex with the biological target. The conformational analysis is key to determine the three- dimensional features, especially in the carbohydrate field as it allows to correlate a molecule with its biological behavior. The conformational search aims to find the minimum local energy for a given molecule by rotating single bonds thus causing the rearrangement of the atoms in space. The relative energies are then evaluated to find the lower energy conformer. The resulting structure can be then compared to information obtained by experimental techniques, such as NMR, for example the NOE derived proton

distances, to obtain a reliable 3D topology of the glycans of interest. Indeed, the combination of modelling/NMR protocols is extremely useful to deduce the conformational and dynamic properties of free and bound carbohydrate molecules and thus obtains the 3D complex of the system under study.[180]

2.3.3 Homology modelling

The availability of the 3D structure of a given protein is a limiting step in the drug design and development. Usually, the 3D structures of a given macromolecules are elucidated by NMR or X-ray crystallography techniques, and then deposited in the public database Protein Data Bank (PDB).[181] When the 3D structure of the target protein has not been resolved by any experimental techniques, homology modelling allow to predict the 3D structure from its amino acid sequence. The protein sequence (target) is available on databases such as *Universal Protein Resource* database (UNIPROT),[182] or National Center for biotechnology information (NCBI)[183]. To build a homology model three different protocols may applied: *ab initio* calculations, fold recognition and comparative modelling. Such protocols are based on the correlation between of the protein structure, and thus its function, to its amino acid sequence, and on the theory that the native structure of the protein has the lowest free energy. Comparative modelling, that has been used within this thesis, implies that one or more structures (templates) from the PDB are homologous to the target and thus the template 3D structure drives the building of the target model. Hence, to perform comparative homology modelling, one (or multiple) experimental 3D structure reasonably similar to the target macromolecule is required. This procedure is done by aligning the target sequence to databases of proteins of known 3D structures. Commonly, the template identification and selection is achieved by *Basic Local Alignment Search Tool* (BLAST),[184] an online resource that provide an estimation of the quality of the

sequence alignment. In template selection, parameters such as the percentage of identity between the searched protein and the template (a good template has minimum 30% of the percentage identity to the target)[185] and the percentage of query cover are considered. Once the templates have been identified, the structure of the target and the templates are submitted to homology modelling programs or servers. In the present thesis, the required models were built using several homology modelling servers, including SWISS MODEL[186], I-TASSER[187] and Phyre2[188].

Prior to further use, a refinement step of the generated model, especially of the side chains, is necessary. This can be achieved by restraining the Cα of the backbone through MD simulation, thus enabling only the side chain rearrangement. Last, the structure needs to be validated by analyzing different parameters including the overall structure geometry and the statistics of non- bonded interactions. In these studies, the online resource PROCHECK[189] has been applied for model validation. This uses the SAVES tool[190] that inspects the compatibility between the predicted secondary structure for a protein according to the primary sequence. PROCHECK also display Ramachandran Plot of the given model, which is useful to check backbone angles of the protein structure by comparing them with observed/experimental angles. In this regard, a good homology model usually has a score of more than 90% meaning that at least 90% of the residues are in the most favorable energy region.

2.3.4 Complex validation: CORCEMA-ST

CORCEMA-ST (Complete Relaxation and Conformational Exchange Matrix Analysis of Saturation Transfer) is a valuable tool for the quantitative structural interpretation of experimental STD NMR data. The program relies on the implementation of the NMR saturation transfer concept within the CORCEMA theory.[191,192]

Based on the fast exchange model discussed above (see section 2.2), CORCEMA-ST, starting from the Cartesian atomic coordinates of a given ligand-receptor complex and other system properties specific to the interaction (e.g., dissociation constant, K_{off}, rotational correlation times of the receptor and ligand) can predict the STD of the ligand protons. Then, the experimental STD values can be compared with those predicted for a model of the complex obtained by X-ray crystallography, NMR or docking. The program performs a matrix calculation which considers the Cartesian coordinates of all ligand protons and protein protons within a provided cut-off distance. The fitness of the molecular model respect to the experimental values is given by the R-NOE factor:

$$R-NOE=\sqrt{\frac{\sigma W_k\left(STD_{exp,k}-STD_{cal,k}\right)^2}{\sigma W_k\left(STD_{exp,k}\right)^2}} \quad \text{[Equation 2.9]}$$

where $STD_{exp,k}$ and $STD_{cal,k}$ are the experimental and calculated STD intensities, respectively, of proton k. The lower the R-NOE factor value, the better the fit between experimental and theoretical data. The R-NOE factor can be used as a scoring function to drive a conformational search for the ligand bound in the receptor's binding pocket.[60] Different structural models obtained from docking can be ranked according to how well they reproduce the experimental STD NMR data, allowing for the selection of a suitable structural model of the complex.

Furthermore, a good starting structure can be optimized via simulated annealing to find the global energy minimum of the ligand in the bound state, further minimizing the R-NOE factor. In this way, the geometry of the ligand inside the binding pocket is, thus, refined by use of experimental STD NMR data.[191]

The combination of experimental NMR techniques, such as transferred NOESY and STD-NMR spectroscopy, together with computational approaches, including Docking, MD simulations and CORCEMA-ST protocol has proven to be effective to provide the three-dimensional structures of ligand-receptor complexes.[193]

2.5 Objectives

In this Thesis, the overall objective is to elucidate the Siglec-sialoglycans interactions at the atomic level by means of ligand based NMR techniques (mostly STD NMR and tr-NOESY) and molecular modelling and implemented with binding assays. In particular, the following Siglecs were inspected: Siglec-2, Siglec-10, and -7. Furthermore, the structural basis of the interaction between glycans and other glycan binding proteins (Macrophage Galactose type Lectin and mumps virus Hemagglutinin neuraminidases) were dissected by applying a similar approach. The final goal is to provide new insights for the understanding of the molecular recognition events underlying the functions of these important biological targets for therapeutic applications. The work has been structured into several chapters that are summarized below:

Chapter 3: Characterization of the characterization of the interactions between *N*-Glycans and CD22/ Siglec-2. The binding modes of naturally occurring sialoglycans were elucidated by biophysical assays, NMR techniques and molecular modelling. Also, the role of different forms of sialic acids in the interaction with human and murine CD22 orthologues was explored. Last, the binding of a potential synthetic sialic acid derivative, potential modulator of CD22 activity, was disclosed. The outcomes provided a structural perspective for the rational design and development of high-affinity ligands to control the receptor functionality.

Chapter 4: Structural study of the molecular mechanisms regulating the interaction between Siglec-10 and naturally occurring sialoglycans. The combined spectroscopic, computational and biophysical approaches provided the glycans' epitope mapping and conformation upon binding and afforded the atomistic description of the interaction interfaces.

Chapter 5: The molecular bases of the recognition of the sialylated LPS of *F. nucleatum* 10953 strain by Siglec-7; through ligand-based NMR techniques, biophysical assays and *in silico* methodologies to propose a structural model of the interaction. Novel structural insights on the binding of the *F. nucleatum* sialylated O-antigen in the interplay with Siglec-7, both important actors in tumorigenesis were afforded. The obtained information will be valuable in the development of therapies effective against the opportunistic pathogen diseases, as well as for the modulation of Siglec-7 activity.

Chapter 6 The molecular interaction between the human macrophage galactose-type lectin (MGL) and the lipooligosaccharide (LOS) of *Escherichia coli* strain R1 was undertaken. Saturation transfer difference NMR spectroscopy analysis, supported by computational studies demonstrated the ability of MGL to recognize glycan moieties exposed on Gram-negative bacterial surfaces.

Chapter 7 The structural characterization of the mumps virus hemagglutinin neuraminidases (MuV-HN) of different viral strains in the interplay with natural and synthetic sialoglycans was addressed. The molecular recognition and enzymatic mechanism were analyzed by a multidisciplinary approach. These outcomes may help in the identification of new inhibitor scaffolds which could prove worthwhile in the fight against mumps virus.

SECTION II-RESULTS AND DISCUSSION

Chapter 3-Structural insights on the recognition of sialoglycans by Siglec-2

3.1 Introduction

CD22 is an inhibitory receptor prominently expressed on B-lymphocytes, where it prevents dysregulation of tolerance mechanisms, self-reactivity and autoimmunity.[68,194] Recently, the crystallographic structure of human CD22 has been solved, both alone and bound to 6'-sialyllactose.[195] The CRD of CD22 selectively recognizes Sias α2-6 linked to Gal residues as terminal part of endogenous glycoproteins expressed on the surfaces of mammalian cells. Further studies demonstrated that the extracellular portion of CD22 adopted an extended conformation that favors the formation of nanoclusters with potential implications in therapeutic antibody engagement (Figure 3.1).

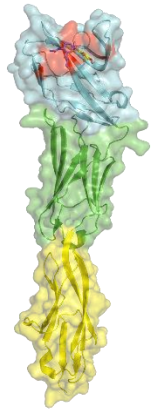


Figure 3.1 Crystal structure of human CD22 in complex with 6' sialyllactose. The three resolved N-terminal Ig domains of human CD22 represented as transparent surface. The V-Set domain is shown as cyan surface, the adjacent C1-type set domain is depicted in green surface, the following C2-set domain is shown as yellow surface. The residues of V-set domain that are known to bind sialic acid and galactose are indicated in red. (PDB ID:5VKM).

The binding between CD22 receptors and *N*-glycans is biologically important in the regulation of the BCR signalling pathway, resulting in several immune consequences. In physiological conditions, on resting B-cells, CD22 is "masked" by *cis* interactions with sialoglycans exposed on the same plasma membrane, resulting in the formation of homo-oligomers (Figure 3.2)[196] that sequester the CD22 away from BCR Conversely, the co-ligation of CD22 to the BCR by a membrane that exposes both antigen and sialylated ligands (*trans* interactions), inhibits the B cell receptor signalling pathway.[197] Upon association with the BCR, indeed, specific protein kinases phosphorylate the CD22 ITIMs and promote the recruitment of the tyrosine phosphatase SHP-1, that, in turn, negatively regulates the B cell signalling.[198] This leads to the suppression of autoimmune B cell responses only to "self" antigens.

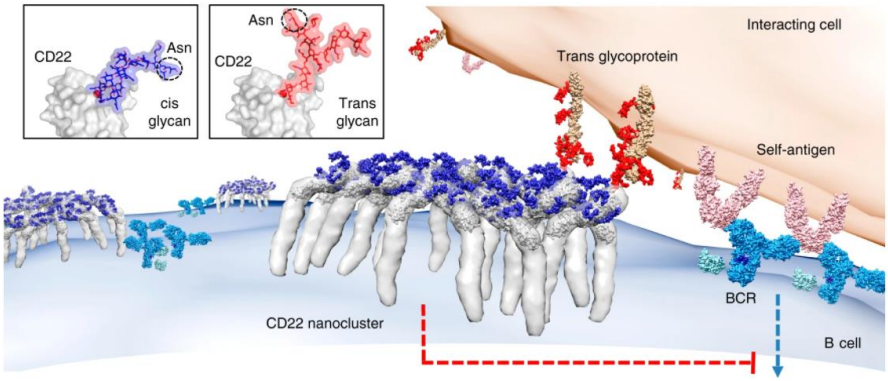


Figure 3.2 Model of CD22 homo-oligomers on B cells (Figure from ref. 195). The illustration depicts the interaction between sialoglycans in cis (blue surface) and trans (red surface), that could be mediated by the flexibility of complex-type glycans (inset). The CD22 is shown as gray surface. BCRs (sky blue and cyan) (PDB IDs: 1HZH and 3KH0)[199,200] are shown binding to self-antigens (pink) (PDB ID:4ZXB)[201] on an interacting cell. Blue arrow represents BCR activation, whereas red dashed line represents BCR inhibition upon the engagement of CD22 with trans glycans (red) promotes recruitment of CD22 oligomers thus resulting in B-cell signalling inhibition.

In this context, CD22 sets the threshold for the proper B-cell activation; hence, reduced sialic acid ligand synthesis or CD22 deficiencies can result in uncontrolled B cell signalling and systemic autoimmunity.

CD22 is conserved in human and murine species, and the ligands on B-cells that act as modulators as well, this is remarkable because high affinity glycan ligands of CD22 are specific depending on the specie. In both species, the high affinity ligands are based on the sequence Siaα2–6Galβ1–4GlcNAc, which typically terminates N-glycans. Nonetheless, the human ligand has Neu5Ac as Sia, whereas the high affinity ligand for mouse CD22 has Neu5Gc.[202]

The role of co-receptor of BCR makes CD22 relevant in the modulation of tolerance to self-antigens that is critical to prevent autoimmune diseases and malignancies including hairy cell leukaemia, marginal zone lymphoma, chronic lymphocytic leukaemia and non-Hodgkin's lymphoma.[93,94,195, 203]

Given the role of CD22 in the etiology of autoimmune diseases and cancer, several tools based on Siglec-2 as candidate target in immunomodulation and tumor therapies have been proposed, with a particular focus on the development of antibody-based approaches.[204,205] For instance, the monoclonal antibody epratuzumab was tested against CD22 for the treatment of Systemic Lupus Erythematosus a B cell inhibition due to this treatment was observed under clinical stages.[206] Parallel to the use of antibody, an emergent approach for the treatment of autoimmune diseases involves glycan-based therapies that relies on the design of specific ligands with high-affinity for the Siglecs binding sites, able to overcome the *cis* interactions of endogenous glycans.[207]Despite different sialic acid mimetics targeting the CD22 have been developed,[207,208] further structural and mechanistic insights are needed to unveil the molecular basis of N-glycans

recognition and binding by CD22, which lays at the basis of the rational design of a new generation of glycomimetics.

To this end, a multidisciplinary approach was applied to shed light on the molecular mechanisms at the basis of sialylated N-glycans recognition by CD22. The overall objectives were i) to understand the conformational behavior of naturally complex-type *N*-glycans and the relationship with CD22 homo-oligomers formation (paragraph 3.2) ii) to elucidate the molecular mechanisms of Neu5Ac and Neu5Gc recognition and binding to the human CD22 (h-CD22) and in comparison with the murine ortholog (m-CD22), paragraph 3.3 iii) to inspect the binding of a sialylated analogue as potential high affinity CD22 ligand as candidate to be employed in therapeutic applications, paragraph 3.4.The use of ligand-based NMR spectroscopy, binding assays combined to MM and MD simulations, as well as docking and CORCEMA-ST methods, provided crucial information concerning the binding epitope and the conformational behaviour of complex sialylated N-glycans and allowed to describe the 3D complexes. The gathered data might provide rational framework opportunities for further developing CD22 potent inhibitors and/or regulators.

3.2 Characterization of the dynamic interactions between complex N-glycans and human CD22

To dissect the molecular mechanisms at the basis of sialylated N-glycans recognition by human CD22, its interaction with various sialo-oligosaccharides (Figure 3.3), including naturally occurring biantennary complex-type N-glycans (ligands **2**-**6**, Figure 3.3) typically exposed on cell surface glycoproteins, was proven. The work was carried out in collaboration with Prof. Sonsoles Martín Santamaría (CSIC, Madrid), Prof. Bruno Pagano (University of Napoli Federico II), Prof. Paul D. Crocker (University of Dundee) and Prof. Koichi Fukase (University of Osaka).

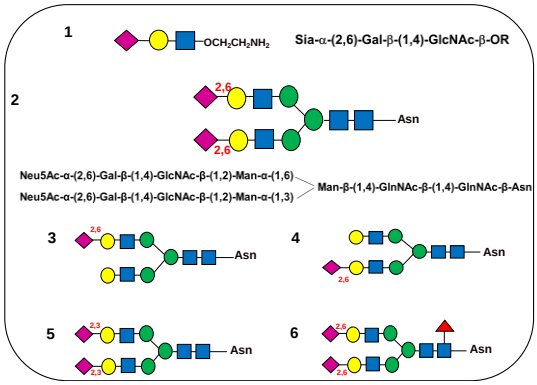


Figure 3.3 Glycan structures (1-6) used in this study.

3.2.1 Affinity of h- CD22 toward different sialoglycans: SPR and Alpha Screen

First, the affinity of the interaction between h-CD22 and different sialoglycans were gathered by SPR and Alpha Screen analysis. SPR equilibrium binding experiments were performed in collaboration with Prof. Bruno Pagano (University of Napoli Federico II) to analyze the interaction between ligand **1** and h-CD22. h-CD22 fused to Fc portion of human IgG1 was provided by Prof. Paul D.

Crocker (University of Dundee, UK). Details concerning the expression and purification methods are reported in section 9.1.1.

h-CD22 was directly immobilized on the sensor surface and a range of increasing concentrations of compound **1** were injected. The difference between the response of the reference cell and the response observed upon **1** addition, represented the actual binding of **1** to the protein. The SPR sensorgrams acquired as results of the binding event are shown in Figure 3.4,a. Trisaccharide **1** clearly displayed very fast association and dissociation profiles to h-CD22, as upon injection of **1**, the binding rapidly reached equilibrium in the first few seconds and returned to baseline rapidly after the end of the injection. The obtained dissociation constant (K_D), calculated from the plot of the response at equilibrium versus the analyte concentration was 250 μM (Figure 3.4,b).

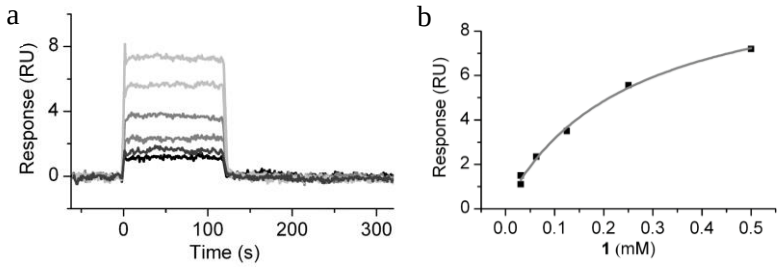
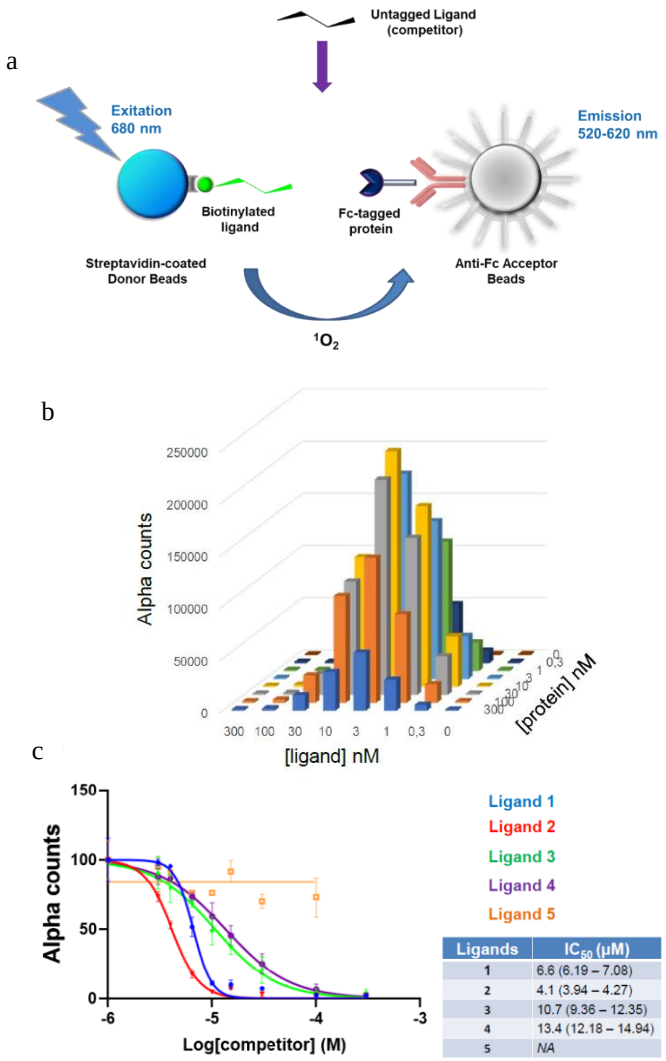


Figure 3.4 SPR analysis of trisaccharide 1 binding by CD22. **a)** Overlay plot of SPR sensorgrams from steady-state affinity analysis of h-CD22 binding to trisaccharide **1** at concentrations ranging from 0.031-0.5 mM. The sensorgrams are reference subtracted and blank subtracted**. b)** The equilibrium SPR response is plotted against the analyte concentration (**1**, x axis). Solid line represents the nonlinear curve fitting to the data (squares).

Then, the apparent IC50 of ligands **1-5** were determined by carrying out the Alpha (Amplified Luminescent Proximity Homogeneous Assay) Screen (PerkinElmer Life Sciences) assay[1], using h-CD22-conjugated acceptor beads,

[1] Performed by PhD Cristina Di Carluccio

streptavidin-coated donor beads, and biotinylated PAA-6'sialylactosammine (6'-SLN-PAA-biotin, PAA=polyacrylamide) as tracer (Figure 3.5). In the absence of competitive ligands, the interaction between 6'-SLN-PAA-biotin and h-CD22 brought the donor beads into proximity of the acceptor beads, resulting in light production (alpha signal) upon illumination of donor beads. The association between the protein and the tracer is affected in the presence of untagged competitors yielding a decrease of alpha signal.



Ligands	IC_{50} (μM)
1	6.6 (6.19 – 7.08)
2	4.1 (3.94 – 4.27)
3	10.7 (9.36 – 12.35)
4	13.4 (12.18 – 14.94)
5	*NA*

Figure 3.5 Binding affinity of h-CD22 binding to sialoglycans by Alpha assay a) Scheme of alpha screen protocol **b)** 3D bar-graph, reporting the alpha counts (in the absence of competitive ligands) vs different concentrations of the tracer and the protein**.** The graph was obtained using Microsof Excel. **c)** Alpha Screen assay of ligands **1**–**5** binding to h-CD22. IC50 values were calculated with a dose-response formula using GraphPad Prism. Each data point represents the mean SD of triplicate measurements. 95% confidential intervals (95% CI) are listed inside parenthesis.

The results of the displacement assay confirmed the specificity of the h-CD22 only for α(2,6) linked sialoglycans, since no inhibition was observed for ligand **5** (α(2,3) linkage)). Conversely, the ligands **1-4** showed binding to h-CD22. Interestingly, the biantennary glycan **2** exhibiting the lower IC50 value, thus the highest inhibitory potency. Additionally, as expected, similar inhibitory potencies for the asymmetrical glycans **3** and **4**, were observed. Their IC50 values were greater than that of the biantennary undecasaccharide, implying the possible involvement of both arms in the receptor binding.

3.2.2 Molecular recognition of sialotrisaccharide 1

STD NMR was employed to define the binding epitope of **1** when interacting with h-CD22. As shown in the spectrum reported in Figure 6a, remarkable changes in the multiplicity and relative intensity of STD signals with respect to those of the reference were observed, indication of specific binding. In detail, the highest STD signals belonged to the Neu5Ac (**K**) and especially to its acetyl group, indicating a major involvement in the interaction. Significant STD enhancements were observed for other protons of Neu5Ac, such as those belonging to the side chain. Intense STD effects were also ascribable to the H4 proton of Gal (**B**) (close to 60%); likewise, H6 proton of Gal received saturation from the protein, although to a lesser extent. These findings suggested that either Neu5Ac and Gal moieties regulated the recognition and interaction processes. Interestingly, a triplet attributed to $H6_R$ of the Gal moiety was distinctly observed around 4 ppm (Figure 3.6a) in the STD spectrum, conversely, in the off-resonance spectrum a complex multiplet was present instead, derived from the overlapped resonances of H6 protons of **A** and **B**. On the other hand, no STD signals were observed for the GlcNAc (**A**), meaning that it pointed far from the protein surface and apparently was not significantly implicated in the interaction. These qualitative results

were then refined by a quantitative analysis achieved by the construction of STD build up curves (Figure 3.6b), which allowed to accurately map the epitope of **1** in its interaction with h-CD22 (Figure 3.6c). The obtained results confirmed the significant involvement of the Neu5Ac unit in the interaction; the methyl group of the N-acetamide moiety, giving the strongest STD effect, was set to 100%, the other relative STD percentages were derived accordingly (Table 3.1).

Table 3.1 STD values derived from the STD-NMR of h-CD22/ligand 1.

^{1}H	*STDmax*	*Ksat*	*STD (fit)*	*STD epitopes (fit)*
K Ac	2.0593	0.3725	0.7671	**100%**
K3ax	0.4342	0.4799	0.2083	**27%**
K3eq	0.4431	0.4853	0.2150	**28%**
K5	0.6731	0.3493	0.2351	**30.6%**
K6	1.1775	0.4381	0.5159	**67.2%**
K7	1.0562	0.4245	0.4483	**58.4%**
K8	0.6790	0.4431	0.3001	**39.1%**
K9R	0.5748	0.5763	0.3313	**43.2%**
H4 Gal	0.9346	0.4757	0.4445	**57.9%**
H6R Gal	0.8157	0.4578	0.3734	**48.6%**
H5 Gal	0.5122	0.6201	0.3176	**41.4%**

.

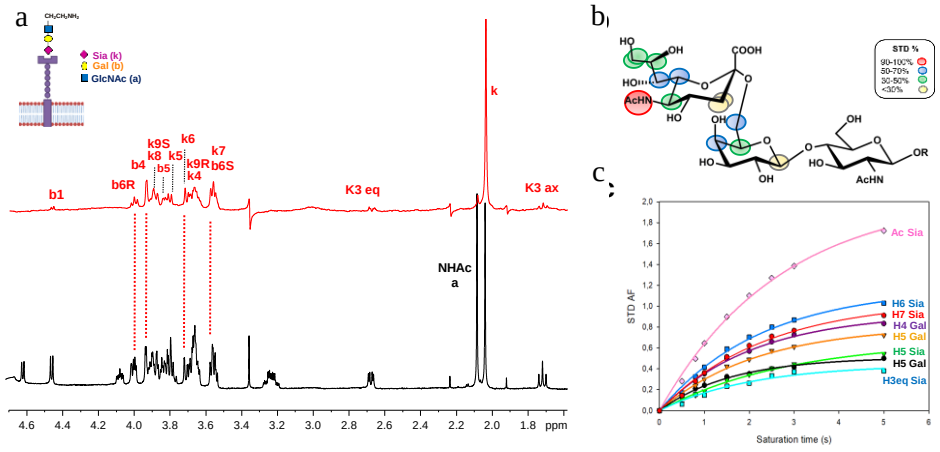


Figure 3.6. STD NMR analysis of the interaction between h-CD22 and trisaccharide 1. a) Reference ^{1}H NMR spectrum (bottom) and STD 1D NMR spectrum (top) of h-CD22/trisaccharide **1** mixture (at saturation time of 2 s) in deuterated phosphate buffer. **b)** Epitope map of ligand **1** calculated by $(I_0-I_{sat})/I_0$, where (I_0-I_{sat}) is the intensity of the signal in the STD NMR spectrum and I_0 is the peak intensity of the unsaturated reference spectrum (off-resonance). **c)** The STD STD build up curves.

Given the relevance of conformation and topology of *N*-glycans in modulating the receptor specificity,[209] the bioactive conformation of **1** in the interaction with h-CD22 has been derived. Firstly, the conformational features of **1** in the free state were assessed by 2D-NOESY in combination with Molecular Mechanics and Dynamics simulations, carried out by Prof. Martín Santamaría group (CSIC, Madrid). The collected results demonstrated that the trisaccharide **1** exhibited a preference for the geometry with ϕ/ψ/ω torsion angles of -60°/180°/60° for the α-Neu5Ac-(2,6)-β-Gal glycosidic linkage, corresponding to the *gt* conformer (Figures 3.7-9, Table 3.2), as experimentally attested by the ROE contact between the proton H5 of the Neu5Ac and the acetyl group of the GlcNAc residue, observed also in the bound state (Figure 3.8).[210]

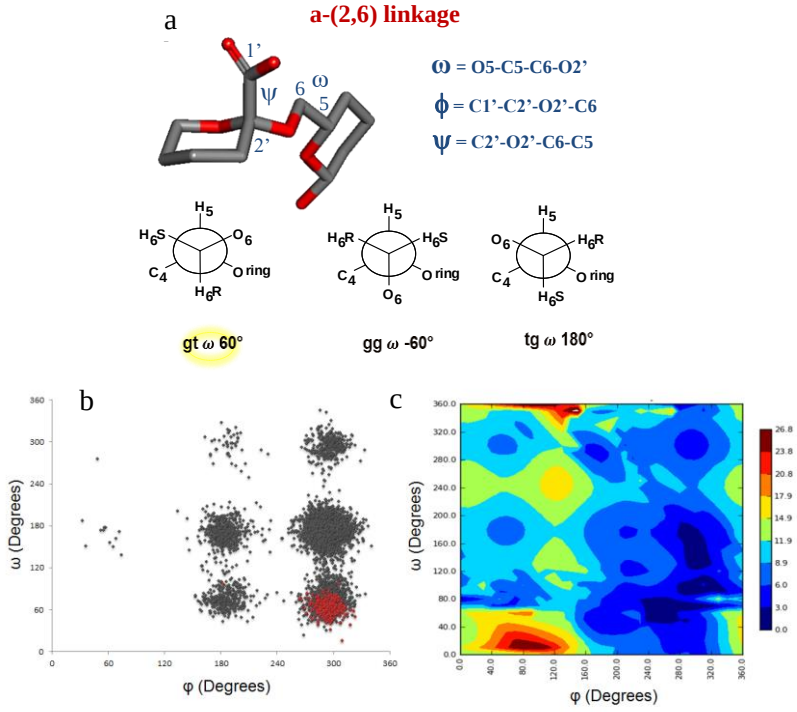


Figure 3.7 MM and MD conformational study of free state trisaccharide 1. a) The definition of the torsional angles for the α(2,6) linkage is reported. The three staggered rotamers, namely gt (ca. 60°), gg (ca. -60°) and tg (ca. 180°) around the C5-C6 bond. **b)** conformations in water of trisaccharide **1** along six MD simulations (50 ns) in the free state. The cluster in red corresponds to the conformation with a H5(Neu5Ac)-CH_3(Gal) distance < 5 Å. c) conformational analysis of φ and ω dihedrals.

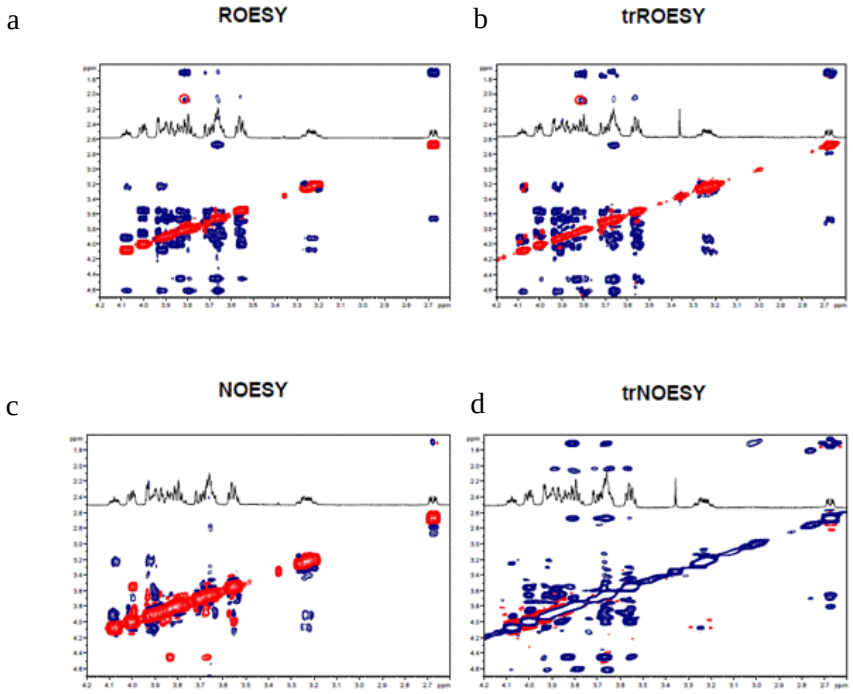


Figure 3.8 Analysis of the conformational behaviour of trisaccharide 1 by NMR. NOESY/ROESY analysis. In the upper panel, ROESY and tr-ROESY spectra (left) of 1, at 298 K, in the free (mixing time of 600ms) and bound (mixing time of 250 ms) state respectively. In the lower panel, NOESY and tr-NOESY spectra of 1 at 298 K, in the free (mixing time of 600 ms) and bound (mixing time of 250 ms) state respectively. In the NOESY spectra a change in the sign of NOE was detected upon h-CD22 addition in solution.

It is worth to remark that the *gt* preference induces the back-folding of the sialic acid residue, leading to a bent conformation, also known as the umbrella form and characterized by a θ angle < 110° (Figure 3.9).[211,212]

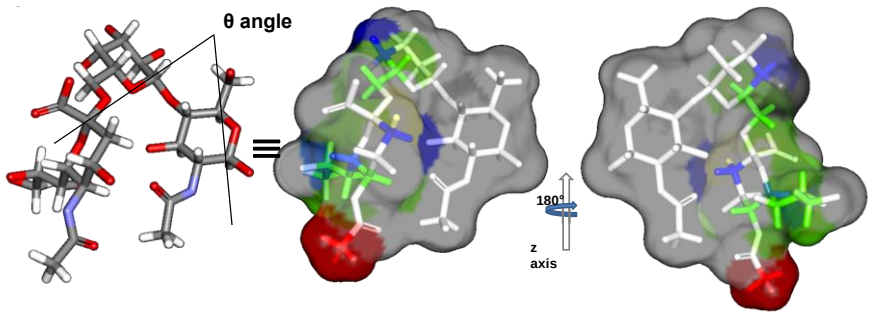


Figure 3.9 Trisaccharide 1 bioactive conformation upon CD22 binding. STD-derived epitope mapping on the molecular envelope of trisaccharide in its bioactive conformation, with color code according to the observed STD effects. The bent conformation of **1** was described by ϕ/ψ/ω torsion angles of -60°/180°/60° for the Neu5Acα (2,6) Gal glycosidic linkage and by the topological angle θ of 110°, defined by the C2, C1, and C1 atoms of the residues Neu5Ac, Gal, and GlcNAc, respectively, going from the nonreducing end to the reducing end.

The comparison of the NOE derived internuclear proton distances of **1**, estimated both in free and bound state, attested that the sugar adopted the above-described bent conformation even upon interaction with the h-CD22 (Table 3.2). Therefore, it can be stated that the protein recognizes the solution conformation of the glycan instead of the ligand and receptor undergoing a "bind and fit" process.

Table 3.2 Experimental and calculated proton distances of ligand 1.
Experimental (NOESY, estimated error 5–10%) *inter*-proton distances for the trisaccharide1 in the free and bound state. The calculated (average from MD simulation) distances for the trisaccharide in the free state are reported in the last columns.

Distance	Exp. Free state	Exp. Bound state	Calc. φ = -60°	Calc. φ = 180°
B1-B5	2.56	2.55	2.6	2.6
B1-A4	2.52	2.4	2.4	2.4
B6S-B5	2.97	2.89	2.9	2.9
K3ax-B6R	4.1	4.1	4.2	2.3
K3ax-B6S	4.2	4.3	4.4	2.5
K3eq-B6R	/	/	4.6	3.37
K3eq-B6S	/	/	4.9	3.8
K3ax-K5	2.7	2.6	2.6	2.6
A1-A5	2.57	2.56	2.6	2.6

To yield a reliable structural model for the h-CD22 and trisaccharide **1** complex in solution docking and MD simulations studies were performed. The crystal structure of human CD22[195] was considered (PDB ID: 5VKM). First, the structure was refined and docking calculations of **1** inside the h-CD22 binding site were performed with AutoDock4.2. The best docked poses were selected according to their ability to represent STD and NOE data. The chosen complexes were submitted to MD simulation and interactions between sialic acid, galactose, and *N*-acetyl glucosamine moieties of **1** with h-CD22 residues were monitored. Some

representative MD complexes were then subjected to calculations with the CORCEMA-ST program. The best fit between the experimental and calculated values corresponded to an R factor of 0.25, was displayed by the h-CD22/**1** model reported in Figure 3.10.

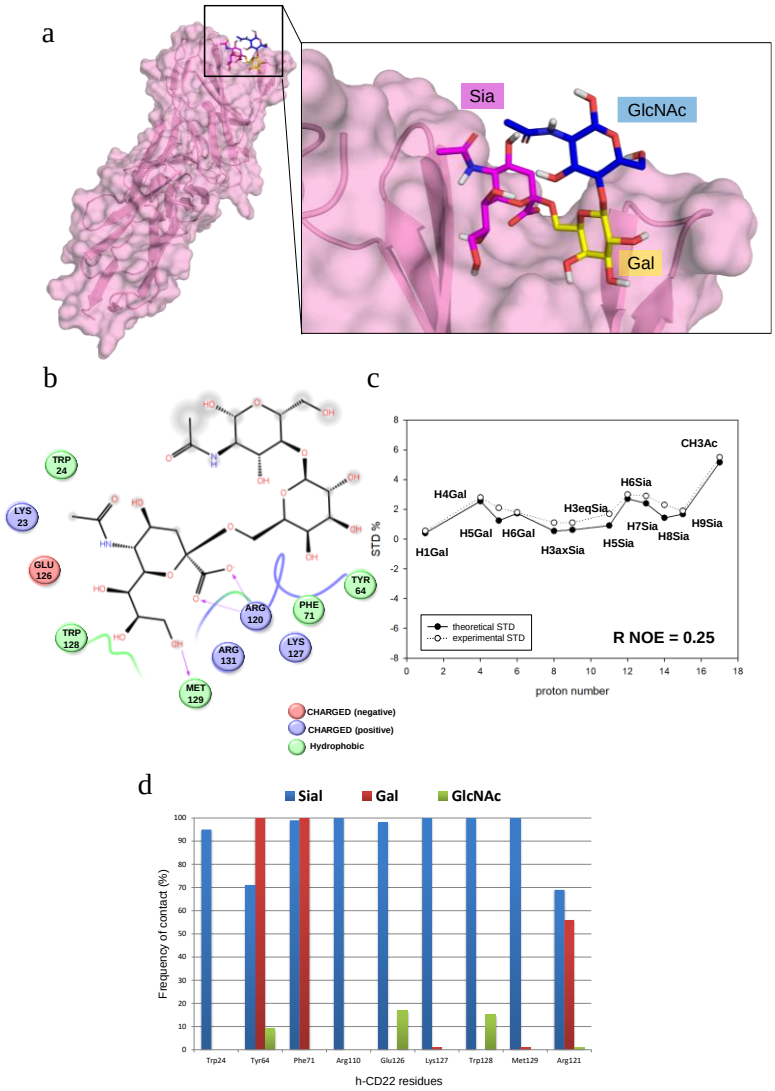


Figure 3.10 Docking/MD and CORCEMA-ST analysis. a) Best binding pose (STD and tr-NOESY based) for the docking and MD simulations of the trisaccharide **1** into binding pocket of h-CD22. **b)** Two-dimensional plot from a representative frame of the MD simulation showing the main interactions of the sialylated trisaccharide with the binding site residues of CD22. Dotted arrows show hydrogen bonds with functional groups from side chains and solid arrows such with functional groups of the backbone. **c)** Comparison between experimental (dashed line) and theoretical (solid line) STD data for the best model of trisaccharide **1** /h-CD22 model as obtained from CORCEMA. **d)** Contact frequencies (distance ≤5 Å) between trisaccharide **1** and binding site residues calculated from 5 to 50 ns of the MD simulations of the best complex. Only residues with contact frequencies ≥ 50% are shown.

In detail, the CORCEMA-ST prediction of the best model showed that protons from both sialic acid and galactose units exhibited STD effects, confirming their involvement in the interaction with h-CD22. The strong STD value of the methyl group of the N-acetamide moiety of sialic acid was consistent with the close contacts between this moiety and Trp24, Trp128 and Glu126 side chains along the simulated trajectory (Figure 3.9d). This enclosing was further promoted by the stable hydrogen bond established between sialic acid NH and Lys127 carbonyl group. Regarding STD data of the sialic acid glycerol side chain, the higher STD effect observed for the H7 in comparison with H8 and H9 seems to be prompted by the close CH-π interaction with Trp128. Although the methylene group at position 9 of the sialic acid also establishes an interaction with this residue, only one of the methylene protons was oriented towards Trp128 indole group, while the other H9 was solvent exposed, as can be observed in Figure 3.9a, hence exhibiting a lower STD effect. In addition, the predicted STD values for the *N*-acetylglucosamine moiety confirmed that it was not in intimate contact with the protein since they were negligibly small.

It is worth to note that the STD effects predicted by CORCEMA-ST when submitting other 3D complexes, mainly differing for the ω and ϕ torsional angles

around Neu5Ac-Gal glycosidic linkage of the ligand (data not shown), were very far from the experimental STD enhancements, with consequently higher R factors. Hence, CORCEMA-ST analysis further suggested that these models could be excluded as plausible conformations. Thus, NMR results (STD and NOE derived), docking, MD simulations and CORCEMA-ST analysis, allowed the selection of the final optimal h-CD22/**1** model.

3.2.3 Molecular recognition of complex-type N-glycans by h-CD22

Given the importance to investigate the interaction processes by using natural ligands that commonly decorate the surfaces of mammalian cells, the study of the binding between h-CD22 and naturally occurring complex-type N-glycan, typically found attached to glycoproteins, as many IgGs,[213]was carried out by NMR and molecular modelling. The computational investigations were performed by Prof. Sonsoles Martín Santamaría group, and further details can be found in ref.[214] In detail, the disialylated biantennary N-glycan **2**, features 6'sialyllactosamine type glycan at terminal ends of both branches (symmetric glycan, Figure 3.3).

First, the interaction between h-CD22 and **2** was probed by STD NMR and other ligand-based NMR techniques (Figure 3.11). Moreover, to resolve the huge signal overlapping and accurately map the interacting epitope, 2D STD NMR experiments, including STD-TOCSY (Figures 3.11a,b), were carried out, and a quantitative analysis, based on the construction of STD build up curves, allowed to establish the epitope mapping of the biantennary N-glycan **2** in the interaction with Siglec-2 (Figure 3.11e). The analysis of the STD enhancements highlighted the contribution of the sialic acid and galactose residues to the recognition process. In detail, the strongest STD enhancements belonged to the acetyl group as well as to the proton at position 6 of the sialic acid residue. The galactose

moiety also gave rise to STD signals, in particular the protons at position 4 and 6, whereas, interestingly, no STD response was observed for all the other sugar unis. These results indicated that the epitope recognized by the h-CD22 was limited to the disaccharide moiety at the terminal end of the N-glycan, with both sugar units significantly contributing to the recognition process.

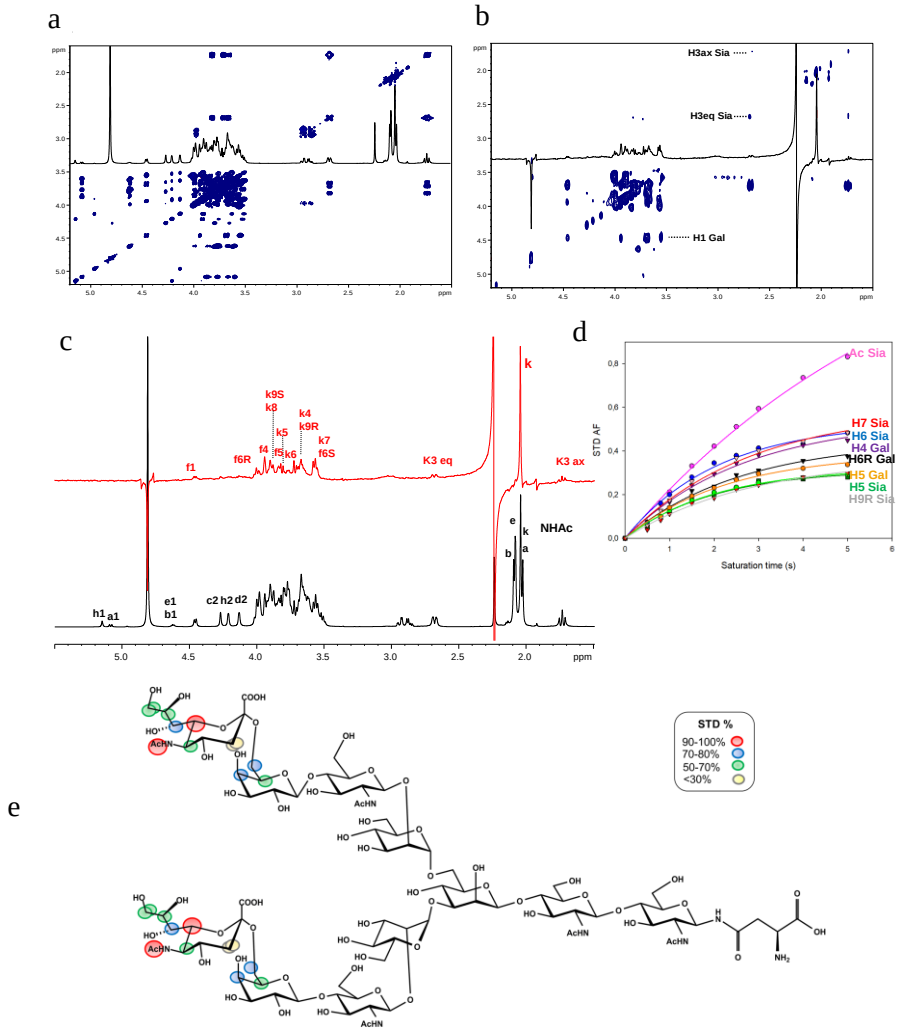
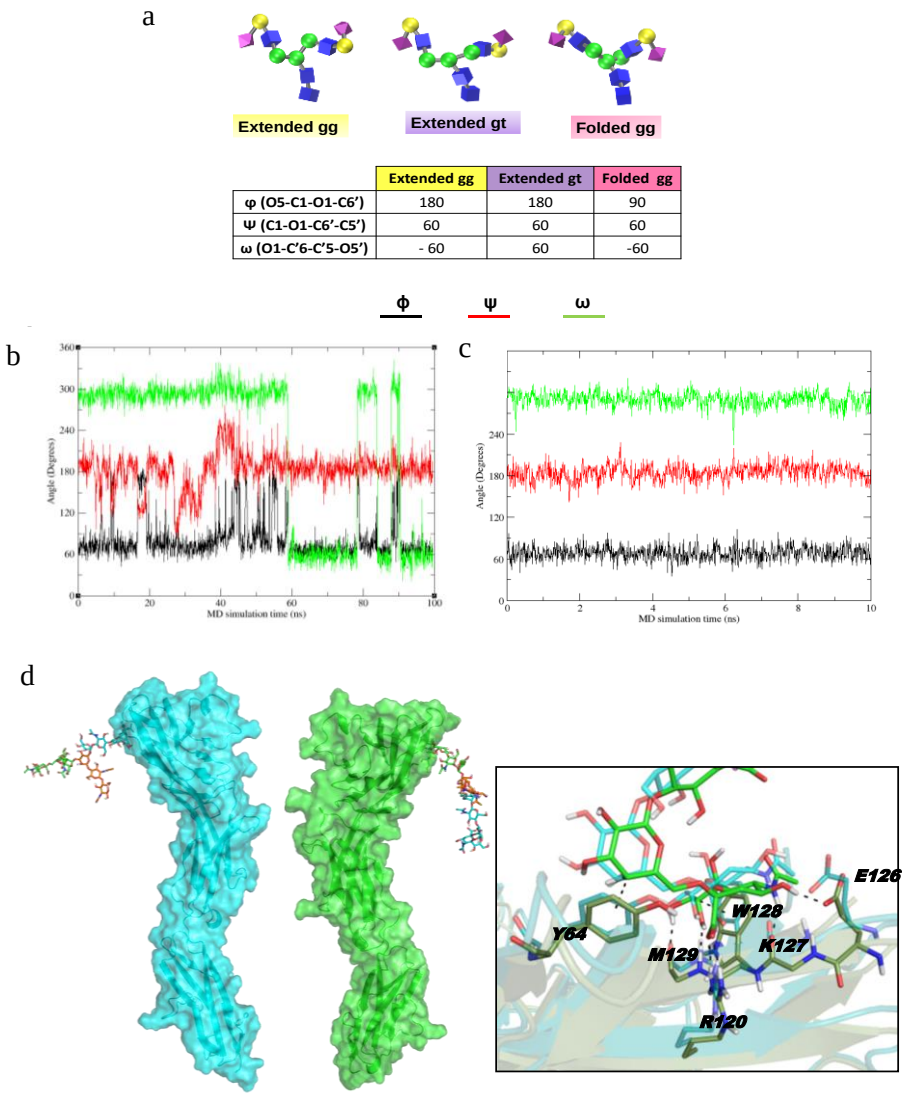


Figure 3.11 STD NMR analysis of the interaction between h-CD22 and glycan 2. 2D TOCSY and reference ^{1}H NMR spectra of glycan **2** alone in solution **a**), and 2D STD-

TOCSY with STD 1D NMR spectrum **b**) of mixture of h-CD22/glycan **2**. **c)** STD 1D NMR spectrum (top) and the corresponding reference ^{1}H NMR spectrum (bottom) of mixture of h-CD22/glycan **2**. **d**) STD build up curves of some protons of glycan **2**. **e)** Epitope map of ligand **2** in the interaction with h-CD22

Computational studies were then undertaken to assess the dynamic behaviour of **2** and to investigate its possible conformational changes upon binding. Concerning the free state, it is worth to note that several conformational ensembles of biantennary *N*-glycan in solution have been described. The glycan chain is known to experience different conformational states, broadly grouped in the "folded" and the "extended" forms, related to the flexibility around the α–Man-(1,6)-β-Man glycosidic linkage (Figure 3.12).[215,216,217]The MD simulation highlighted that the conformational flexibility of **2** was predominantly given by the ψ torsion of this linkage, which alternated along the MD simulation between different values. Further, glycan **2** showed a preference for the *gg* conformation of the ω torsion of the α-Man-(1,6)-β-Man linkage. As for the bound state, a preference for the extended conformers around the α-Man-(1,6)-β-Man linkage was observed (Figure 3.12c).



	Extended gg	Extended gt	Folded gg
φ (O5-C1-O1-C6')	180	180	90
Ψ (C1-O1-C6'-C5')	60	60	60
ω (O1-C'6-C'5-O5')	- 60	60	-60

Figure 3.12 Molecular recognition of biantennary N-glycans by hCD22. **a)** The three major conformations in solution of the biantennary undecasaccharide **2** that differ from the value of the ω and ψ torsions of the Man-α-(1,6)- Man linkage: the extended *gg* and *gt* and the folded *gg*. The glycosidic torsional angles definition were: ω= O1-C'6-C'5-O'5, ϕ= O5-C1-O1-C6', ψ= C1-O1-C6'-C5'. The SNFG nomenclature has been used. **b)** Man-α-(1-6)-Man Dihedral fluctuation along MD simulation (free-state) **c)** Man-α(1,6)-

Man dihedral fluctuation along MD simulation (bound state) **d)** Complexes of **2** with h-CD22 *via* its 1-6 branch (green) and its 1-3 branch (blue). Interactions between one of the Neu5Ac-α-(2,6)-Gal-β moiety and the protein are represented with dashed lines. On the right, a snapshot from MD simulation showing H-bond interactions (dashed lines) between each branch of undecasaccharide **2** (blue and green, respectively) and h-CD22 binding site residues is reported.

Given the biantennary character of **2**, MD simulations highlighted the potential of this glycan to simultaneously bind two h-CD22 molecules ideally located in proximity on the cellular surface (Figure 3.13).

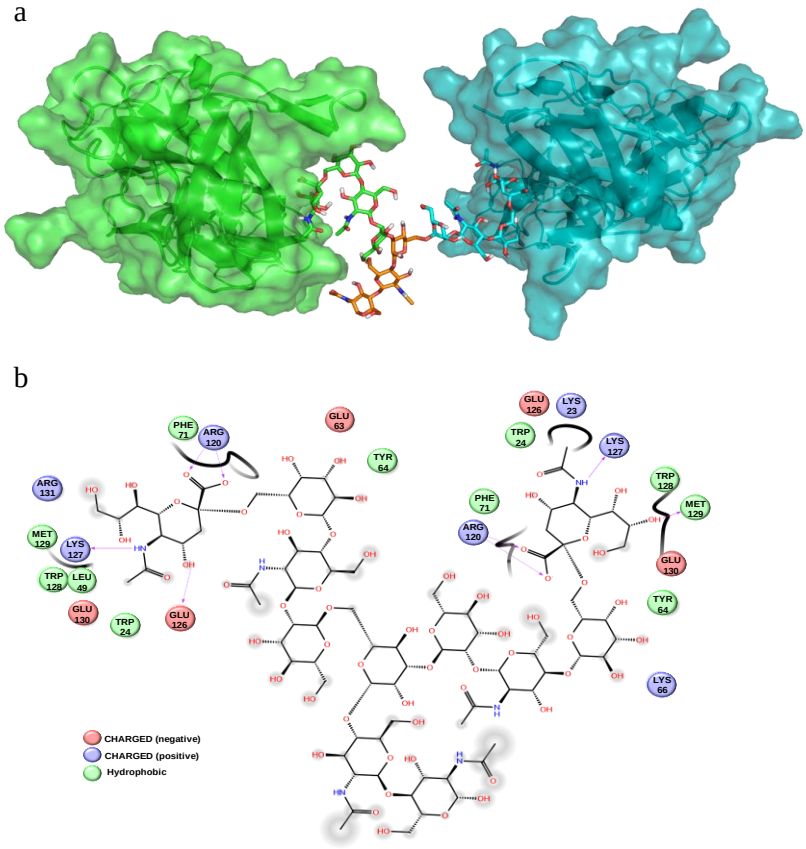


Figure 3.13 MD simulations of the ternary complex. a) Top view of the ternary complex model $(h\text{-}CD22)_2$ **-2** extraccted from MD simulation. The extended conformation of the

glycan allowed the simultaneous binding with two h-CD22 proteins. **b)** Two-dimensional plot taken from a representative frame of the MD simulation illustrating the interactions of the sialylated undecasaccharide with residues in the binding site of CD22. The residues shown are close to the ligand and participate in hydrophobic and polar interactions.

These results show that a bivalent binding of glycan **2** to two h-CD22 proteins could be possible when belonging to the glycosylation pattern of another h-CD22 molecule. The STD-NMR analysis performed on the asymmetric ligands **3** and **4** (Figure 3.14), confirmed that both antennae of **2** could equally bind h-CD22 *via* the corresponding Neu5Ac-α-(2,6)-Gal-β moiety in a stable way and with an almost identical epitope. The potential simultaneous binding of glycan **2** to two h-CD22 molecules was therefore evaluated.

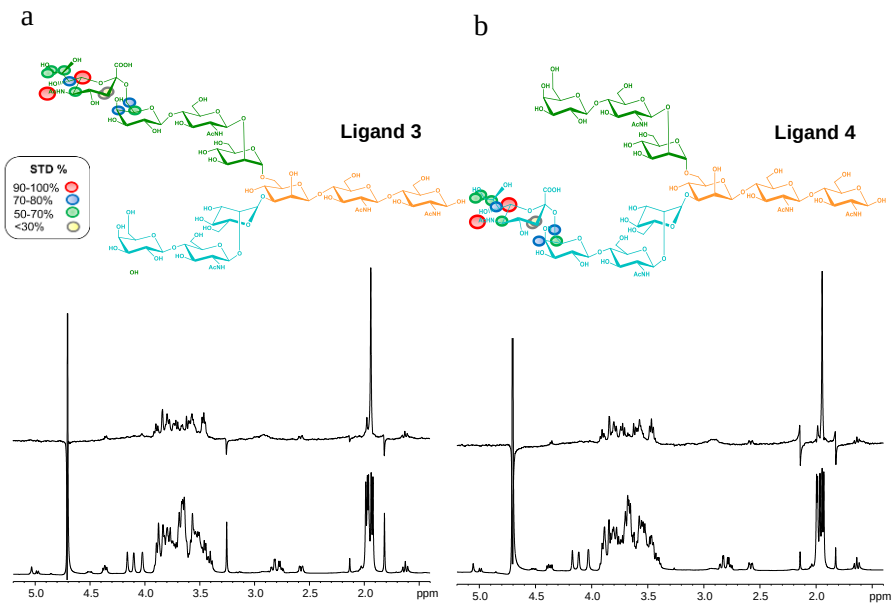


Figure 3.14 STD NMR analysis of glycan 2 and 4 binding to h-CD22. a) STD NMR analysis of the binding between h-CD22 and the glycan **3**.**b)** STD NMR analysis of the binding between h-CD22 and the glycan **4**.

These results, in agreement with computational data, indicated that h-CD22 can accommodate, with the same binding mode, the sialic acid – galactose epitope whether it is presented on one glycan branch or on the other further suggesting that it does not discriminate between the two different glycan antennae (Figure 3.14).

The ability of **2** to interact with two h-CD22 molecules simultaneously when belonging to the glycosylation pattern of another h-CD22 molecule was also explored by MD simulations. (Figure 3.15).

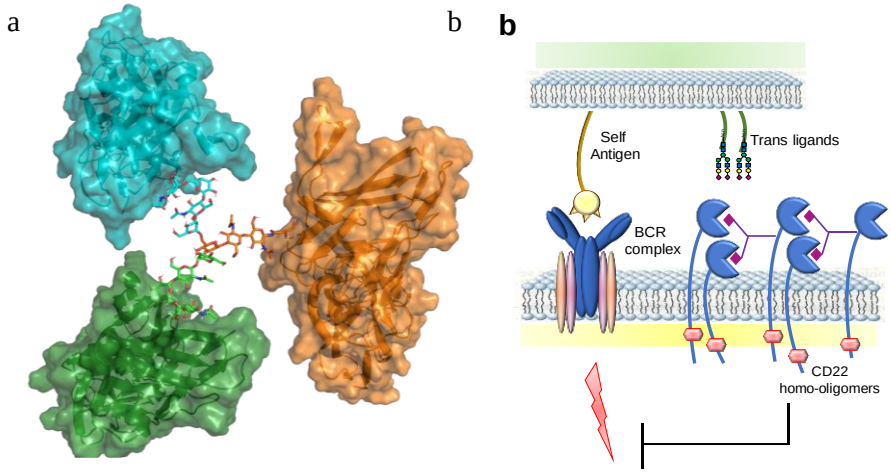


Figure 3.15 Model of B-cell receptor surface. a) Top view of the quaternary complex $(h\text{-}CD22)_2$ -**2**/(h-CD22) from MD simulation. The extended conformation of the glycan allowed the simultaneous binding with two h-CD22 proteins. **b)** Representation of the quaternary complex (h-CD22)-**2**/$(h\text{-}CD22)_2$ The conformation of biantennary N-glycans allows the simultaneous interactions with two molecules of CD22 and facilitates the formation of CD22 homo-oligomers. The binding with *trans* ligands, which decorates membranes that expose both «self» antigens and Siglec ligands, recruits CD22 to the BCR and inhibit the B-cell activation

To gain further insights into the ligand specificity of h-CD22, the recognition features of h-CD22 to differently sialylated oligosaccharides were investigated (ligands **5**-**6**). STD NMR binding experiments conducted on the biantennary α2,3 sialo-undecasaccharide **5** (Figure 3.3) revealed no binding, confirming the selective recognition of α(2,6) sialylated glycans by h-CD22 (data not shown). Among the glycans modification the core fucosylation on GlcNAc, has attracted considerable attention for its involvement in several physiological events including immune response, signal transduction and cell adhesion.[218]

Thus, to understand if the presence of α1,6 linked fucose in the sialylated dodecasaccharide **6** (Figure 3.3) influenced the presentation of the N-glycan to h-CD22, thus modulating the interaction, STD NMR analysis on the mixture h-CD22- dodecasaccharide **6** was performed. The STD NMR spectrum clearly showed that the human receptor CD22 recognized the fucosylated N-glycan **6**. However**,** no differences in the STD enhancements were observed by comparing it to the STD NMR spectrum of the mixture h-CD22-undecasaccharide **2**. These data suggested that, in the case under study, the core fucosylation did not change the binding mode of the Siglec-2 when interacting with sialylated N-glycans (data not shown).

3.3 Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22

The interaction between different sialoglycans bearing acetylated or glycolylated sialic acid, namely Neu5Ac and Neu5Gc and human (h-CD22) and murine CD22 (m-CD22) was here investigated by ligand based NMR techniques and molecular modelling. As the crystal structure of m-CD22 is not known, homology modelling studies were carried out using as templates the crystal structure of h-

CD22 (see above). The work has been carried out in collaboration with Prof. Sara Sattin (University of Milano), Prof. Koichi Fukase (University of Osaka, Japan) and Prof. Paul D. Crocker (University of Dundee, UK).

3.3.1 Binding specificity of m- and h- CD22 toward Neu5Ac and Neu5Gc ligands

The binding affinities of both m-CD22 and h-CD22 toward Neu5Gc and Neu5Ac containing ligands were evaluated by fluorescence analyses. In detail, fluorescence titrations of increasing amounts of sialoglycans into a fixed concentration of the proteins were performed. The results confirmed the ability of m- and h-CD22 to recognize acetylated and glycolylated sialoglycans (Figures 3.16). As shown in the Figure 15, the derived values of binding constants (Kb), all in the micromolar range, were comparable suggesting that the proteins similarly recognized all the sialylated conjugates tested.

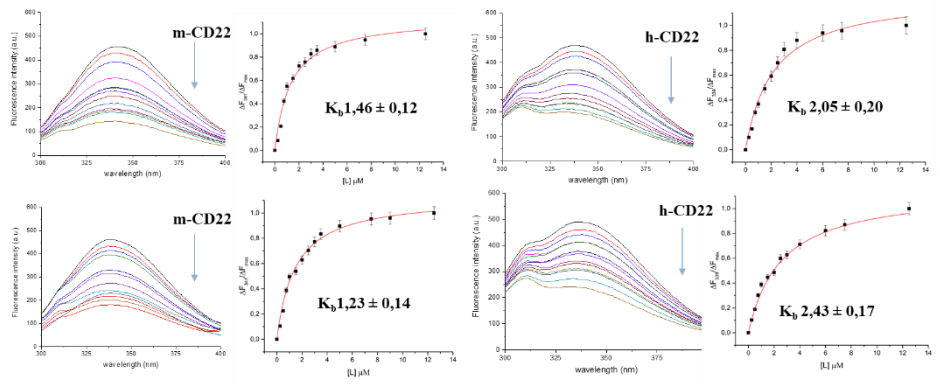


Figure 3.16 Intrinsic fluorescence quenching titrations. Fluorescence spectra of m-CD22 (upper panel, black lines) or h-CD22 (lower panel, black lines) in the presence of increasing amounts of Neu5Gc containing ligand (colored lines) or Neu5Ac containing ligand (colored lines), respectively. The binding isotherm and the values of the binding constants (K_b) are also reported.

3.3.2 STD NMR analysis of Neu5Ac and Neu5Gc containing glycans by human and murine CD22.

A detailed STD NMR analysis (Figure 3.17) described the ligand portion entailed in the binding with m-CD22, providing the epitope mapping of the acetylated trisaccharide [Neu5Ac-α-(2,6)-Gal-β-(1,4)-GlcNAc], 6'SLN (trisaccharide **1**).

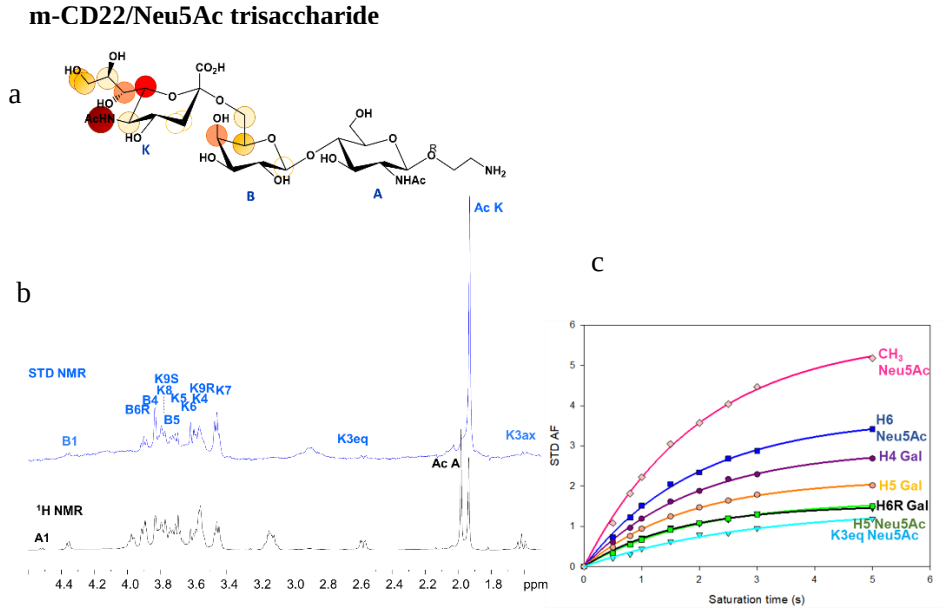


Figure 3.17 STD analysis of Neu5Ac-α-(2,6)-Gal-β-(1,4)-GlcNAc-β–OR interacting with murine CD22. a) Epitope map of the acetylated ligand interacting with h-CD22. STD effects lower than 10% are not indicated in the epitope mapping. **b)** STD NMR analysis of the trisaccharide bound to m-CD22 using a protein/ligand molecular ratio of a 1:100. The STD NMR spectrum obtained in the presence of h-CD22 was nearly identical to that obtained in the presence of h-CD22 **c)** STD build up curves derived from STD spectra acquired at different saturation times.

The fingerprint of the STD NMR spectrum clearly indicated the specificity of the binding between m-CD22 and the sialylated ligand (Figure 3.17b), revealing relevant differences in the multiplicity and relative intensity of the STD signals (blue spectrum) compared with the corresponding off-resonance (black spectrum). The Neu5Ac (**K**) mostly participated to the interaction with m-CD22; interestingly, its acetyl group was the highest STD signal in the spectrum, highlighting its strong involvement in the binding with the receptor protein, while the acetyl group of the glucosamine residue (**A**) disappeared from the STD spectrum. Additionally, H6 proton of Neu5Gc gave a good STD signal, close to 70%, followed by the protons belonging to the glycerol chain and the H5 (range of 30-50%); diastereotopic (two atoms in a molecule which, if replaced, would generate compounds that are diastereomers, i.e. non-mirror image stereoisomers) protons at position 3 ($H3_{ax}$ and $H3_{eq}$) of the Neu5Ac instead contributed to the binding at a lower extent (less than 30%). Moreover, a good amount of saturation transfer was also detected to Gal (**B**), mainly to protons H4, H5 and H6. Similarly to h-CD22, the multiplet around 3.9 ppm in the off-resonance, deriving from the overlapping of the resonances of H6 **A** and H6 **B**, was converted into a triplet corresponding to the only H-6 GlcNAc **A** in the STD spectrum. This was also a further evidence that the *N*-acetylglucosamine moiety was excluded from the recognition process and gave no STD signals. These data were supported by the quantitative analysis obtained by the construction of STD build up curves. (Figure 3.17c and Table 3.3).[219]

Table 3.3 Experimental STD intensities of the acetylated ligand bound to m-CD22. STDmax were evaluated by fitting the data to a monoexponential equation: STD = STDmax * (1 – exp (- ksat * tsat)).

^{1}H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH_3 Neu5Ac	5.7320	0.4884	2.7995	**100**
H6 Neu5Ac	3.7105	0.5080	1.8849	**67.3**
H4 Gal	2.9244	0.5187	1.5169	**54.2**
H5 Gal	2.1719	0.5580	1.2119	**43.3**
H6R Gal	1.5247	0.6170	0.9407	**33.6**
H5 Neu5Ac	1.6331	0.5256	0.8584	**30.6**
$H3_{eq}$ Neu5Ac	1.4283	0.3632	0.5188	**18.5**

The comparison of the results obtained on m-CD22 with the data previously published on h-CD22 highlighted a totally comparable binding mode of Neu5Ac epitope between the two proteins. In both cases, experimental results evidenced the major participation of the sialic acid – galactose moieties in the interaction with h- and m-CD22, while the full absence of STD signals belonging to the *N*-acetylglucosamine residue demonstrated the complete exclusion from the binding process when interacting either with m-CD22 or with h-CD22.

The hydroxylated derivative of *N*-acetylneuraminic acid, the *N*-glycolylneuraminic acid Neu5Gc binding to human and murine CD22 was analyzed by STD NMR. The resulting experiments are showed in Figures 3.18 and 3.19. It can be observed that the glycolylated ligand similarly interacted with both receptors as indicated by the same recognition pattern. The highest signal in the STD spectra belonged to the glycolyl moiety of Neu5Gc, whose signal was set to 100%; the other protons contributing to the binding were normalized to this value (Tables 3.4 and 3.5). The sialic acid – galactose moiety of the glycolylated ligand were

again the main determinant of the binding to both h-CD22 and m-CD22. In detail, H7 of Neu5Gc was saturated more than 50% as well as H-5 and H-4 protons of galactose residue; STD values around 40-50% were reached by most of the protons belonging to the Neu5Gc-Gal unit. Further, the STD signals of the $H3_{eq}$ and $H3_{ax}$ showed the lowest effects for both receptors. The GlcNAc residue (**A**) was completely excluded from the CD22 binding pocket, as already shown with Neu5Ac containing glycans, confirming the solvent exposure of the third sugar. The additional data gathered from the construction of STD build-up curves confirmed the results obtained from the qualitative STD NMR analysis (Figures 3.18 and 3.19).

m-CD22/Neu5Gc trisaccharide

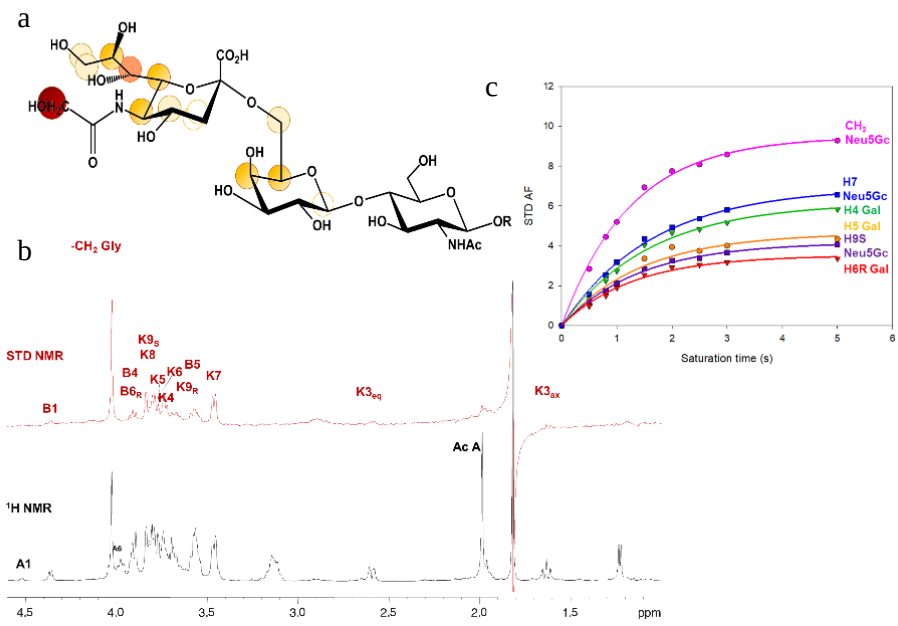


Figure 3.18 STD analysis of Neu5Gc-α-(2,6)-Gal-β-(1,4)-GlcNAc-β–OR interacting with murine CD22. a) Epitope map of the acetylated ligand interacting with m-CD22. STD effects lower than 10% are not depicted in the epitope mapping. **b)** STD NMR analysis of the glycolylated trisaccharide bound to m-CD22 using a protein/ligand molecular ratio of a 1:100. The STD NMR spectrum obtained in the presence of h-CD22 was totally

comparable to that obtained in the presence of m-CD22 **c)** STD build up curves derived from STD spectra acquired at different saturation times.

Table 3.4 Experimental STD intensities of glycolylated ligand bound to m-CD22. STDmax were evaluated by fitting the data to a monoexponential equation: STD = STDmax * (1 – exp (- ksat * tsat)).

^{1}H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH_2 Neu5Gc	9.4963	0.8017	7.6132	**100**
H7 Neu5Gc	6.9388	0.6038	4.1897	**55.0**
H4 Gal	6.1757	0.6239	3.8530	**50.6**
H5 Gal	5.7199	0.6108	3.4937	**45.9**
H1 Gal	4.6474	0.6891	3.2025	**42.1**
H9S Neu5Gc	4.1764	0.7117	2.9723	**39.0**
H4 Neu5Gc	6.2336	0.4618	2.8787	**37.8**
H6R Gal	3.5266	0.7722	2.7232	**35.8**
H9R Neu5Gc	3.5970	0.6542	2.3531	**30.9**
H3 Gal	4.0638	0.5758	2.3399	**30.7**

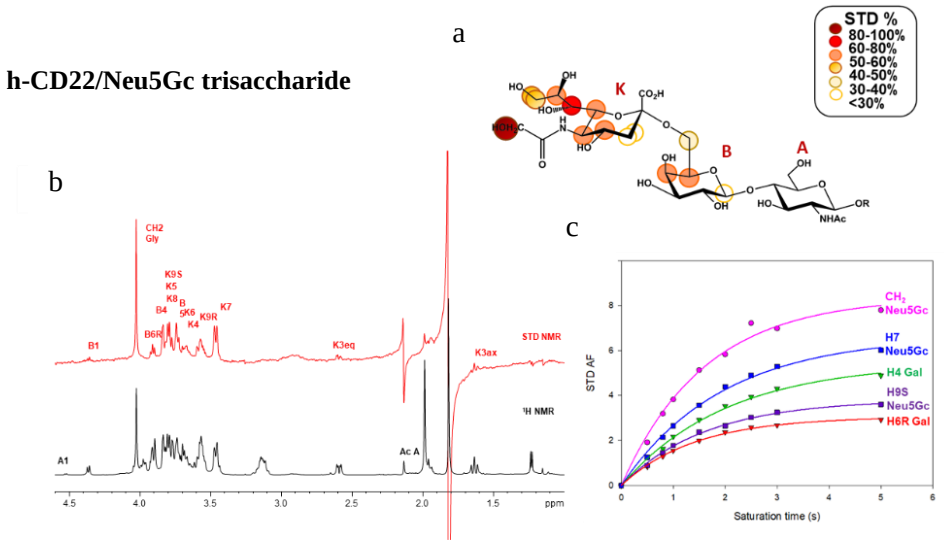


Figure 3.19 STD analysis of Neu5Gc-α-(2,6)-Gal-β-(1,4)-GlcNAc-β–OR interacting with human CD22. a) Epitope map of the glycolylated ligand interacting with h-CD22. STD effects lower than 10% are not depicted in the epitope mapping. **b)** STD NMR analysis of the glycolylated trisaccharide bound to h-CD22 using a protein/ligand molecular ratio of a 1:100. The STD NMR spectrum obtained in the presence of h-CD22 was totally comparable to that obtained in the presence of m-CD22 **c)** STD build up curves derived from STD spectra acquired at different saturation times.

Table 3.5 Experimental STD intensities of glycolylated ligand bound to h-CD22.

STDmax were evaluated by fitting the data to a monoexponential equation: STD = STDmax * (1 – exp (- ksat * tsat)).

^{1}H	STD_{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH2 GLY	8.3838	0.6205	5.2021	**100**
H7 Neu5Gc	6.6503	0.5120	3.4049	**65.4**
H5 Gal	5.1684	0.5315	2.7470	**52.8**
H4 Gal	5.4644	0.4935	2.6967	**51.8**
H1 Gal	3.6867	0.6772	2.4966	**48.0**
H9S Neu5Gc	3.8238	0.6113	2.3375	**44.9**
H9R Neu5Gc	3.6682	0.6236	2.2875	**44.0**
H6R Gal	3.0547	0.7042	2.1511	**41.3**
H3 Gal	3.9695	0.4872	1.9339	**37.2**

Key information about the topology and conformation adopted by glycolylated *N*-glycans when interacting with h- and m-CD22 (the bioactive conformation) was achieved by tr-NOESY analyses. The ligand containing the Neu5Gc–α-(2,6)-Gal glycosidic linkage is defined by three different torsion angle: ϕ (C1-C2-O-C'6), ψ (C2-O-C'6-C'5) and the additional ω (O−C'6−C'5−O'5).[220] As previously reported for Neu5Ac containing glycans, the *gt* conformer (ω value ~60°) was the most populated also for the glycolylated trisaccharide. Hence, the key NOE established between the acetyl group of GlcNAc residue and H5 of Neu5Gc (Table 3.6) observed in the tr-NOESY spectra (Figures 3.20 and 3.21) revealed a bent conformation of the ligand assuming a shape characterized by an umbrella-like topology when bound either to h-CD22 and m-CD22.[209]

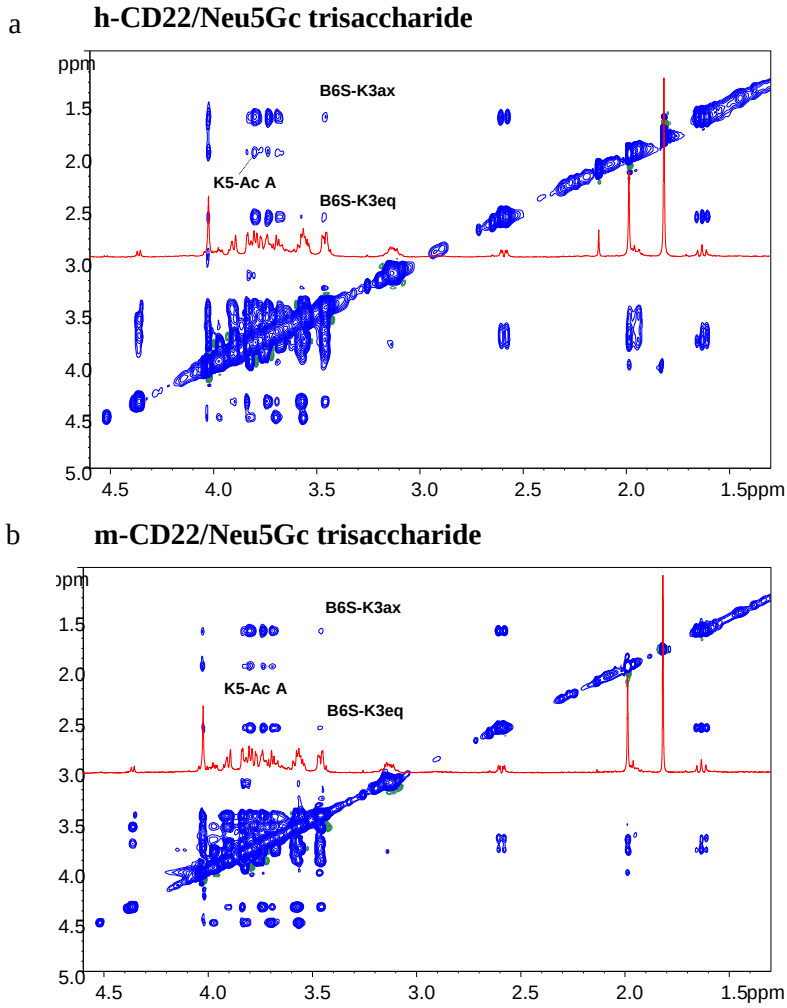


Figure 3.20 Tr-NOESY analysis of h-and m-CD22 binding to h- and m-CD22 a) Tr-NOESY NMR spectrum of the glycolylated trisaccharide in the bound state with h-CD22, using a mixing time of 400ms. **b)** Tr-NOESY NMR of the glycolylated trisaccharide bound to m-CD22, using a mixing time of 400ms. The ligand 6'SLN upon binding with both h- and m-CD22 adopts an umbrella-like topology.

Table 3.6 Experimental (NOESY, estimated error 5–10%) *inter*-proton distances for the glycolylated in the free and bound state. The calculated (average from MD simulation) distances for the trisaccharide in the free state are reported in the last columns.

Distances	Family I Φ = -60° Ψ = 180° ω = 60°	Family II Φ = 180° Ψ = 180° ω = 60°	Exp Lig. free state	Exp. h-CD22 bound state	Exp. m-CD22 bound state
$H3_{eq}$ Neu5Gc - H6S Gal	4.93	3.84	4.50	4.61	4.72
$H3_{eq}$ Neu5Gc - H6R Gal	4.58	3.37	nd	nd	nd
$H3_{ax}$ Neu5Gc - H6S Gal	4.43	2.53	4.11	4.80	4.90
$H3_{ax}$ Neu5Gc - H6R Gal	4.25	2.35	nd	nd	nd
H5 Neu5Gc - CH_3 GlcNAc	4.30	9.60	nd	4.96	4.81

The stability of ϕ and ψ dihedral angles of the glycolylated trisaccharide in the free state was also monitored during MD simulations (Figure 3.21).

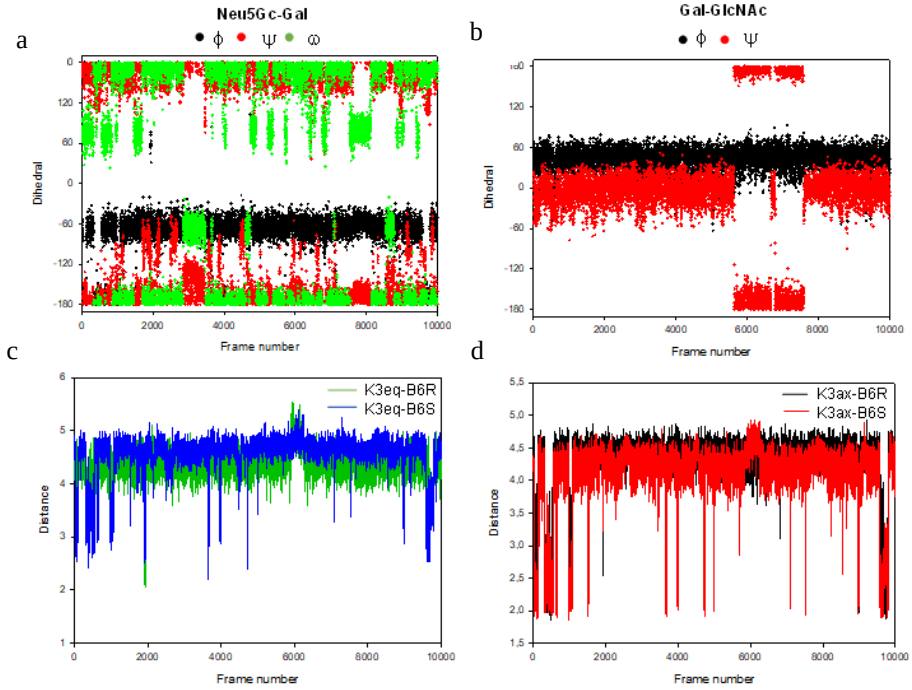


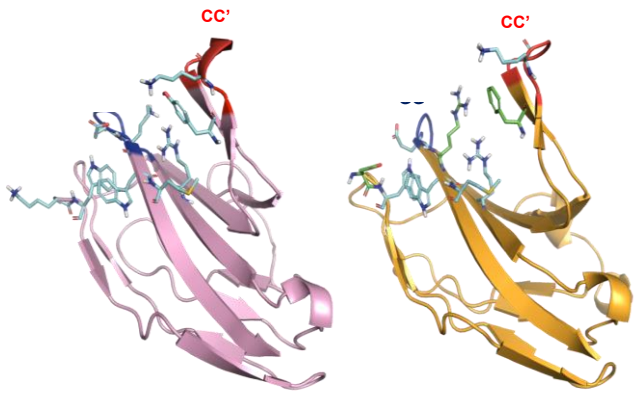
Figure 3.21 MD simulation analysis of the glycolylated trisaccharide in its free state a) φ/ψ/ω dihedral angles of Neu5Gc-Gal linkage along the MD trajectory. **b)** φ/ψ/ dihedral angles of Gal-GlcNAc linkage along the MD trajectory. **c)** $H3_{eq}$ Neu5Gc – H6S/H6R Gal intra-ligand distances**. d)** $H3_{ax}$ Neu5Gc – H6S/H6R Gal intra-ligand distances.The torsion angles were defined as follows: ϕ (C1-C2-O-C'6), ψ (C2-O-C'6- C'5), ω (O−C'6−C'5−O'5).

While the acetylated trisaccharide in solution can explore different populations depending on the values of ϕ torsion angle (-60°/180°),[214] the glycolylated ligand preferentially adopts a conformation with ϕ around -60° (see Table 3.6 and Figure 3.21). Furthermore, the tr-NOESY analyses in the presence of human and murine CD22 (Figure 3.21) confirmed the preference of the glycolylated glycan for the energetic minimum characterized by ϕ/ψ dihedral angles of -60°/180; the absence of NOE contacts between the H6R of galactose and the diastereotopic (axial an equatorial) H3 protons of Neu5Gc confirmed the above conformation (Table 3.6).

3.3.3 Molecular overview of the interaction of glycolylated glycans with murine and human CD22

Computational studies including homology modelling, docking and MD simulations were carried out to describe the interaction between CD22 and Neu5Gc ligands. The 3D structure of m-CD22, not available, was modelled based on h-CD22 structural template.[195]First, m-CD22 sequence encoding for m-CD22 extracellular V-set and C2-set domains was aligned to the template sequence[195] (PDB ID: 5VKM) using BLAST. According to the sequence alignment displayed in Figure 21a, m-CD22 showed significant sequence identity (above 58%) relatively to h-CD22, in agreement with the conserved nature of CD22. The target template alignment was submitted to SWISS-MODEL server to obtain m-CD22 three-dimensional structure. Focusing on the obtained m-CD22 structural model, the typical Siglec Ig-like folding can be noted, with the sialic acid binding site located at the summit of the *N*-terminal V-set domain (Figure 3.22).

b



hCD22	mCD22
Lys23	Asp25
Trp24	Trp26
Tyr64	Phe68
Lys66	Lys72
Arg120	Arg124
Glu126	Glu130
Lys127	Arg131
Trp128	Trp132
Met129	Met133

Figure 3.22 BLAST alignment of the extracellular regions of murine CD22 and human CD22. Key amino acids are highlighted in blue and Cys forming disulfide bridges in green. Sequence corresponding to the V-set domain is evidenced in pink, sequence of C2 – set domains in purple. Conservation between the two sequences is evaluated using Jalview[221] **b)** Comparison of the *N*-terminal V-set domains of h-CD22 (pink), PDB ID: 5VKM, and m-CD22 homology model (orange). Common residues constituting the binding sites are highlighted in cyan. Residues of m-CD22 pocket differing from h-CD22 are colored in green. A direct comparison of the binding site residues can be found in the table on the right.

Similarly to other Siglecs,[94] m-CD22 binding site architecture featured a shallow pocket constituted by the A, F and G strands and bound by the CC' and GG' variable loops. Compared to h-CD22, the composition of the binding site residues entitled to sialylated epitopes binding was overall conserved. The most relevant differences lied in the replacement of Lys23, Tyr64 and Lys127 with Asp25, Phe68 and Arg131 in m-CD22, respectively, which slightly affects the polarity of the binding region. Prior to docking calculations, the structural model stability was assessed by MD simulation using AMBER 18 package, the root mean square fluctuation of carbon alpha (RMSF) and the mobility of the variable CC' loop and GG' loop were monitored along the trajectory (Figure 3.23). The analysis of the

potential energy along the simulation allowed to identify the m-CD22 structure at lowest energy, which was subsequently employed for docking calculations.

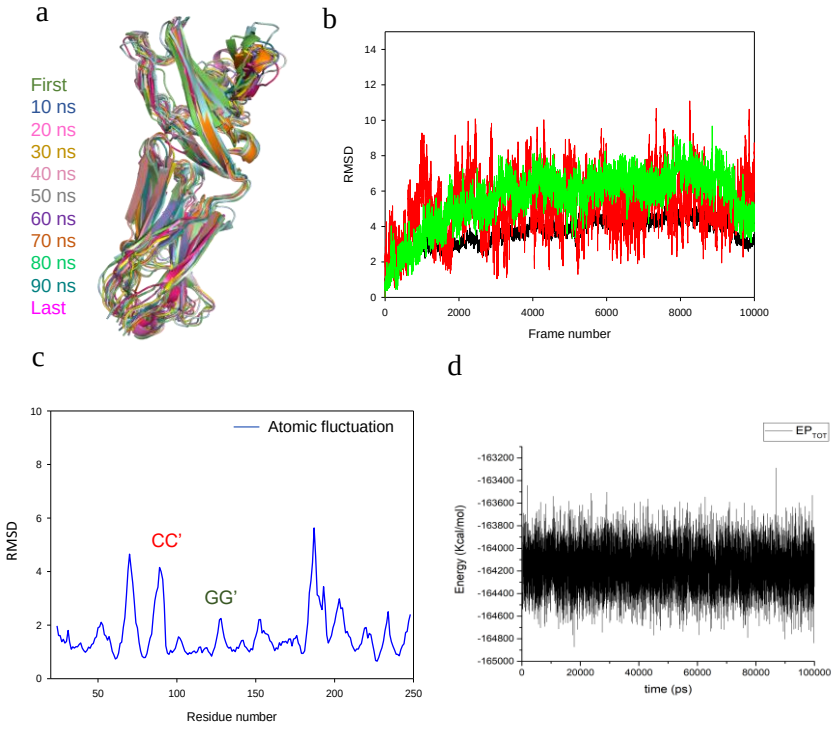


Figure 3.23 MD simulation analysis of m-CD22 homology modelling. The peaks in the RMSF plot corresponded to the mobile loops connecting the ß-strands, in both V-set and C2-set Ig-like domains. **a)** Superimposition of the m-CD22 structures each 10 ns of the MD simulations. Along the MD simulation, no relevant conformational changes emerged. **b)** Backbone root mean square fluctuation (RMSD) of the protein, CC' loop (res 69-74), GG' loop (res 127-130), depicted in black, red and green respectively. The fluctuations of the backbone RMSD of the CC' loop, can be attributed to a dynamic equilibrium between a disordered (high RMSD) and partially ordered (low RMSD) forms of the region. **c)** Atomic fluctuation of the structure by residue, calculated using the protein Cα atoms. The peaks in the RMSF plot corresponded to the mobile loops connecting the ß-strands, in both V-set and C2-set Ig-like domains. **d)** Plot of the potential energy variation of m-CD22 structure along the MD. The structure with the lowest potential energy was considered for the docking calculations.

To analyze its binding mode, the Neu5Gc ligand was docked into h-CD22 and m-CD22 binding sites by means of AutoDock 4.2. As result, for both receptors, Neu5Gc ligand binding poses with favourable theoretical energies were found. Also, among the two receptors, the energies and populations of the top clusters were very close, ranging from −3.5 to −2 kcal mol^{-1} (Table 3.7). Thus, based on energy and cluster populations, two promising h-CD22/ and m-CD22/ligand complexes were chosen.

Table 3.7 Cluster rank, Cluster population, computed binding energy and RMSD (Root Mean Square Deviation) for the molecular docking (AutoDock) of m-CD22/ligand and h-CD22/ligand complexes.

Complex	Cluster Rank	No cluster conformations	Estimated free energy of binding (kcal/mol)	RMSD from reference structure (Å)
h-CD22	1	154	-2.57	2.37
m-CD22	3	85	-1.94	3.23

Notably, Neu5Gc ligand displayed a similar binding mode inside the receptors pocket, in accordance with NMR data, showing the involvement of the following major determinants of sialylated ligands binding, i.e. the conserved arginine (Arg120 and Arg124 for h-CD22 and m-CD22) and aromatic residues (Trp24, Trp128 and Trp26, Trp132 for h-CD22 and m-CD22).[195]It is worth to note that in the selected binding modes the ligand assumed the *gt* conformation in the chosen clusters.

To finely describe the interaction between Neu5Gc containing glycans and h- and m-CD22, the aforementioned structures were used as starting point to run MD simulations throughout 100 ns. Along the trajectory, the receptor and

ligand RMSD, the ligand dihedral angles fluctuation, as well as hydrogen bonds and contacts between the ligand and the receptor were monitored (Figure 3.26). In both complexes, the ligand remained anchored to both h-CD22 and m-CD22 receptors until the end of the simulation, as demonstrated by the ligand RMSD values within ~1.5/2 Å, further suggesting the stability of the binding poses.

The final models of the interactions between Neu5Gc containing ligand and h-CD22/m-CD22 were further validated by means of the CORCEMA-ST program, that allowed the comparison between the theoretical and the experimental STD data for each protein ligand complex by means of the R-NOE factor value. To this purpose, different structures of h-CD22 and m-CD22 complexes were extracted from the MD simulations and then considered for CORCEMA-ST calculations. The three-dimensional complexes showing the best fit between theoretical and experimental data, resulting in R-NOE values of 0.24 and 0.21 respectively, are depicted in Figures 3.24 and 3.25.

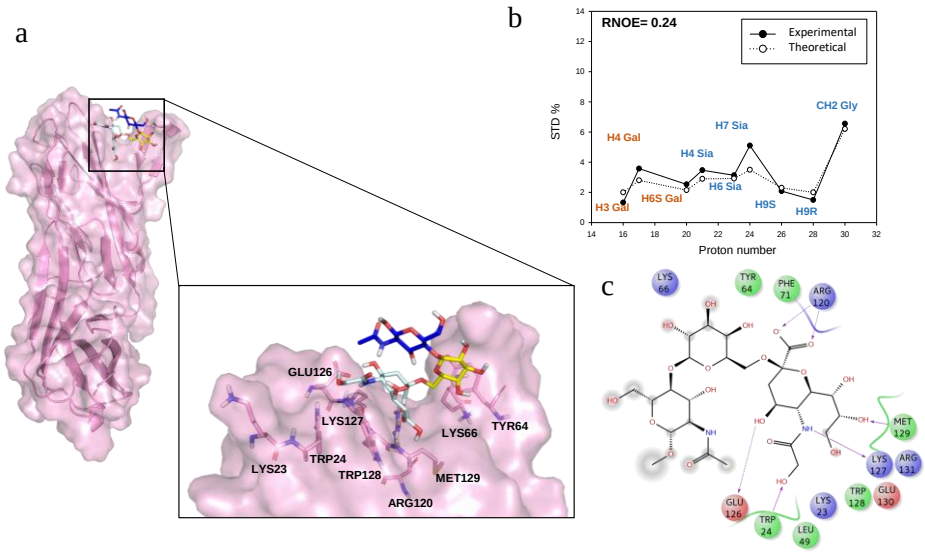


Figure 3.24 Interaction between h-CD22 and Neu5Gc ligand. a) Three-dimensional model of Neu5Gc ligand bound to h-CD22 V-set domain as derived by STD, tr-NOESY and MD. **b)** The three-dimensional h-CD22/Neu5Gc complex showing the best fit between theoretical (solid line) and experimental (dashed line) STD data derived by CORCEMA-ST analysis. (R-NOE values of 0.24). **c)** Two-dimensional plots representing the interactions between the glycolylated trisaccharide and h-CD22 binding site residues, derived from a representative frame of the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone.

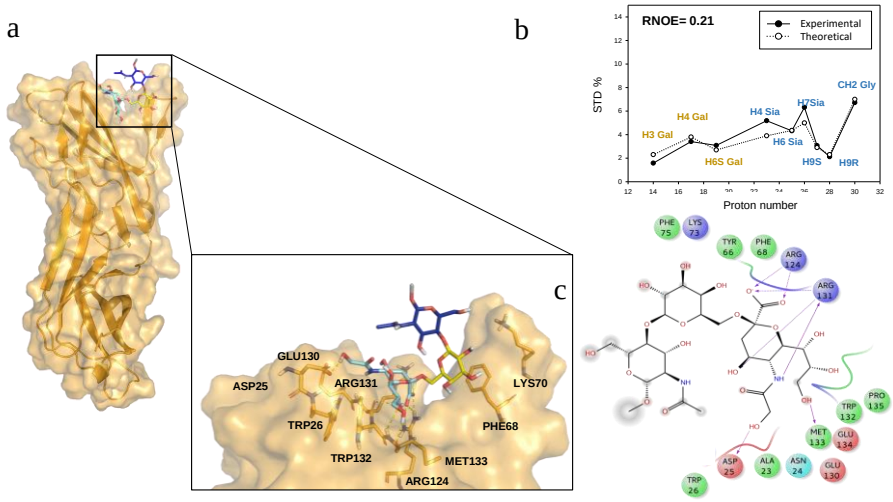


Figure 3.25 Interaction between m-CD22 and Neu5Gc ligand. a) Three-dimensional model of Neu5Gc ligand bound to m-CD22 V-set domain as derived by STD, tr-NOESY and MD. **b)** The three-dimensional m- CD22/Neu5Gc complex showing the best fit between theoretical (solid line) and experimental (dashed line) STD data derived by CORCEMA-ST analysis. (R-NOE values of 0.21). **c)** Two-dimensional plots representing the interactions between the glycolylated trisaccharide and m-CD22 binding site residues, derived from a representative frame of the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone.

Concerning h-CD22/Neu5Gc ligand final complex, the Neu5Gc containing glycan established several strong interactions with the receptor binding pocket residues, retained for most of the time of the simulation (Figure 3.26).

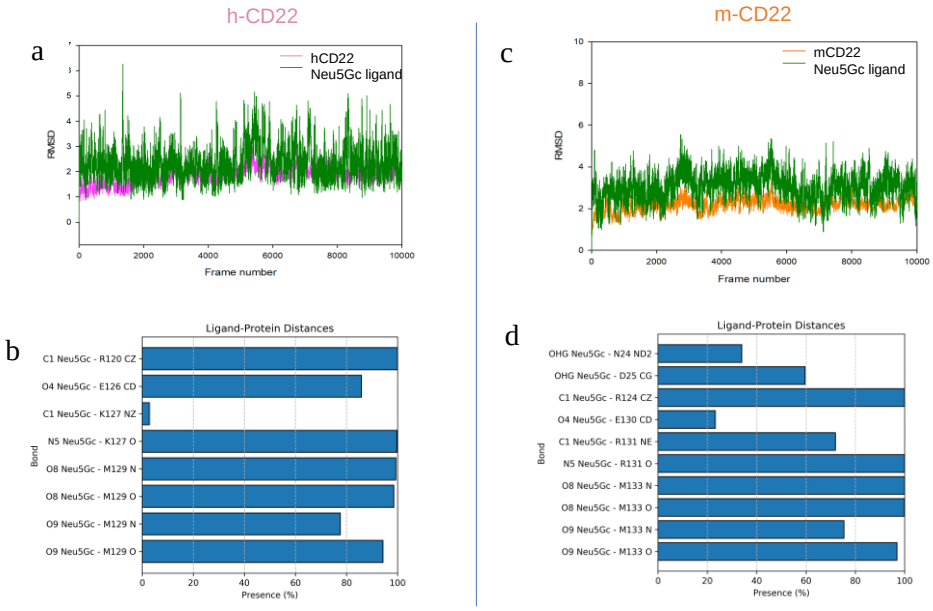


Figure 3.26 MD simulation analysis of h-CD22/m-CD22 complexes with the glycolylated ligand a) h-CD22 and Neu5Gc ligand RMSD variation along the MD. The ligand RMSD was measured with respect to the protein. **b-c)** Frequency of most representative h-CD22/Neu5Gc ligand distances. A distance cut-off of 5Å was considered for the calculation. m-CD22 and Neu5Gc ligand RMSD variation along the MD. A distance cutoff of 5Å was concsidered for the calculations

In particular, the polar network between h-CD22 and the glycolylated ligand was similar to that described for the corresponding Neu5Ac ligand. Indeed, the highly conserved Arg120 established a salt bridge between its guanidine group and Neu5Gc carboxylate (Figure 3.24). Neu5Gc glycerol moiety was involved in hydrogen bonds with Met129 backbone oxygen and amide, as well as in CH–π interactions with Trp128. Also, the hydroxyl group at position 4 of Neu5Gc formed a strong polar interaction with Glu126 residue. Most importantly, the *N*-glycolyl group of Neu5Gc engaged a stable hydrogen bond with Lys127 and hydrophobic contacts between its methylene protons and Trp24 and Trp128 aromatic moieties. It is worth to note that no relevant interactions or contacts specifically involving

the hydroxymethyl group of the *N*-glycolyl moiety were found. Gal unit contributed to the receptor binding by means of CH–π interaction with Tyr64 aromatic residue and polar contacts with Lys43. On the contrary, the GlcNAc unit was far from the h-CD22 surface for the most part of the simulation. Devoid of key interactions with the receptor, this sugar moiety displayed high mobility; despite this, the distance between the CH_3 group of the *N*-acetyl glucosamine and the H-5 of Neu5Gc assumed an average value of 4.9 Å (data not shown), in accordance with the NOE derived distance (Table 3.6).

According to the CORCEMA-ST (Figure 3.24b), the strongest STD effects in the h-CD22 complex were predicted for protons belonging to Neu5Gc unit; significant saturation was also estimated for some protons of Gal unit, although no saturation was observed for protons of GlcNAc units, in full agreement with the experimental STD data. The high STD value of the glycolyl group of the ligand was consistent with the close contacts with Tyr24, Trp128, Glu126 and Lys127 side chains. Regarding STD data of the Neu5Gc glycerol moiety, the higher STD effect observed for the H7 is due to the strong CH-π interaction of this proton with Trp128 indole group. Concerning the hydroxymethylene group at position 9, only the H9S was oriented towards Trp128, thus exhibiting a higher STD effect with respect to H9R. Also, H4 and H6 of Neu5Gc displayed significant STD effects for their vicinity to the receptor surface. For the Gal unit, considerable saturation was predicted especially for the proton in position 4, due to its close contact with Tyr64 aromatic ring.

As for m-CD22/Neu5Gc selected complex, the ligand interaction pattern showed many similarities with its human ortholog. Still, Neu5Gc majorly contributed to the binding, interacting with Arg124, Arg131, Trp132, Met-133 receptor residues, whereas no participation of GlcNAc residue was observed (Figure 3.25). Specifically, the carboxylate of Neu5Gc formed the key electrostatic interactions with

the Arg124 guanidine group. The Met133 entailed numerous polar interactions with the hydroxyl groups of the ligand glycerol moiety. Additionally, the Arg131 performed the same role as Lys127 in h-CD22 receptor, thus forming a solid hydrogen bond with the amide nitrogen of Neu5Gc *N*-glycolyl moiety, and simultaneously establishing a polar contact with the hydroxyl group at position 4. Differently from h-CD22 complex, the hydroxyl group of the *N*-glycolyl moiety was involved in hydrogen bonds with Trp26 and Asp25 pocket residues. Furthermore, similarly to h-CD22 binding mode, the Gal unit mostly entailed in CH–π interaction with Phe68 residue; also, the GlcNAc residue did not interact with m-CD22, exhibiting a higher degree of fluctuation with respect to the other residues. The average distance between the CH_3 group of the *N*-acetyl glucosamine and the H5 of Neu5Gc along the simulation was 5 Å, in agreement with the NOE derived distance, describing the umbrella topology (data not shown).

Considering the CORCEMA-ST of m-CD22 and h-CD22 in complex with Neu5Gc containing ligand, the orientation of the glycan inside the receptors binding pockets seemed to be very similar. Nonetheless, the predicted STD effects were higher for the m-CD22-ligand complex. In detail, the stronger STD value observed for the hydroxymethyl group of the *N*-glycolyl moiety of Neu5Gc could be explained by its contacts with Trp26, Trp132. Also, a great contribution to the higher STD effect was given by the interaction between Asp25 and the hydroxymethyl group of the *N*-glycolyl moiety. The importance of this interaction for *N*-glycolyl trisaccharide to m-CD22 recognition was confirmed by performing CORCEMA-ST analysis on another complex extracted from the MD simulation, having good potential energy but lacking the hydrogen bond involving the Asp25 residue. Indeed, the outcome resulted in higher R-NOE (0.7) due to the significantly lower STD value attributed to the *N*-glycolyl moiety protons (data not shown). Considering the Gal moiety, similar STD effects were predicted for the

protons directed toward the residue aromatic side chain, namely H4 and H3, despite the presence of Ph68 in place of Tyr64.

3.3.4 Comparing acetylated and glycolylated glycans in the interaction with murine and human CD22

To directly compare the mode of interaction of Neu5Ac/Neu5Gc with m-CD22, molecular modelling studies, including docking and MD simulations (Figure 3.27), were performed to establish a three-dimensional complex of m-CD22 and Neu5Ac glycans, thus defining a reliable model of interaction.

a b

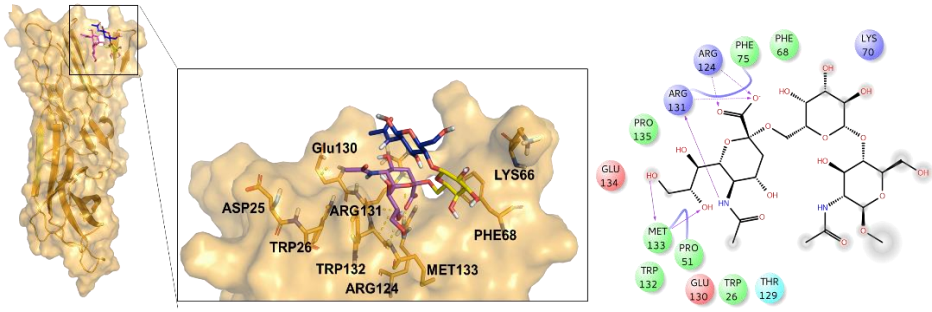


Figure 3.27 Interaction between m-CD22 and Neu5Ac ligand a) Three-dimensional model derived by STD, tr-NOESY and MD for the Neu5Ac ligand bound form (*gt* conformer) to m-CD22 homology model. The representative frame from the most populated MD cluster was considered to depict the complex. **b)** Two-dimensional plots representing the interactions between the glycolylated trisaccharide and the binding site residues of mCD22. The representative frame of the most populated MD cluster was considered to depict the complex.

According to our model, the Neu5Ac ligand displayed a similar orientation with respect to Neu5Gc ligand, establishing the crucial salt bridge with Arg124 through its carboxylate. The glycerol lateral chain hydroxyl groups interacted through hydrogen bonds with the Met133 backbone. The *N*-glycolyl group

was involved in the binding with Arg131 as well as hydrophobic interactions with Trp26 and Trp132 aromatic residues. The Gal residue, comparably to Neu5Gc ligand, Gal residue was involved in CH-pi interactions with Phe68, and the Glc-NAc was far from the receptor binding region. Thus, except for the additional hydrogen bond formed by the *N*-glycolyl moiety with Asp25 residue, it can be assessed that m-CD22 interacts with Neu5Ac and Neu5Gc ligands in analogous manner. Comparing the three-dimensional structures of the complexes, it was possible to note a slightly different shape and polarity of the receptor cavities which accommodate the Sia *N*-acetyl/*N*-glycolyl moieties (Figure 3.28 a,b).

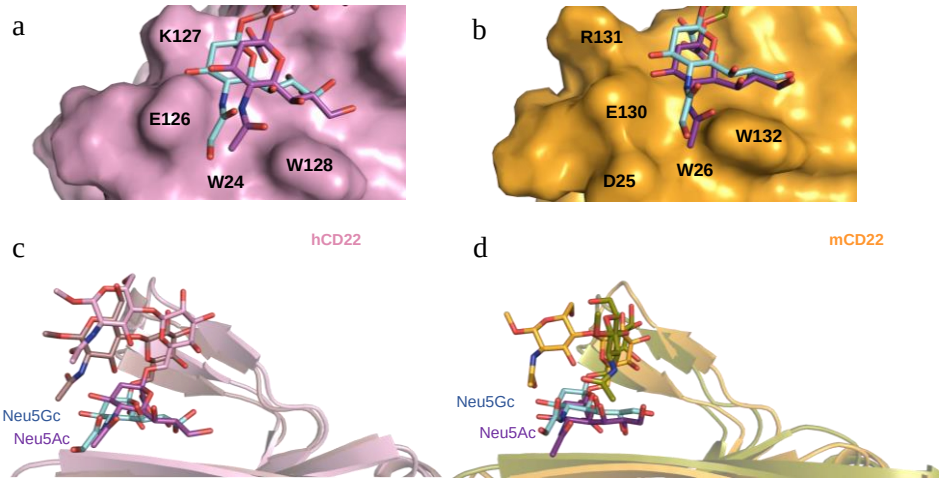


Figure 3.28 Comparison of the interaction of Neu5Ac/Neu5Gc ligands with m-CD22 and h-CD22. a) Close up view of *N*-acetyl (purple) and *N*-glycolyl (cyan) binding region of h-CD22, showing the protein surface (pink). **b)** Close up view of *N*-acetyl (purple) and *N*-glycolyl (cyan) binding region of m-CD22, showing the protein surface (orange). **c)** Superimposition of h-CD22/Neu5Gc (pink) and h-CD22/Neu5Ac complexes (dirty violet) **d)** Superimposition of m-CD22/Neu5Gc (bright orange) and m-CD22/Neu5Ac complexes (olive). NeuG5= NeuGc sialotrisaccharide, Neu5Ac= Neu5Ac sialothisaccharide.

Indeed, whereas the h-CD22 region responsible for *N*-acetyl and *N*-glycolyl binding is essentially constituted by aromatic residues, m-CD22 also comprises the

Asp25 residue in optimal position to interact with the longer *N*-glycolyl chain. On the other hand, the *N*-glycolyl hydroxymethilene group moiety does not interact with the h-CD22 receptor and is exposed to the solvent. Last, MM/GBSA and MM/PBSA analysis was performed using AMBER, to have an indication of the relative binding energy of the complexes (Table 3.8).

Table 3.8 Relative binding energies of h-CD22 and m-CD22 with acetylated and glycolylated ligands. All units are expressed in kcal/mol.

Complex	ΔGb (MM/GBSA)	ΔGb MM/PBSA)
h-CD22/Neu5Gc	-41.24 ± 0.15	-12.66 ± 0.16
h-CD22/Neu5Ac	-38.43 ± 0.18	-10.75 ± 0.18
m-CD22/Neu5Gc	-38.29 ± 0.16	-11.73 ± 0.19
m-CD22/Neu5Ac	-34.69 ± 0.26	-10.72 ± 0.18

As result, all four complexes exhibited comparable ΔGb values, in agreement with the similar binding properties of the receptors toward the different forms of sialic acids here discussed (Figure 28c,d). Therefore, it is possible to conclude that, despite some differences in the binding regions, human and murine CD22 similarly recognizes Neu5Ac and Neu5Gc ligands.

3.4 Molecular recognition of a sialic acid analog by CD22

In this section, the description of the binding features of a Sialyl Tn antigen (Sia-α-(2,6)-Galα–Thr) derivative in the interplay with h-CD22, achieved by means of NMR and molecular modelling, is portrayed. The sialylated analogue, depicted in Figure 3.29, has been developed by Prof. Cristina Nativi group (University of Florence), and the detailed synthetic procedure is described in ref. 222.

From the structural viewpoint, the compound is a constrained tricyclic glycoside that retains the α-O-glycosidic linkage at the Gal moiety. As this class of compounds exhibited interesting inhibitory activity toward other biological targets,[223,224] it was inspected whether the rigid galactoside moiety would affect the binding specificity of CD22, with the aim to identify novel ligands to employ within therapeutic and diagnostic purposes, as potential cancer vaccines with increased immunogenicity.

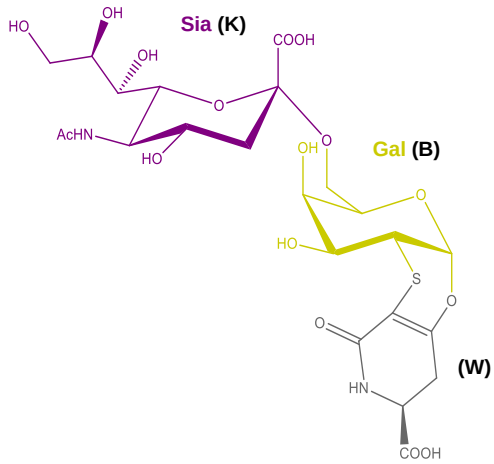


Figure 3.29 Structure of the studied STnThra analogue (STnThra). K= Nue5Ac B=Gal and W=Aglycon. Neu5Ac and Gal units were colored according to SNFG nomenclature.

At first, intrinsic fluorescence studies of h-CD22 upon STnThra mimetic addition, showed that the receptor tryptophane residues were quenched by the ligand, thus proving that the sialylated ligand formed a complex with CD22 receptor. From the fluorescence data analysis, the binding constant of the process was evaluated, showing value in the range of those obtained for the natural epitopes (Figure 3.30).

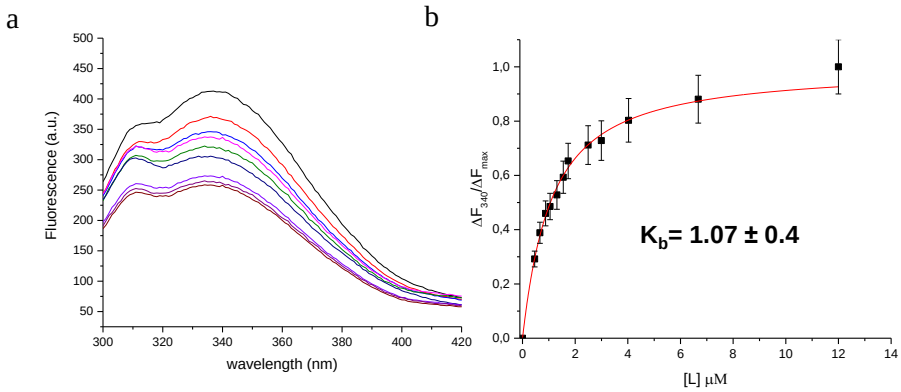


Figure 3.30 Binding affinity of STnThra ligand and h-CD22 by steady state fluorescence. a) Fluorescence spectra of h-CD22 (left panel, black line) in the presence of increasing amounts of STnThra solution (colored lines). **b)** Binding curve, the value of the binding constant (K_b) is reported.

Then, tr-NOESY and STD NMR have been employed to analyze the interaction between STnThra and the h-CD22. In detail, TROESY and tr-NOESY experiments were performed to obtain information on the bioactive conformation of the ligand (Figure 3.31). It is worth to note that the free state ligand NOESY intensity was close to zero, thus TROESY experiment was performed.

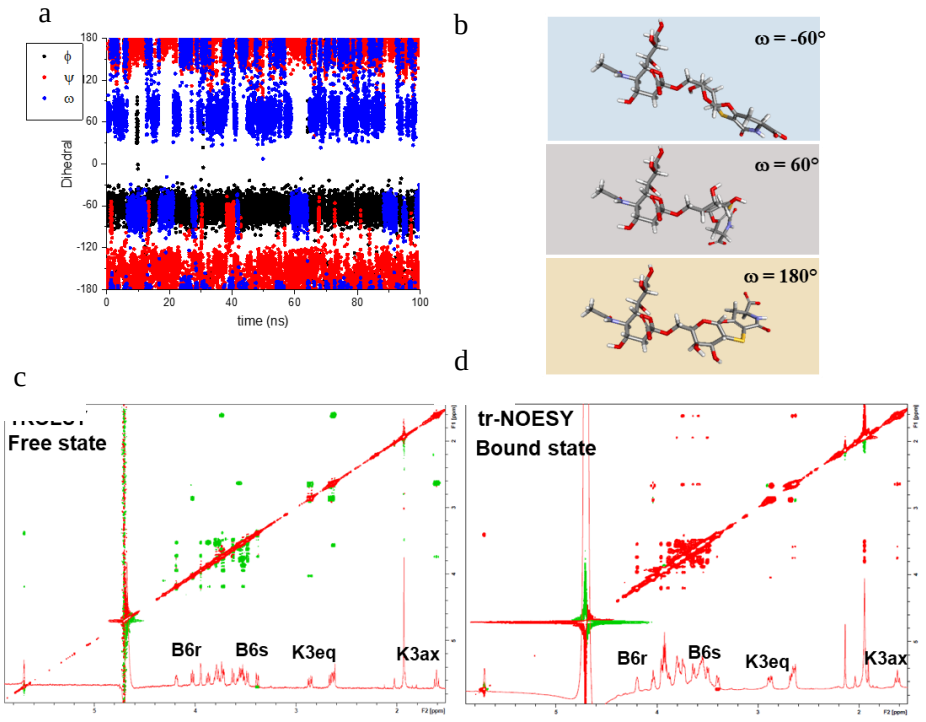


Figure 3.31 Conformational analysis of STnThra. a) Dihedral angle fluctuation around the Sia-Gal linkage along the MD simulation of free STnThra analog. **b)** 3D representation of STnThra analog conformations according to the three ω value 2D NOESY was also recorded but produced signals close to zero. **c)** TROESY on free state STnThra **d)** tr-NOESY of 1:30 h-CD22/STnThra mixture (T=298K, Tm= 300 ms).

The qualitative analysis of the free state TROESY and bound state tr-NOESY of the ligand allowed to deduce a similar conformation between the free and the bound state, and hence indicating that, upon binding to CD22, the bioactive conformation of the ligand was nearly identical to the most populated conformation in the free state. In detail, the absence of the cross-peaks relative to the B6R/BS6 and K3ax/K3eq demonstrated that the glycosidic linkage between Neu5Ac-Gal preferentially assumed a ϕ angle value of -60 (see previous

paragraph, Table 3.6). It is worth to note, that, due to the rigid nature of W moiety, and differently from the natural ligand, the conformation is not bent but rater extended, independently from the Neu5Ac-Gal dihedral ω values, therefore no inter-residues distances, diagnostic for ω value, were experimentally observable (Figure 3.31). Thus, to gain further conformational information, a 3D model of the ligand was built, and its dynamic behavior was explored by MD simulations using AMBER, confirming the selection of ϕ glycosidic torsion angle of -60, and showing a fluctuation of the ω value along the simulation in the free state (Figure 3.31).

To determine the binding epitope of the Neu5Ac analogue, relative STD effects have been calculated from the STD amplification factors (Table 3.9, Figure 3.32). The analysis of the STD NMR spectrum allowed to infer that the Ne5Ac derivative was recognized by h-CD22, and to map the interaction epitope (Figure 3.32).

Table 3.9 Experimental STD intensities of the sialylated analogue bound to h-CD22.

Proton	STD epitopes fit%	Proton	STD epitopes fit%
B1	40,7	**K6**	70,4
B5	66,1	**K4**	57,3
Wb	38,0	**K9R**	51,5
B4	67,1	**B6S**	49,5
B6R	54,1	**K7**	81,4
K8	59,6	**B2**	47,1
K9S	50,6	**Wa**	38,8
K5	60,2	**Wa'**	26,8
B3	68,3	**K3eq**	23,2
K3ax	31,9	**HNAc**	100,0

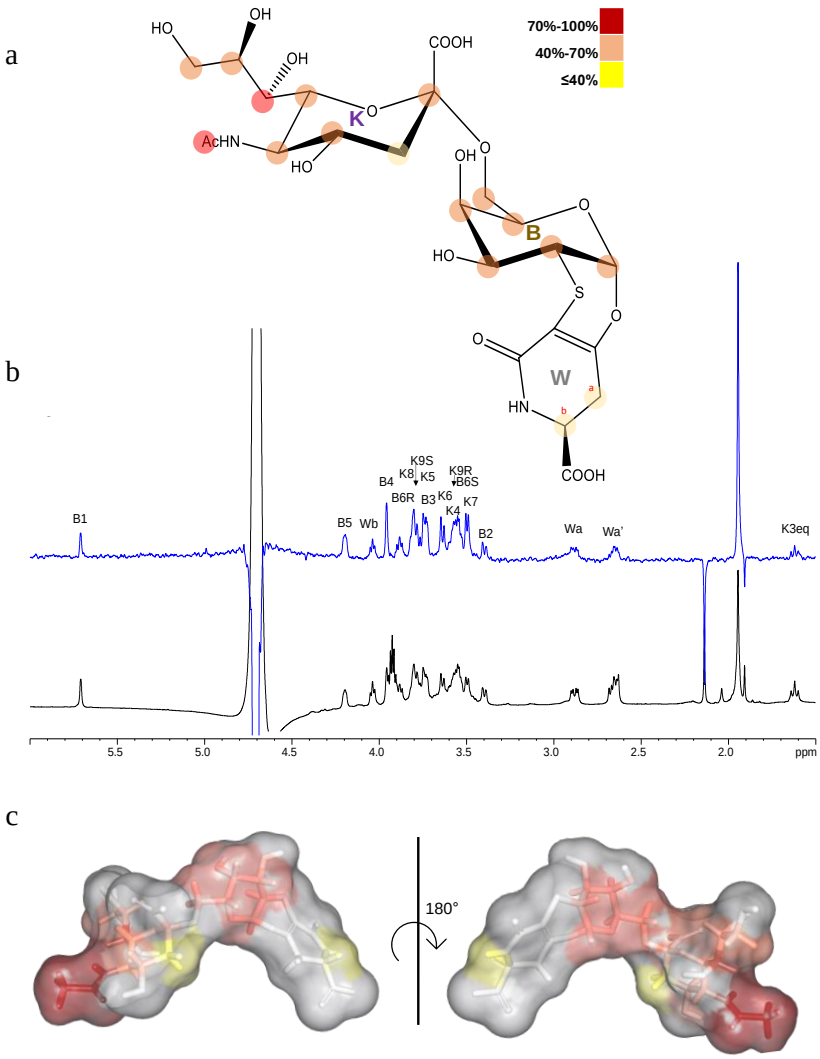


Figure 3.32 STD analysis of STnThra interacting with human CD22. a) Epitope map of the ligand. **b)** STD NMR analysis of the STnThra upon the interaction with h-CD22 using a protein/ligand molecular ratio of a 1:100. **c)** 3D representation of the ligand in the bioactive conformation obtained by tr-NOESY with molecular surface colored according to STD enhancements.

From the analysis of the relative enhancements in the STD-spectrum, a strong contribution from the Neu5Ac moiety was detected, with the highest observed for its Acetyl group. Furthermore, the involvement of the Neu5Ac lateral chain was deduced from the strong enhancements (between 50% and 70%) observed for K5, K6, K7, K8 and K9. Significant increases were exhibited by all protons of the Gal residue, in particular for H3, H4, H5. Interestingly, also the Wa and Wa' of the ligand showed STD enhancements even if to a lesser extent. This observation suggested that also the non-glycoside fragment was involved in the binding. Computational approaches were then used to get a 3D perspective of the complex (Figure 3.33).

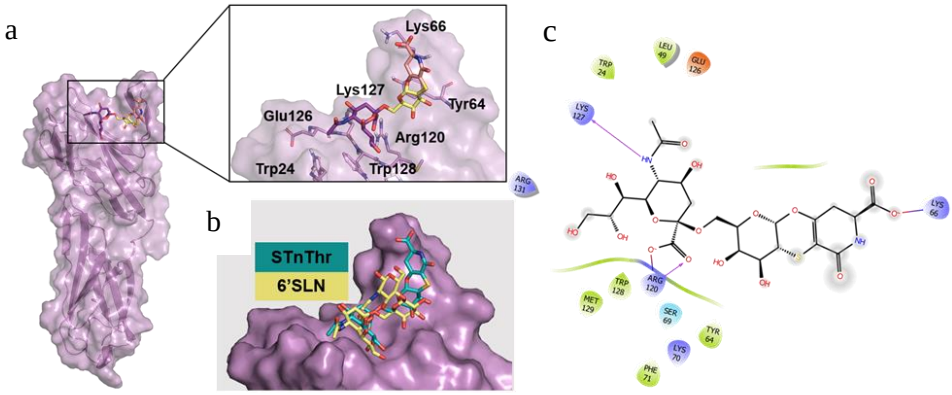
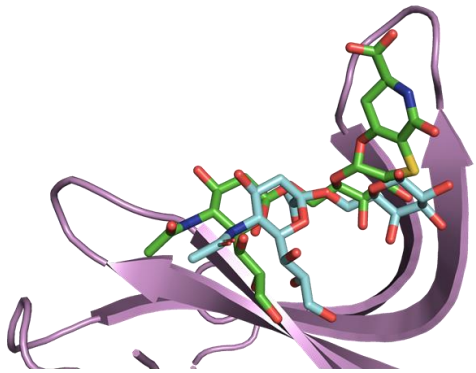


Figure 3.33 Interaction between h-CD22 and STnThra ligand a) Three-dimensional model derived by STD, tr-NOESY and MD for the STnThra ligand bound form to h-CD22 crystal structure. The representative frame from the most populated MD cluster was considered to depict the complex. **b)** Superimposition of the 6'SLN bound model previously obtained and the STnThra bound model **c)** Two-dimensional plots representing the interactions between the STnThra and the binding site residues of h-CD22. The representative frame of the most populated MD cluster was considered to depict the complex.

First, docking calculation showed that the cluster possessing the higher number of poses contained the models that were comparable to the h-CD22 canonical binding motif (Figure 3.34).



Docking pose	Cluster rank	n. of cluster conformations	Est. Free energy of binding (Kcal/mol)	Est. Inhibition constant	RMSD from ref.
STNThra	1	85	-5.0	75.54	1.371

Figure 3.34 Docking of StnThra/h-CD22. Docking of STnThra (green) in the crystal structure of h-CD22(violet) in complex with 6'siayllactose (cyan, PDB ID: 5VKM). The relative docking parameters are also reported.

Therefore, a pose from this cluster was subjected to MD simulation to investigate the complex dynamic behavior and overall stability. Consequently, the RMSD analysis of the representative models obtained from MD trajectory clustering showed that the ligand remained stable in the binding pocket of the receptor for the entire simulation time (Figure 3.35).

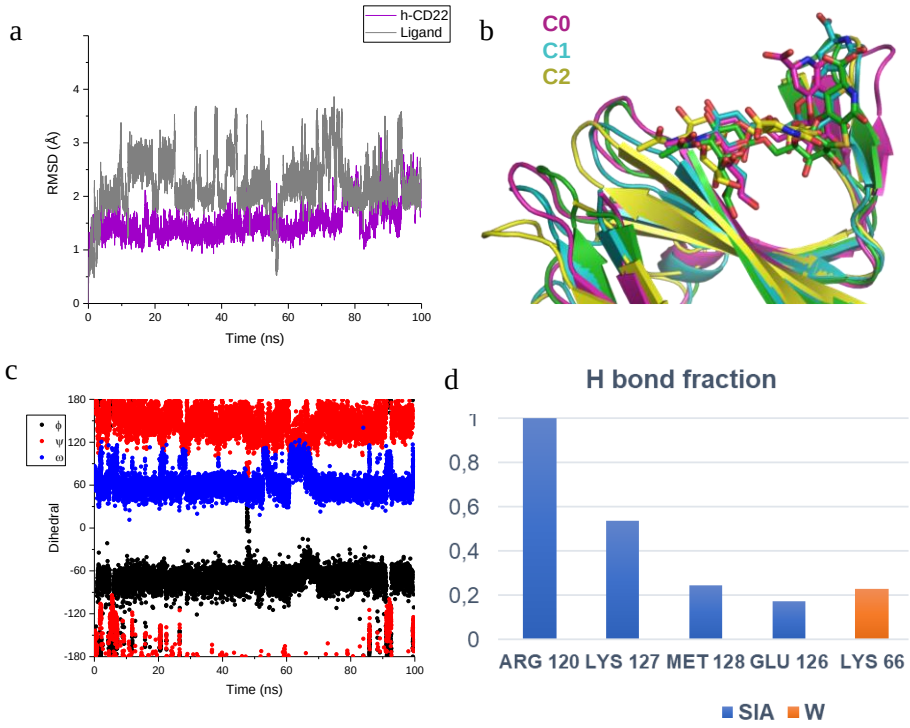


Figure 3.35 MD simulation analysis of STnThra/h-CD22 complex. a) Protein and ligand RMSD. The ligand RMSD was calculated in reference to the protein. **b)** Superimposition of the three most populated clusters from the MD simulation. Kmean algorithm was considered for clustering **c)** Neu5Ac-Gal dihedral angles of STnThra fluctuation along the MD simulation**. d)** Protein/ligand H bonds fraction.

From the ligand/ receptor interaction analysis, it was noted that the sialic acid moiety interacted with Arg120, to form the carboxylate bridge, essential determinant for Siglecs recognition. The lateral glycerol chain protons were involved with stacking with Trp128 residue, as well as hydrogen bonds with Arg131, additionally, the N-acetyl group formed hydrogen bond with Lys127 residue. The Gal pyranose ring stacked with the Tyr64 residue, thus explaining the significant STD effects. Notably, the non-glycosidic moiety was in close proximity with the CC' loop of CD22 and engaged a strong hydrogen bond between its carboxylate group and the Lys66 residue of the loop.

3.5 Discussion

Here, new structural insights into the molecular basis of N-glycans recognition by human CD22 have been provided by a combined use of NMR, binding assays, docking and molecular modelling.

In detail, the recognition and binding processes between CD22 and natural and synthetic ligands have been dissected by profiling the ligands' epitopes in their bound conformation and by model them into the CD22 binding pocket.

The STD NMR investigation defined the binding epitope, namely the ligands' region recognized and interacting with CD22. Then, the bound conformation of sialoglycans was defined using NOE based data in combination with a computational approach including MM and MD simulations; last step, docking and molecular modelling studies, together with CORCEMA-ST protocol, were used to achieve key information into the fine structural characteristics of the complexes. In detail, our three-dimensional structure of the CD22-sialoglycan complex confirmed the architecture of the binding site essentially composed by hydrophobic (Phe21, Trp24, Trp128 and Met129), polar (Tyr64), basic (Arg120, Lys127) and

acid (Glu126) residues. According to the deduced model, the *N*-acetamide moiety of sialic acid, giving the strongest STD effect, was in close contact with Trp24, Trp128 and Glu126 side chains and established a stable hydrogen bond with the carbonyl group of Lys127. The sialic acid glycerol side chain was also involved in the binding, by close CH-π interaction with Trp128 which in fact explains why modifications of either the acetamido- or the 8- 9- hydroxy groups of Sia's influence the binding affinities. Aromatic extension at C9, as example, results in a strong reduction on IC50 values for these sialosides.[205]

Moreover, in accordance with crystallography data, we demonstrated that the CD22 V-set domain established key interactions not only with the terminal sialic acid but also with the adjacent galactose moiety that appears to play a pivotal role in the interaction process; no contacts were indeed observed with all the others sugar residues.

In addition, our results elucidated the conformational behavior of biantennary sialylated N-glycans when bound to the h-CD22, suggesting that they preferentially adopt an extended conformation and that the interacting surface of both antennae was totally comparable. Noteworthy, it has been shown that the orientation of the two antennae of the glycan chain in the extended conformational state allowed a stable simultaneous cross binding of two Siglec molecules, favoring the formation of CD22 homo-oligomers on the B-cell surface. Likewise, changes in the glycan shape, for example as consequence of the closure or the opening of the α-1-6 antenna, modify the covered area provided by the glycan on the protein surface might contribute to tune the recognition and interaction events between sialyloglycans and CD22, governing *cis* or *trans* interactions.

Next, by applying a similar approach, the effect of sialic acid glycolylation on the binding with CD22, was studied by comparing the behavior of different ligands in complex with murine and human CD22, thus improving the knowledge of

the structural basis of the recognition of sialylated *N*-glycans from by CD22 receptor.

The STD NMR analysis revealed a comparable binding epitope of both murine and human CD22 towards glycolylated glycans, involving an almost invariant set of residues within CD22 binding pocket. In all resulting complexes a bent umbrella-like conformation of the ligand was adopted, as supported by the results achieved from the NMR data and MD simulations. Moreover, it was noted the possibility of m-CD22 receptor to form additional interactions with Neu5Gc ligand, mainly involving the hydroxymethyl group of *N*-glycolyl moiety; this was further supported by a comparison of the molecular surfaces of h-CD22 and m-CD22 where deputed to interact with *N*-acetyl/*N*-glycolyl Sia chains; Indeed, it was revealed that the substitution of Asp25 in place of Lys23 in the binding subsite of m-CD22 influences the possibility to establish hydrogen bonds with the hydroxyl group of the glycolyl moiety. Overall, our studies indicate that, despite the different nature of Sia residue, the recognition region of h-CD22 is almost invariant comparing Neu5Ac and Neu5Gc containing glycans. These results agree with the similar affinity of h-CD22 toward Neu5Gc/Neu5Ac structures, as recently reported by Varki et al.[225]

Many efforts have been made to developing selective, potent CD22 glycan ligands, mostly through modifications at the Sia's C4, C5 and C9 positions, leading to the development of high affinity ligand with therapeutic applications. With the aim to improve CD22 binding by multivalent ligands, the interaction of the structurally constrained STnThra mimetic was investigated in the interplay with h-CD22. Indeed, such glycomimetic has a carboxylate group at the aglycon moiety that may allow for the easy conjugation to bioavailable nanomaterials such as nanoparticles, dendrimers and liposomes.[226]As result, it was demonstrated that the compound was effectively recognized by the target receptor. Molecular

modelling showed that CD22 bound the ligand, without alteration of the interaction pattern, demonstrating that the rigid aglycone did not affect the presentation of the Sia-Gal epitope to the enzyme. This finding suggests that this analogue may be a promising candidate for the glycoimmunotherapy applications targeting h-CD22. Indeed, the obtained results suggest that these structurally constrained sialic acid analogs have the potential to improve the recognition by the h-CD22 with respect to the natural ligands, thus a class of related compounds may be designed to find high affinity and selective CD22 ligands, by combining virtual screening with synthetic and *in vitro* binding assays to compare the IC_{50} of hit compounds, thus leading to *in vivo* testing of the best leads. Furthermore, the functionalization and characterization of suitable nanomaterials coated with STnThra ligands may lead to interesting therapeutic applications, for example to efficiently probe the cell surfaces.[229]

Globally, these findings provided new hints for the design and synthesis of high-affinity ligands of potential therapeutic relevance, with pharmacophores that are preorganized in their bioactive conformation, leading to reduced entropy costs upon binding. For instance, these information could be used to develop novel compounds effective against B-cell derived tumors, acting to prevent propagation of tumor cells and their development at secondary sites or to improve the currently employed cancer combination therapies, that may greatly benefit to those patients who do not answer to such regimes.

Chapter 4- Characterization of Siglec-10/ Sialoglycans interactions by means of biophysical, computational and spectroscopic techniques

4.1 Introduction

Siglec-10, human homologue of Siglec-G, is an inhibitory receptor,[227] expressed on DCs, macrophages and B cells.[228] It is encoded, together with other Siglec proteins, on chromosome 7 and is characterized by five extracellular domains and three tyrosine-based motifs in its cytoplasmic tail and belongs to CD33-related Siglecs sub-category.[229]

Siglec-10 is known to regulate the tolerance of B1 sub-class of B cells, dampening autoimmune responses upon recognition of α(2,6) or α(2,3) linked sialoglycans. Even if Siglec-10 is mainly expressed on B cells, it also can be found on dendritic and myeloid cells and on subsets of human leukocytes, including macrophages and neutrophils.[230,231,232]

Given the wide expression pattern with respect to other Siglecs, Siglec-10/G perform various roles within the immune system. First, it is responsible for BCR signalling regulation, thus modulating both expansion of B1 cells and tolerance of B cells towards self-antigens. Indeed, as CD22, Siglec-10/G act as a co-receptor of BCR, dampening the calcium signal on B1 cells and consequently impedes the activity of the transcription factor NFATc1 and NFκB through the enrollment of ITIM-binding protein SHP-1 (or Grb2).[233]Differently from CD22, it is believed that Siglec-10 undergoes to the direct recruitment to Sias on BCR after antigen stimulation. [234] (Figure 4.1)

a

b

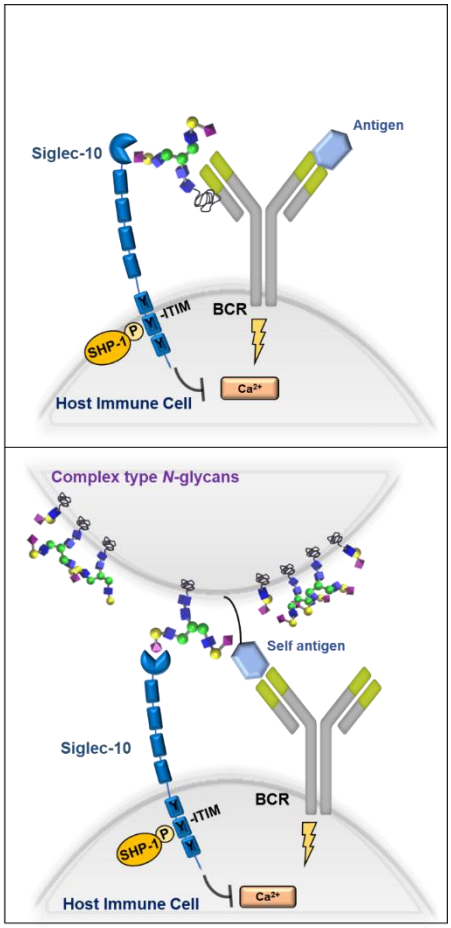


Figure 4.1 Immune response mechanisms mediated by Siglec-10 on B1 cells. a) Regulation of BCR activation by Siglec-10 binding to *cis* ligands on BCR, leading to downstream immune inhibitory signalling. The *cis* interactions regulates B-cells proliferation **b)** Self-antigens are expressed on cells with abundance of Siglec-10 ligands and possibly engage both BCR and Siglec-10/G, and inhibit BCR signalling. The *trans* sialic acid-Siglec interaction maintains tolerance to self antigens.

Second, it was shown that Siglec-10/G on DCs selectively modulates innate immune response to DAMPs but not PAMPs, thus suggesting an exclusive role in mediating sensing of infection vs. tissue injuries by the immune system.[218]

The discrimination between DAMPs and PAMPs in dendritic cells by Siglec-10/G occurs through recognition of CD24, a glycosylphosphatidylinositol-anchored membrane receptor that is heavily sialylated.[229] The interaction between CD24-associated DAMPs and Siglec-10 dampens the activation of TLR4 by NF-kB signal pathway, leading to a repression of inflammatory response (Figure 4.2a).[235] Intriguingly, the interplay between Siglec-10 and CD24 has been associated with reduced infiltration and weakened NK cells function in patients affected by hepatocellular carcinoma.[236] Noteworthy, it has been also observed at fetal-maternal interface during the first three months of pregnancy, suggesting a possible role of Siglec-10 in establishing the immune tolerance in the placenta between the fetus and the mother.[237]

On the other hand, it has been demonstrated that[238] Siglec-10 can be exploited by the Gram-negative bacterium *C. jejuni* to evade host immune responses, since the pseudaminic acid residues (Pse, a sugar highly related to Sia) decorating the bacterial flagellum are recognized by Siglec-10, this leading to the production of anti-inflammatory cytokines as interleukine-10 (IL-10).[239] Correlated to this promotion of anti-inflammatory axis on DCs Siglec-10 by *C. jejuni*[118]flagella is the Guillain-Barré syndrome (Figure 4.2b).

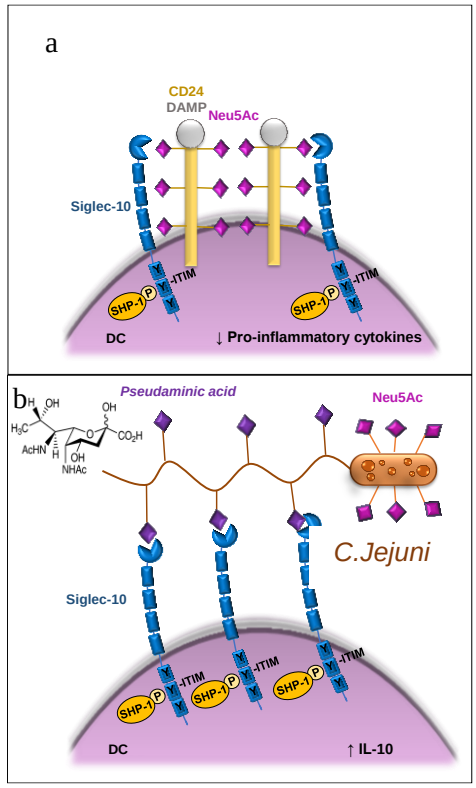


Figure 4.2 Siglec-10 immune regulation through sepsis. a) Siglec-10 on DCs interacts with CD24 to impede DAMP-mediated TLR signalling **b**) Siglec-10 binding of pseudaminic acid on the flagella of *C. jejuni* promotes IL-10 production by DCs.

To date, there are no structural information on Siglec-10, either via X-ray and NMR, and therefore the molecular basis of Siglec-10 recognition by endogenous ligands still needs to be clarified. Analyzing the molecular events that regulate host immune suppression upon sialylated *N*-glycans recognition by Siglec-10 represents a fundamental prerequisite for deciphering its mechanism of action and may promote the rational design and synthesis of highly specific ligands to exploit its anti-inflammatory pathway for the modulation of immune cellular responses. In this context, the important question of how Siglec-10 recognizes and binds

different sialylated glycans has been here faced. By coupling NMR approaches to biochemical and computational methodologies a 3D model of Siglec-10 interaction with the cognate ligands was achieved and validated, providing a dynamic characterization of the molecular interactions in solution, describing the binding epitopes of the ligands in their bioactive conformation.

As part of the PhD project, the computational studies have been performed at CSIC in Madrid under the supervision of Prof. Martín Santamaría. Furthermore, with the aim to produce the labeled Siglec-10 CRD domain for NMR investigation, efforts have been made to carry out the recombinant expression of Siglec-10 in *E. coli* during the period spent at Giotto Biotech under the supervision of Prof. Marco Fragai (University of Florence).

The overall achieved results have provided the first molecular clues into the mechanism of sialoglycans recognition by Siglec-10, demonstrating the critical role played by the glycan conformation in the protein – ligand interaction

4.2 Unveiling molecular recognition of sialoglycans by Siglec-10

To address the important matter of how Siglec-10 detects sialylated glycans, it has been explored the interaction of Siglec-10 and different kind of sialylated ligands, containing Siaα(2,3)Gal and Siaα(2,6)Gal glycosidic linkages,[240]displayed in Figure 4.3.

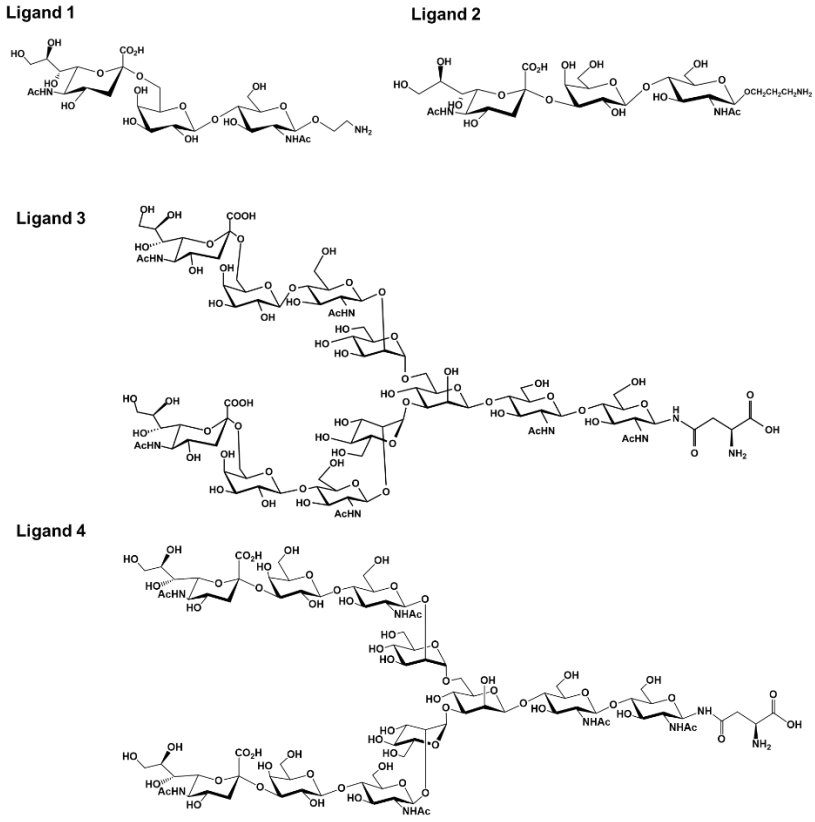


Figure 4.3 Structures of sialylated ligands employed in this work. The two trisaccharides, Sia-α-(2→3)-Gal-β-(1→4)-GlcNAc, ligand **1**, and Sia-α–(2,6)-Gal-β-(1,4)-GlcNAc, ligand **2** differ in the type of glycosidic linkage at the non-reducing end; the longer glycans, ligands **3** and **4**, are commonly exposed on mammalian cells.

The interaction between different sialylated conjugates and Siglec-10 was primarily evaluated by ELISA in collaboration with Dr. Fabrizio Chiodo (CNR Pozzuoli, Napoli), see section 9 for further details. The results, depicted in Figure 4.4a, showed that Siglec-10 recognized all the sialylated conjugates tested. Similarly, intrinsic fluorescence experiments showed the decrease of Siglec-10 tryptophan fluorescence intensity when different amounts of ligands were titrated into a fixed concentration of the protein. This indicated that, independently from the nature of the Sia linkage to the Gal residue, both sialoglycans could interact with Siglec-10 (Figure 4.4b).

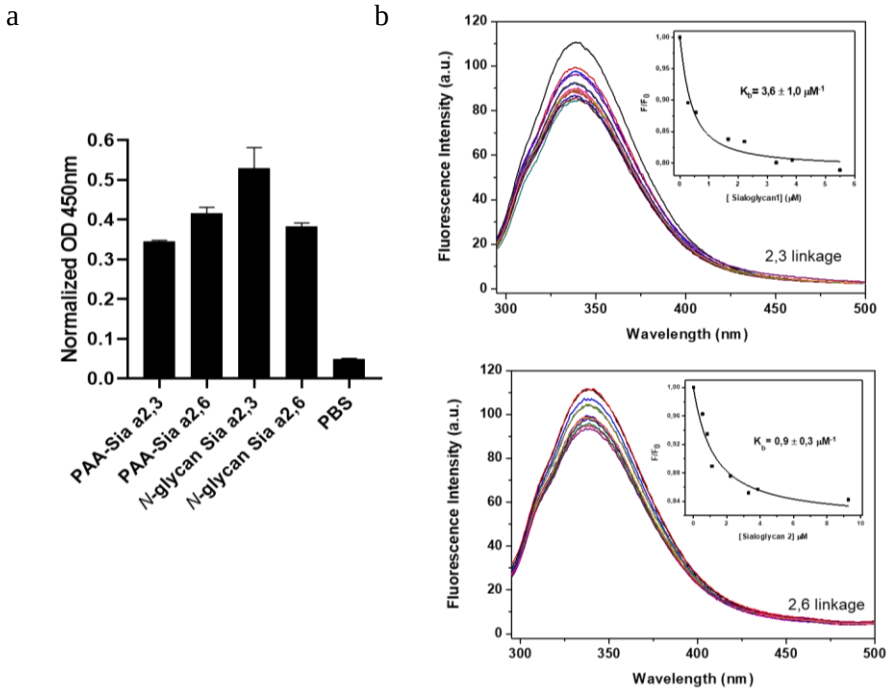


Figure 4.4 a) ELISA tests. Siglec-10 binding to sialylated conjugates detected by ELISA. ELISA plates were coated with synthetic conjugates carrying terminal sialic acids with α23 or α26 linkages, followed by the addition of a human Fc chimera of Siglec-10. The detection of the interaction was evaluated with an anti-Human IgG-HRP (horseradish

peroxidase) conjugate followed by substrate addition. Siglec-10 shows binding to all the sialylated conjugates tested. Error bars indicate standard deviations of the experiment performed in duplicate. **b)** Fluorescence analysis. Fluorescence spectra of Siglec-10 in the presence of increasing concentrations of sialoglycan **1** (top panel)- and **2** (bottom panel) respectively. Insets displaying the binding isotherm and the obtained binding constant are showed.

4.2.1 Molecular interaction between Siglec-10 and trisaccharide 1

The molecular recognition of trisaccharide **1**, Sia-α-(2,3)-Gal-β-(1,4)-GlcNAc-OR, by Siglec-10 was then analyzed through a combination of ligand-based NMR techniques and computational studies.

The STD NMR analysis (Figure 4.5) further confirmed the binding and allowed to provide a detailed description of the structural elements of **1** when accommodated into the Siglec-10 binding pocket.

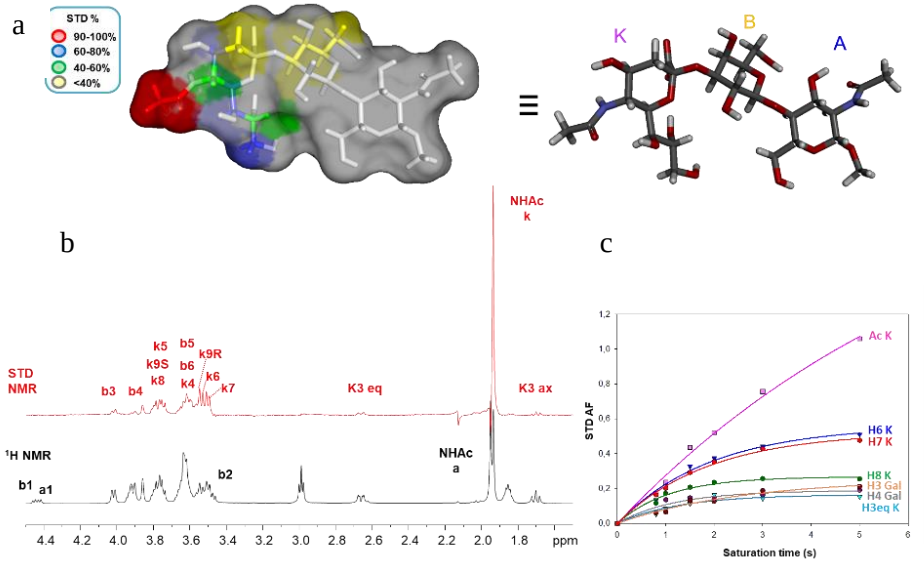


Figure 4.5 STD NMR analysis of the interaction between Siglec-10 and trisaccharide 1. **a)** STD-derived epitope mapping on the molecular envelope of ligand **1** in its bioactive minus gauche conformation. STD effects were calculated as difference between the intensity of the relative signal in the STD-NMR spectrum, (I_0–I_{sat}), and the off-resonance

(reference spectrum), I_0. Siglec-10 gives the maximum saturation transfer to the acetyl group protons belonging to K residue of trisaccharide **1**, so it was set at 100% and the other ligand **1** protons percentages were derived accordingly. **b)** Reference ^{1}H NMR spectrum (black) and STD 1D NMR spectrum (red) of the Siglec-10/ligand 1 mixture (molecular ratio 1:100, temperature of 298K, magnetic field of 600MHz). **c)** STD build-up curves were derived by the monoexponential equation $STD = STD_{max}[1 - \exp(-k_{sat}t)]$, constructed by fitting the STD AF at different saturation times, from 1s to 5s.

The comparison of the STD NMR spectrum with the corresponding reference (off-resonance) showed that only certain protons of ligand **1** displayed STD enhancements (Figure 4.5b). In detail, Sia (**K**) was essential for the interaction with Siglec-10, giving the most significant contribution to the molecular recognition process. Indeed, the *N*-acetyl group of **K** exhibited the highest STD signal, the protons at position 6 and 7 affected as well by a moderately high saturation transfer, close to 80%. Significant STD signals, around 60%, were also observed for protons H5, H8 and H9, indicating the contribution in the interaction, whereas weaker STD increases were observed for the diastereotopic H3 protons of **K**, and for few protons of the Gal unit **B**. Conversely, the GlcNAc (**A**) almost did not contribute to the binding process, as confirmed by the absence of any corresponding STD signal. Additionally, STD build-up curves were constructed (Figure 4.5c, Table 4.1) to precisely map the interacting epitope of ligand **1**. The strong involvement in the binding of **K** was confirmed, as well as the involvement of **B** unit, suggesting the proximity to the Siglec-10 binding site.

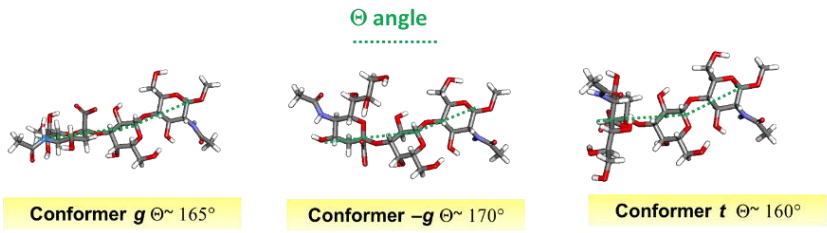
Table 4.1 Experimental STD epitopes of trisaccharide 1 bound to Siglec-10.
STD intensities of ligand **1** bound to Siglec-10 experimentally calculated at different saturation times, using the monoexponential equation $STD(t_{sat}) = STD_{max} \times [1-exp(-k_{sat} \times t_{sat})]$.

^{1}H	STD_{max}	K_{sat}	STD (fit)	STD epitopes (fit)
Ac Sia	1.8331	0.1871	0.342973	**100%**
H6 Sia	0.5615	0.5144	0.288836	**84.2%**
H7 Sia	0.5251	0.5249	0.275624	**80.4%**
H8 Sia	0.263	0.7105	0.186862	**54.4%**
H9S Sia	0.3128	0.5849	0.182957	**53.3%**
H4 Gal	0.2017	0.5974	0.120496	**35.1%**
H3 Gal	0.2919	0.3848	0.112323	**32.7%**
H5 Sia	0.2243	0.6552	0.146961	**42.8%**
$H3_{eq}$ Sia	0.1651	0.7792	0.128646	**37.5%**

The determination of the epitope map of trisaccharide **1** was complemented with the study of the conformational properties in the free state and upon binding with Siglec-10. The accessible conformational space of **1** was sampled by MM and MD simulations and NOE-derived experimental data. This permitted to define the preferred torsional angles adopted in the free and bound states by ligand **1**.

First, from the MD simulation analysis, it was observed that the ligand **1** mainly explored the *–g* (Φ=-60º) and *t* (Φ=180º) conformations (Figure 4.6). Conversely, the Ψ angle remained stable along the dynamics around an average value of -11º. Thus, a conformational equilibrium was observed in the free state between three main conformers namely *–g*/ *g*/ *t* differing by the value of the Φ torsion angle (H1-C1-O-CX') between Sia and Gal residues (-60°/ 60°/ 180°, Figures 4.6).[211]

a



b

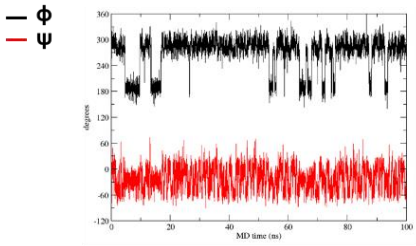


Figure 4.6 Conformational analysis of ligand 1 by MD a) The three major conformations in solution of the trisaccharide **1**. **b)** Dihedrals ϕ, ψ of Sia-α-(2,3)-Gal linkage fluctuation in MD simulation of the free state.

It is worth to remark that all the conformers defined around the Sia-α-(2,3)-Gal linkage adopt a rather extended conformation in which the sugar units occupy a cone-like region of space.[212]This topology is defined by a value of θ higher than 110°, where θ represents the angle between the carbon C2 of Sia residue and C1 atoms of the adjoined Gal and GlcNAc units.[209]MD results were supported by NOESY NMR analysis, that confirmed the existence of an equilibrium between different conformations of **1** in its free state (Table 4.2 and Figure 4.7a).

Table 4.2 Intra-molecular ^{1}HSia $-^{1}$HGal distances for the investigation of conformational behaviour of ligand 1. The distances were obtained by the integration of signals in selective (sel-NOE) experiments at different mixing times irradiating the B3 proton, using B1-B3 as reference. In the free state, the conformation of ligand **1** was defined as an equilibrium of three geometries, changing in Φ (C_1-C_2-O-C_3') torsion angle that could assume -60°/60°/180° values, corresponding to –g/g/t populations respectively. Passing to the bound state, a preference for –g conformation was investigated.

Distance ^{1}HSia – ^{1}H Gal	**Exp. Free state**	**Exp. Bound state**	**Conformer –g Φ = -60° Ψ = -11°**	**Conformer +g Φ = 60° Ψ = -11°**	**Conformer t Φ = 180° Ψ = -11°**
$H3_{ax}$ Sia-H3 Gal	2.86	3.9	**4.1**	3.11	2.18
$H3_{eq}$ Sia-H3 Gal	3.65	----	**4.4**	2.03	3.40
H8 Sia-H3 Gal	3.41	3.19	**3.36**	6.17	4.31

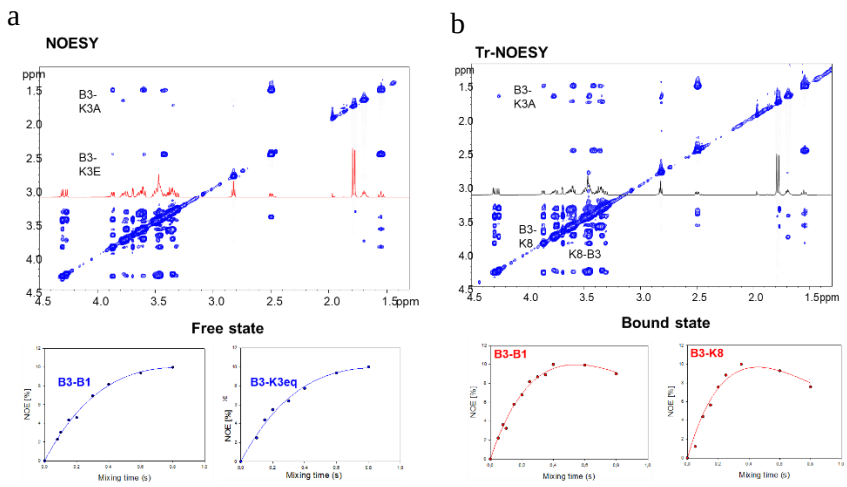


Figure 4.7 NOESY and tr-NOESY analysis of ligand 1. NOESY spectrum **a)** of ligand 1 in its free state (mixing time of 600ms) and tr-NOESY spectrum **b)** of ligand 1 bound to Siglec-10 (mixing time of 400ms) at 283K. The protein-ligand molar ration was set at 1:10. Some differences in terms of signal intensities were observed comparing the free

and bound state. At the bottom panel the selected NOE build-up curves of ligand **1** in the free and bound states by selective excitation of H3 of Gal residue

As for the bound state, tr-NOESY NMR experiments were acquired to identify possible changes of the ligand **1** conformation when interacting with Siglec-10. Remarkably, the analysis of NOE cross-peaks demonstrated a selection of the *–g* conformer (Φ/Ψ torsion angles of -60°/-11°) of **1** upon binding (Figures 4.6 and 4.7b, Table 4.2); in particular, a decrease of B3-$K3_{ax}$ cross-peak intensity and the absence of B3-$K3_{eq}$ signal was observed in the tr-NOESY spectrum (Figure 4.7b). On the contrary, an increase of the B3-K8 crosspeak intensity was revealed. A precise integration of key signals obtained from selective-NOE experiments, allowed to extract crucial *inter*-residual ^{1}H-^{1}H distances (Table 4.2) by the construction of NOE build-up curves further confirmed a preference for the *–g* conformation of **1** when bound to Siglec-10 (Figure 4.5 and Figure 4.7).

4.2.2 3D-View of Siglec10 in interaction with trisaccharide 1

A 3D view of the interaction between the Siglec-10 and trisaccharide **1** was achieved by computational studies. Since the crystal structure of Siglec-10 has not been solved yet, the homology model of Siglec-10 CRD several was performed using several Siglecs structural templates. The models were computed with different homology modelling servers and then refined by MD simulations as implemented within the AMBER 14 package.

In detail, the Siglec-10 CRD (aa 17-144) was considered for computational 3D structures calculation by homology modelling. Human Siglec-10 encoding sequence was extracted from NCBI. Siglec-8 (PDB ID: 2N7A), Siglec-7 (PDB ID: 1NKO), Siglec-5 (PDB ID: 2ZG2) and Siglec-3 (PDB ID: 5IHB) structures were used as templates. For each template, four homology models were generated using four different servers: SWISS-MODEL, I-TASSER, PHYRE2 and RAPTOR-

X.[241] To select the best models, detailed analysis was performed, focusing on the orientation of the binding site residues and the flaking loops. The models derived from I-TASSER, and PHYRE-2 were discarded because the structures did not have coherence with the experimental 3D structures of the homologous proteins. Thus, computational studies were performed using the 3D structures obtained from SWISS-MODEL and RAPTOR-X, which were submitted to molecular dynamics simulations of 100 ns to optimize the geometry.

To assess the quality of the homology models, the results were compared to the MD simulation carried out using the crystal structure of Siglec-3/CD33, which was performed with the same simulation parameters (data not shown). The Siglec-3 crystal complex was chosen as reference because of the similarity to Siglec-10, in terms of binding specificity and affinity toward sialylated glycans.[89]
The structural models of Siglec-10 displayed high similarity in terms of 2D and 3D structure among them and also in comparison to the corresponding templates, showing a nearly invariant canonical V-set Ig-domain fold. As can be expected by the sequence alignment (Figure 4.8a), the superimposition of the Siglec-10 models to the crystal structures of other Siglecs revealed some diversity concerning the amino acids constituting the binding site (Figure 4.8b), suggesting a putative role in the specificity of recognition of precise sialylated epitopes.

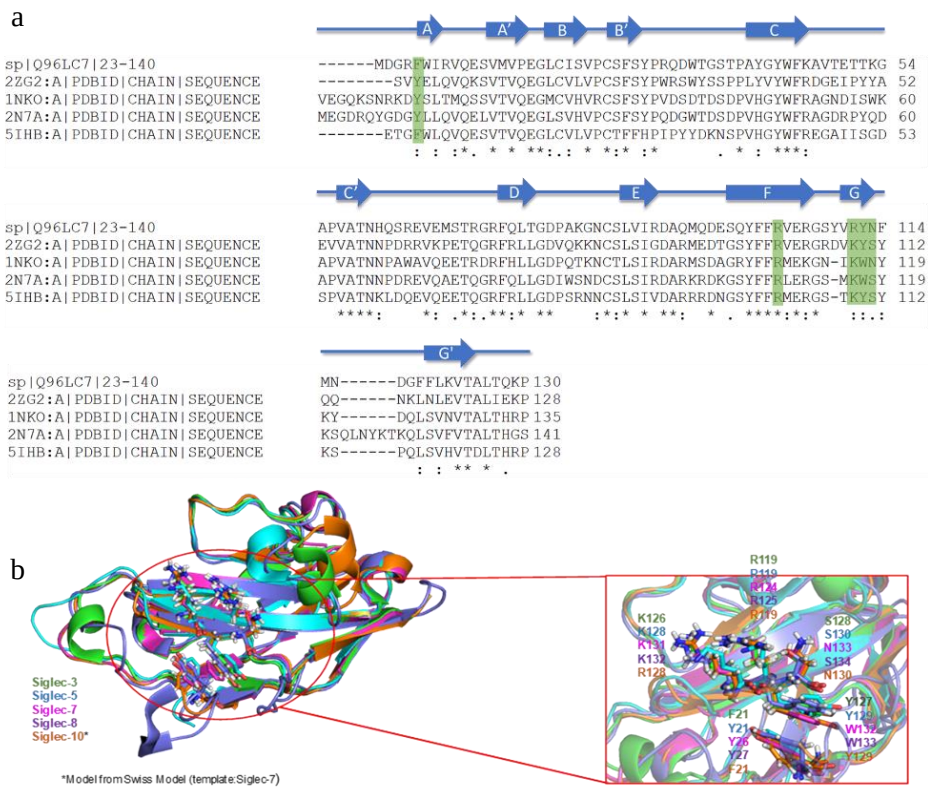


Figure 4.8 Features of the Siglec-10 CRD a) Sequence alignment of Siglec-10 CRD respect to the CD33-like Siglecs templates, namely Siglec-3 (PDB ID: 5IHB), Siglec-5 (PDB ID: 2ZG2), Siglec-7(PDB ID: 2HRL), Siglec-8 (PDB ID: 2N7A). The residues responsible for the binding of sialylated glycans are highlighted in green. **b)** Superimposition of the best Siglec-10 homology-based model (orange) and the X-ray crystal structures from other Siglec members. The residues of Siglec-10 that interacted with the sialic acid moiety which are reported to be crucial for the interaction with sialylated glycans are shown as sticks. The Arg in position 119 in Siglec-10 was indeed conserved in all the Siglecs, along with the aromatic residues in position 21 and 129 (Phe and Tyr in Siglec-10).

Nevertheless, from the superimposition of the selected 3D structure, it was noted a different shape of the variable region of Siglecs, namely the *inter* -strand CC' loop,[88] probably due to the lower sequence similarity with respect to

the structural templates (Figure 4.9a), that led to difficulties in the loop conformation prediction. For this reason, the mobility of the CC' and GG' loop of the selected models has been investigated during the dynamic simulation (Figure 4.9 e-f).

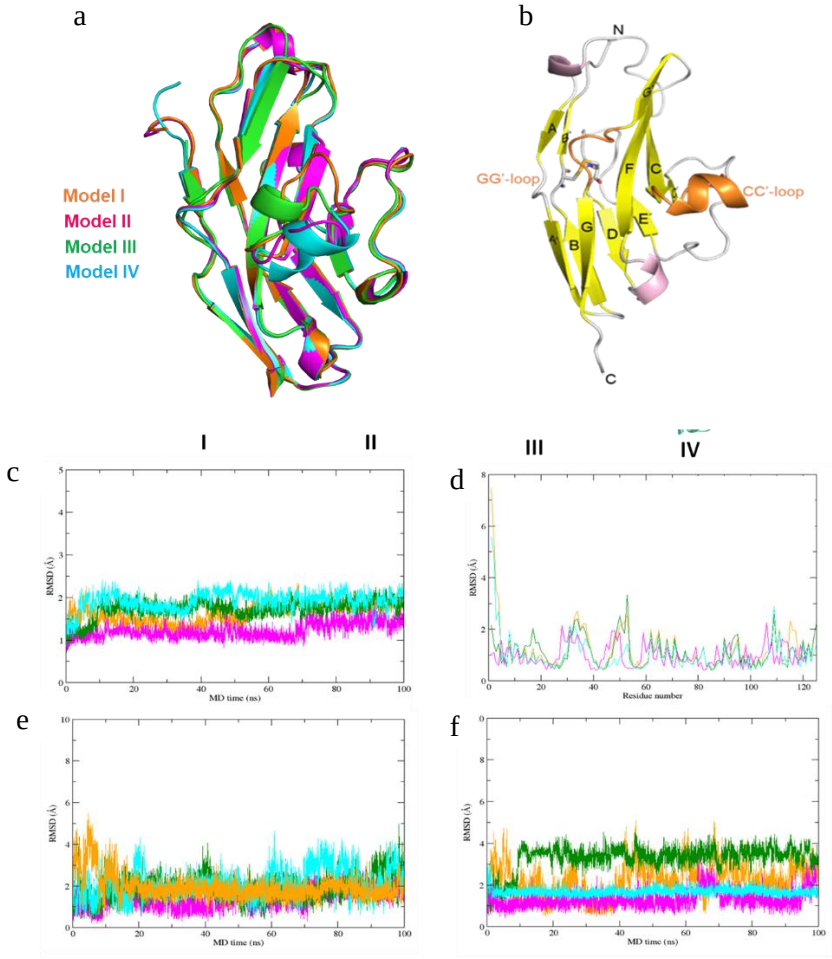


Figure 4.9 Comparison of Siglec-10 homology models upon MD a) Superimposition of the selected homology models. Relevant differences can be observed only in one of the variable Siglec region that is the C-C' loop. **b)** Structure of the best Siglec-10 model, obtained by means of Swiss Model by using Siglec-7 as template. The secondary structure

elements are indicated according to Siglecs V-set domain nomenclature. **c)** Analysis from the MD simulations of the best homology models. Superimposition of the average structures of each model at different time of the simulation. **d)** RMSD of the backbone of Model I (orange), Model II (purple), Model III (green), Model IV (cyan) **c)** RMSF of the backbone of the different models of Siglec-10 **e-f)** RMSD of the backbone of the important loop regions (CC'-loop) on the left and (GG' -loop) on the right.

To select the most accurate model, a detailed analysis was performed, focusing the attention on the orientation of the binding site residues and the flaking loops (Figure 4.9). As result, it was observed that the models obtained using Siglec-3 and Siglec-7 as templates (both the Swiss-Model and the Raptor-X models) displayed higher stability along the simulations and therefore they were used for the subsequent docking calculations using ligand **1**.

Thus, four promising models (Figure 4.9a) were subsequently used for docking calculations by means of AutoDock 4.2 with trisaccharide **1**. The ligand protein complexes selected from the analysis of the binding poses were then submitted to MD simulations to monitor the stability in solution. The most consistent protein ligand complexes obtained from docking and MD were subjected to CORCEMA-ST calculations, with the aim to predict the STD-NMR intensities for a proposed molecular model of a ligand-receptor complex; and measure the STD values can be then compared and used to validate a given complex.
The combined analysis permitted the selection of MODEL IV as the optimal/best Siglec-10 model, predicted by using the template of Siglec-7 (Figure 4.10).

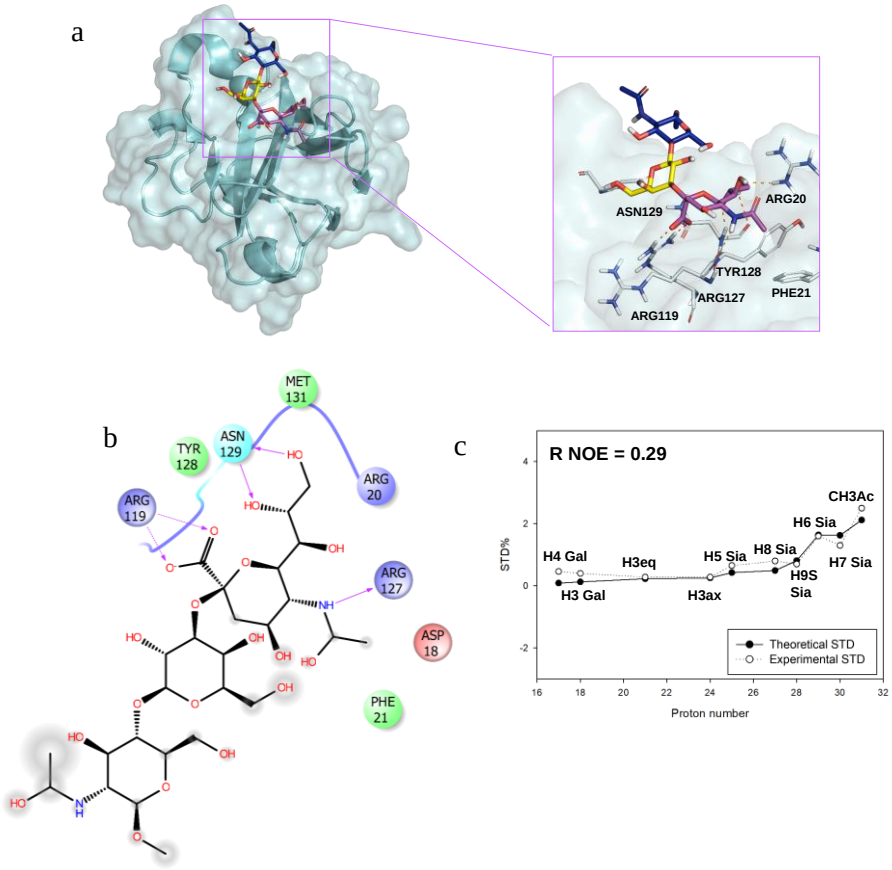


Figure 4.10 Docking/MD and CORCEMA-ST analysis of trisaccharide 1. a) Best binding pose (STD, tr-NOESY and MD based) for the docking of the trisaccharide **1** bound form into binding site of Siglec-10 model derived by using Siglec-7 as structural template. The carbohydrate and interacting amino acids are shown as sticks. Dashed black lines represents hydrogen bonds. **b)** Two-dimensional plots representing the interactions between the sialylated trisaccharide and the residues of the binding site of Siglec-10 model. Dotted arrows indicate hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone. The residues shown are close to the ligand and participate in hydrophobic and polar interactions **c)** Comparison between experimental (dashed line) and theoretical (solid line) STD data for the best model of trisaccharide **1** bound to the structural model of Siglec-10 CRD derived by CORCEMA-ST analysis.

In detail, docking calculations were performed using trisaccharide **1** in each of its three possible conformations (*-g*, *+g*, *-t*) and the best models according to AutoDock scoring were subjected to CORCEMA-ST calculations (Figure 4.10). The best fit between predicted and experimental data, corresponding to the lowest R factor value, was given by the MODEL IV/**1** complex reported in Figure 4.10, carrying the trisaccharide **1** in the extended *-g* conformation. Of note, the profiles of the theoretical STD data achieved for the other two models, bearing the trisaccharide **1** in *g* and *t* conformation respectively, were far-off from the experimental STD effects, ensuing higher R factor values and therefore that these models were excluded (data not shown). The best Siglec-10/ligand **1** mode (Figure 4.10) was then subjected to MD simulation, further validating the preference of the ligand for the *-g* conformation in the bound state, as suggested by the stability of the Φ glycosidic torsion angle between the sialic acid and the galactose moiety during the simulation time (Figure 4.11). During the 100 ns MD simulation of the model IV/**1** complex, it was possible to observe high stability of the complex until 50 ns, without applying any restriction (Figure 4.11).

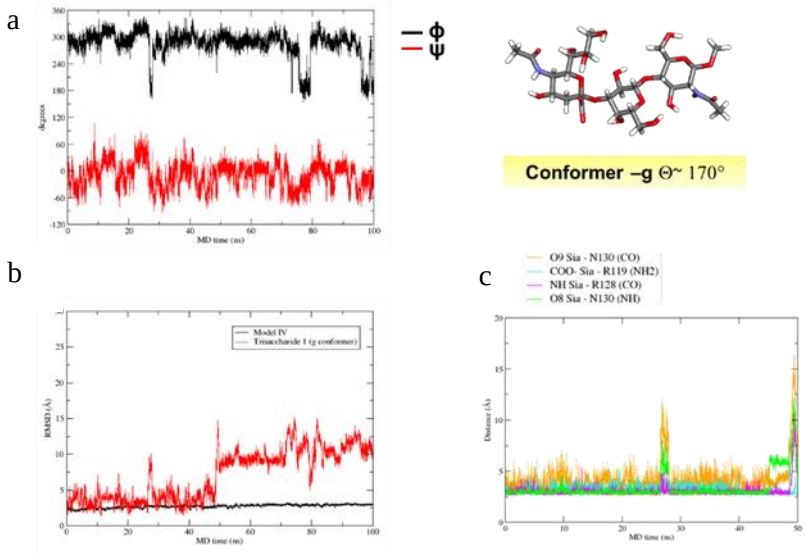


Figure 4.11 MD simulation trajectory analysis of trisaccharide 1/model IV complex. a) Dihedrals ϕ, ψ of Sia-α-(2,3)-Gal linkage fluctuation in MD simulation of the bound state (model IV) **b)** RMSD plot of trisaccharide **1** (-*g* conformer) in complex with Siglec/10 CRD (model IV) **c)** Representative intermolecular distances between Sia unit and the Siglec 10 residues calculated along the simulation.

Examination of the binding mode of trisaccharide **1** by Siglec-10 (model IV) revealed the typical architecture of Siglecs binding site, characterized by a solvent-exposed, highly positively charged cavity constituted by the C, F and G β-strands and surrounded by the CC′- and GG′-loops (Figures 4.8 and 4.10). The Sia residue was placed at the top side of the G-strand, making several polar interactions with the receptor amino acids. Such interactions established by the Sia moiety were previously reported to be crucial for recognition of sialylated glycans by other CD33-related Siglecs.[242,243,244]Along the simulation, ligand-protein interaction analysis identified relevant polar contacts between the trisaccharide **1** Sia unit and the residues located into the binding site of Siglec-10, namely Arg119, Arg127, Asn129, along with hydrophobic interactions involving Phe21 and

Tyr128. In detail, the Sia carboxylate established a stable ionic interaction with the ammonium group from the highly conserved Arg119. Furthermore, the hydroxyl groups of the glycerol chain of Sia made hydrogen bonds with Asn129 and Arg38 residues of Siglec-10 V-set domain. Also, a polar contact between the *N*-acetamide moiety of Sia and Arg127 residue was found. Notably, the hydrogen bonds between OH group at C8 position of **1** and Asn129 amide, and OH group at C9 and Arg119 CO group were maintained along most of the simulation. The NH group of Sia N-acetamide moiety formed a hydrogen bond with the Arg127 backbone which was steadily held along the simulation. Additionally, Tyr128 established key CH-π interactions with the C7 and C9 hydrogens of sialyl glycerol moiety. The only buried portion of **1** was the methyl group of Sia unit, which lied in a cleft constituted by Phe21 and Tyr128, thereby establishing strong CH-π interactions. In the complex, the Gal residue formed polar contacts with the binding site residues of the protein, as Asn129 and Glu66; whereas the GlcNAc residue pointed far from the receptor surface, consistently with STD NMR analysis.

The pattern of interactions of the MODEL IV/**1** complex was supported by the CORCEMA-ST results, that produced good agreement between experimental and predicted STD data, with an R NOE value of 0.29 (Figure 4.10c. The protons belonging to the Sia units exhibited the highest STD effects, confirming that Sia was the major determinant of the interaction with Siglec-10. The strongest STD value observed for the methyl group of the *N*-acetamide moiety of Sia corresponded to the close contacts between this moiety and Phe21 and Tyr128 side chains. This was also ascribable to the stable hydrogen bond established between *N*-acetamide moiety and Arg127 backbone that reinforced this interaction. Regarding STD data of the Sia glycerol side chain, the higher STD effect observed for the H7 in comparison with H8 and H9 was likely due to the CH-π interaction between H7 and Tyr128. Although the hydroxyl groups in position 8 and 9 made

direct hydrogen bonds with the backbone of Asn129, the protons H8 and H9S pointed away respect to the Tyr128 group, consequently leading to a STD effect lower than the other Sia protons. In spite of the contacts between the CC' loop of the protein with the galactose ring of ligand **1** bound conformation, weaker STD effect were predicted for the Gal protons. Moreover, the STD values estimated for the *N*-acetylglucosamine moiety were negligibly small, confirming that the residue was not involved in the interaction with the protein.
Overall, the NMR results (STD and NOE derived) combined with docking calculations, MD simulations and CORCEMA analysis allowed to describe the interaction between ligand **1** and Siglec-10 at molecular level.

4.2.3 Molecular interaction between Siglec-10 and trisaccharide 2

The binding of α(2,6) linked sialoglycan **2** by Siglec-10 was investigated by means of STD NMR technique (Figure 4.12) As expected, several relative enhancements and differences in the multiplicity of STD signals (Figure 4.12a) were detected. The epitope mapping of ligand **2** (Figure 4.12b) highlighted the protons specifically involved in the interaction with Siglec-10. The strongest STD signal belonged to the *N*-acetyl group protons of non-reducing **K** residue, providing an $STD_{epitope\ fit}$ value of 100% (Figure 4.12c, Table 4.3).

Table 4.3 Experimental STD epitopes of trisaccharide **2** bound to Siglec-10. STD intensities of ligand **2** bound to Siglec-10 experimentally calculated at different saturation times, using the monoexponential equation $STD(t_{sat}) = STD_{max} \times [1-exp(-k_{sat} \times t_{sat})]$.

^{1}H	STD_{max}	K_{sat}	STD (fit)	STD epitopes (fit)
Ac Sia	9.0345	0.2863	2.586577	**100%**
H6 Sia	4.4377	0.4472	1.984539	**76.7%**
H7 Sia	3.7108	0.4639	1.72144	**66.6%**
$H3_{eq}$ Sia	0.8736	0.5546	0.484499	**18.7%**
$H3_{ax}$ Sia	1.1098	0.7396	0.820808	**31.7%**
H9S Sia	1.4138	0.5461	0.772076	**29.8%**
H5 Sia	1.684	0.3123	0.525913	**20.3%**
H6R Sia	0.5481	0.7821	0.428669	**16.6%**
H4 Sia	0.9568	0.4301	0.41152	**15.9%**

Regarding the other protons belonging to Sia, only the H6 and H7 possessed significantly high STD values, thus implying a major role into the Siglec-10 recognition, while the other protons exhibited STD values around or below 30%. Furthermore, the Gal moiety (**B**) of ligand **2** seemed to be less involved in the interaction with the protein if compared to ligand **1**, since it received a weaker magnetization transfer from the protein, as indicated by the lower value of STD percentages. The GlcNAc (**A**) moiety of ligand **2** revealed no STD signal, suggesting that it was solvent exposed.

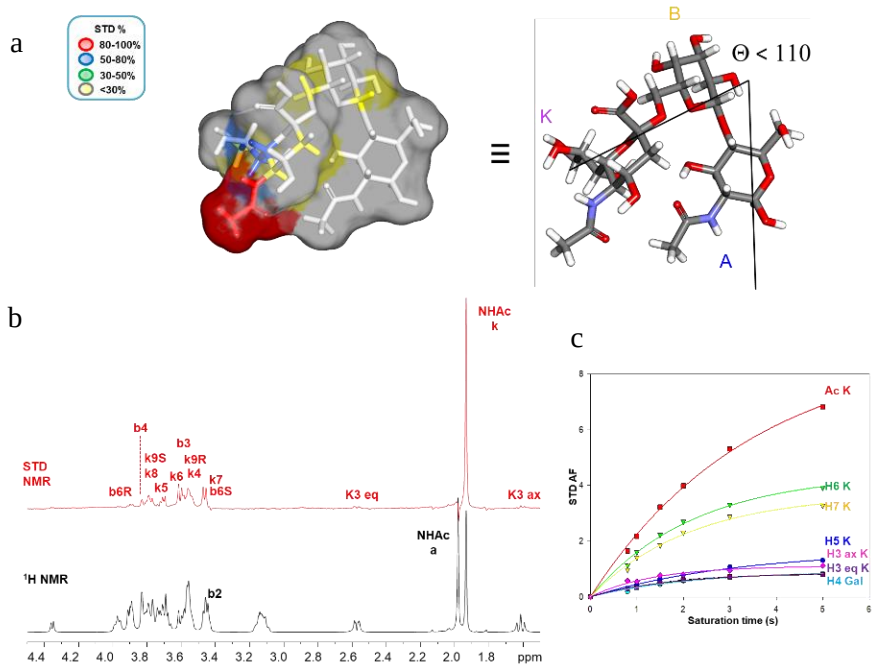


Figure 4.12 STD NMR analysis of the interaction between Siglec-10 and trisaccharide 2. a) Epitope map of ligand 2 as derived by STD data. The STD effects were calculated by $(I_0-I_{sat})/ I_0$ ratio. **b)** Reference ^{1}H NMR spectrum (black) and STD 1D NMR spectrum (red) of the Siglec-10/ligand 2 mixture (molecular ratio 1:150, temperature of 298K, magnetic field of 600MHz). **c)** STD build-up curves, derived by the monoexponential equation $STD = STD_{max}[1 - \exp(-k_{sat}t)]$. The acetyl group protons belonging to K residue of trisaccharide 2 gave the maximum signal, was set at 100% and the percentages of the other protons were normalized respect to the maximum**. d)** STD-derived epitope mapping on the molecular envelope of ligand 2 in its bioactive bent gt conformation.

Similarly to ligand **1,** the conformational behavior of ligand **2** in its free state and upon binding with Siglec-10 was determined by combining NMR techniques with Molecular Mechanics and Dynamics simulations. In this case, the Sia-α-(2-6)-Gal glycosidic linkage was defined by an extra torsion angle,[245,] named ω (O5−C5−C6−O2'), as already described in the previous chapter, in which the MD simulation of the free state ligand **2** was performed. This additional flexibility

element affected the entire 3D structure of trisaccharide **2** in terms of conformation. In detail, it has been extensively described that three different rotamers, namely *gg*/ *tg*/ *gt*[246] based on ω value of -60°/180°/60° respectively, could be defined, being the conformer characterized by ω= 60º (*gt* conformer) the most populated in the free state (Figures 4.12 and 4.13, see also Chapter 3).[214]

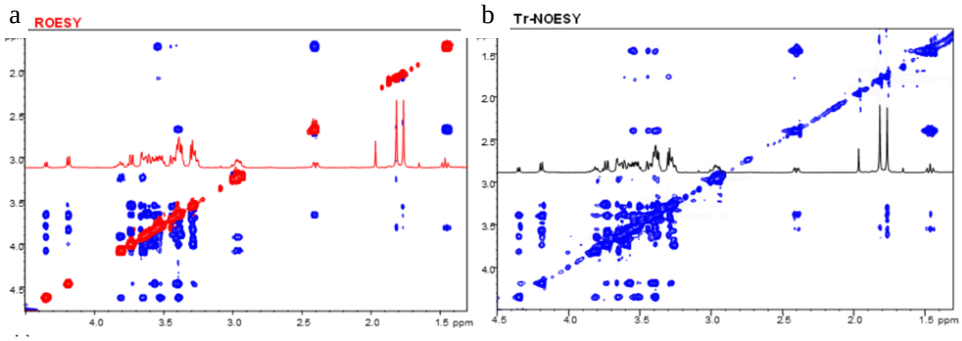


Figure 4.13 Conformational analysis of ligand 2. ROESY spectrum **a**) of ligand **2** in its free state (mixing time of 450ms) and tr-NOESY spectrum **b**) of ligand **2** bound to Siglec-10 (mixing time of 400ms) at 283K. The protein-ligand molar ration was set at 1:10.

Comparing the experimental NOE-derived and theoretical distances, (Table 4.4, Figure 4.13) a preference for the geometry with ϕ/ ψ/ ω torsion angles of -60°/180°/60° around Sia-Gal linkage (Family I, *gt* conformer, see also chapter 3)[211] was detected both in the free and bound state.

Table 4.4 Investigation of conformational behavior of ligand 2.

Distance	Exp Free state	Exp Bound state	Family I Φ = -60° Ψ = 180° Ω = 60°	Family II Φ = 180° Ψ = 180° Ω = 60°
H1 Gal -H4 GlcNAc	2.52	2.5	**2.4**	2.4
$H3_{eq}$ Sia -H4 Gal	/	/	**6.53**	5.62
$H3_{eq}$ Sia -H6S Gal	/	/	**4.93**	3.84
$H3_{eq}$ Sia -H6R Gal	/	/	**4.58**	3.37
$H3_{ax}$ Sia -H4 Gal	/	/	**6.34**	4.78
$H3_{ax}$ Sia -H6S Gal	4.2	4.3	**4.43**	2.53
$H3_{ax}$ Sia -H6R Gal	4.1	4.3	**4.25**	2.35
H5 Sia -CH_3GlcNAc	4	4	**4.3**	9.6

The *gt* preference of **2** corresponded to a bent conformation that implicated the back-folding of the sialic acid residue and thus characterized by an open umbrella-like topology (θ angle below than 110°),[212]Figure 4.12. Experimentally, this bent conformation was supported by the small NOE contact between H-5 of sialic acid residue and the methyl group of N-acetylglucosamine.

4.2.4 3D-View of Siglec10 in interaction with trisaccharide 2

To get a 3D perspective of the interaction of trisaccharide **2** with Siglec-10 CRD, docking calculations were performed by using the previously selected Siglec-10 model, considering different trisaccharide **1** starting geometries. MD simulations of the most representative docked complexes were performed to assess the stability of the binding poses. Then, to select the best 3D complex

according to STD data, such complexes were submitted to CORCEMA-ST calculations. The lowest R factor, with the value of 0.26, was obtained for the complex reported in Figure 4.14 in which the trisaccharide **2** adopted a *gt* conformation.

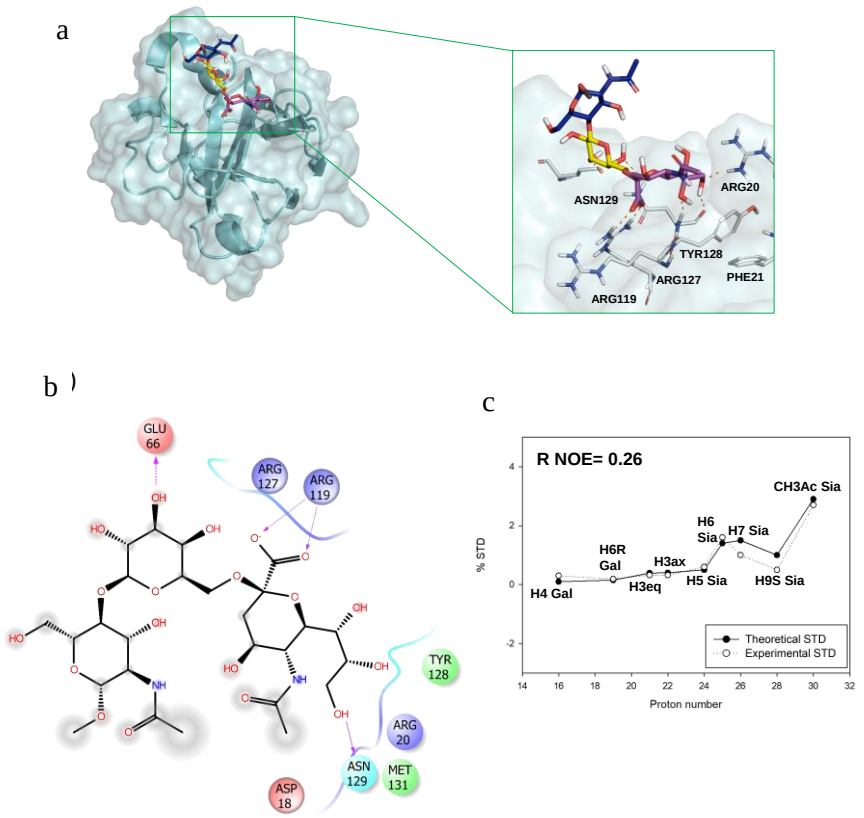


Figure 4.14 Docking/MD and CORCEMA-ST analysis of trisaccharide 2. a) Best binding pose (STD, tr-NOESY and MD based) predicted by the docking of the trisaccharide **2** bound form into the binding pocket of Siglec-10 model derived by using Siglec-7 as structural template. The carbohydrate and interacting amino acids are shown as sticks. Dashed black lines represents hydrogen bonds. **b)** Two-dimensional plot illustrating the interactions of the sialylated trisaccharide 2 with the residues in the binding pocket of Siglec-10 CRD model. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows such with functional groups of the backbone. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions. **c)** Comparison between experimental (dashed line) and theoretical (solid line) STD data for

the best 3D complex of the interaction between trisaccharide **2** and the homology model of Siglec-10 CRD, calculated by CORCEMA-ST program.

The terminal sialic acid was again the main determinant of ligand binding, establishing several polar interactions within Siglec-10 pocket, and that were comparable to those observed for the trisaccharide **1**/MODEL IV complex (Figure 15).

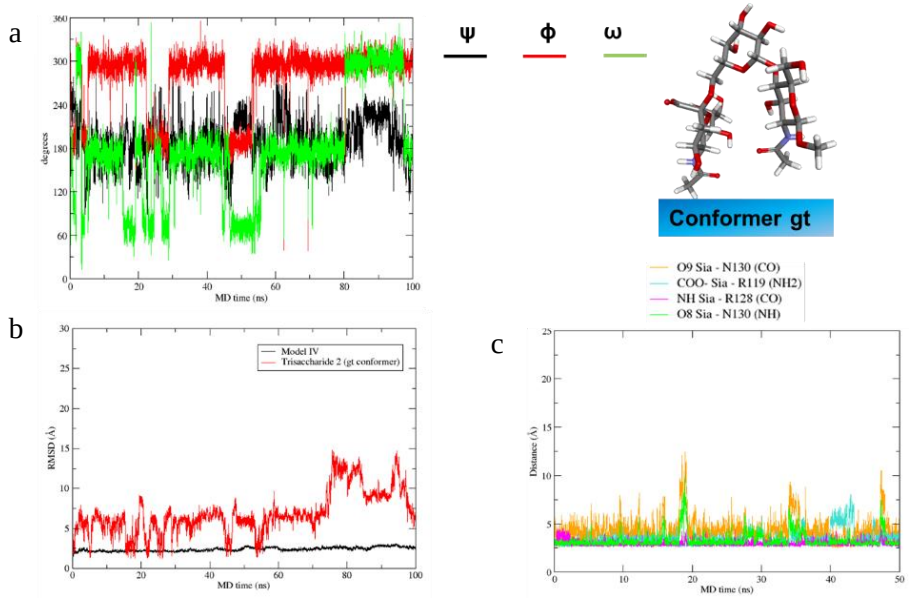


Figure 4.15 MD simulation trajectory analysis of trisaccharide 2/model IV complex. a) Dihedrals ϕ, ψ, ω of Sia-α-(26)-Gal linkage fluctuation in MD simulation of the bound state **b)** RMSD plot of trisaccharide **2** (gt conformer) in complex with Siglec/10 CRD (model IV) **c)** Representative intermolecular distances between Sia unit and the Siglec 10 residues calculated along the simulation.

Analogously to trisaccharide **1**, the methyl group of Siac lied in the hydrophobic groove defined by Phe21 and Tyr128. Furthermore, the Sia residue of **2** formed the key salt bridge between its carboxylate and the guanidinium group of Arg119, and hydrogen bonds between Asn129 backbone and the hydroxyl groups of the glycerol chain, namely OH8 and OH9 were also present. Notably,

these interactions remained stable along the MD simulation. As detected for trisaccharide **1**, it was detected a direct involvement of the Gal unit of **2** in the binding. Indeed, a strong hydrogen bond occurred between the Gal hydroxyl group in position 3 with Glu66 residue. As for ligand **1**, the protein did not engage interactions with the GlcNAc unit of **2**, in full conformity with the STD-NMR data.

From the MD simulation performed on the selected **2**/MODEL IV complex, it was possible to observe that, although the trisaccharide **2** remained steadily anchored to the protein via the sialic acid moiety, the bound ligand explored both *gt* and *tg* conformations along the dynamic. The ω angle, indeed, fluctuated between ω=60º and ω=180º. This suggested that ligand **2** displayed higher conformational flexibility, even upon binding, respect to ligand **1**. However, from the CORCEMA analysis, the best fit between experimental and theoretical data was obtained with complexes in which the ligand was in the *gt* conformation. In detail, from the CORCEMA prediction of the best-fit complex, the highest theoretical STD value was given by the methyl group of the *N*-acetamide moiety of Sia. This was imputable to the vicinity of the acetamide group toward the aromatic residues Phe21 and Tyr129. For the same reasons discussed for the trisaccharide **1**, the H-7 of the Sia glycerol side chain of trisaccharide **2** displayed higher STD effects respect to the protons in position 8 and 9. Concerning the galactoside ring of the trisaccharide **2**, lower STD effects were observed, despite the vicinity to the CC' loop of Siglec-10. In complete agreement with the experimental STD data, no remarkable STD effects were predicted for the GalNAc moiety. Thus, MD simulation and CORCEMA results fully supported the tr-NOE and STD-NMR analysis thus confirming the reliability of the 3-D complex here proposed.

4.3 Molecular interaction between Siglec-10 and complex-type N-glycans

The interaction between naturally occurring glycans and the Siglec-10 has also been investigated with the aim to disclose if a different presentation of the glycan epitope could influence its recognition by Siglec-10 Two biantennary complex-type undecasaccharide-Asn (ligands **3** and **4**, Figure 4. 3), bearing α(2,3) and α(2,6) linked sialic acid at glycan antennae termini, were used (Figure 4.16).

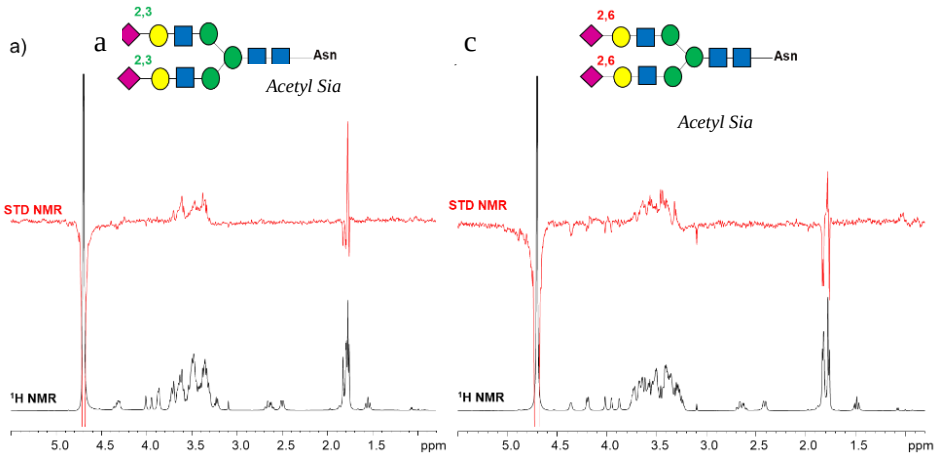


Figure 4.16 Analysis of the interactions between complex-type N-glycan and Siglec-10. a) STD NMR interaction between Siglec-10 and undecasaccharide **3**. **b)** STD NMR analysis of the interaction between Siglec-10 and undecasaccharide **4**.

Analysis of STD NMR spectra acquired in the presence of Siglec-10 revealed that the Sia residue was again the key sugar in the recognition and binding process. The results were indeed very similar to those obtained with the shorter ligands **1** and **2** respectively; the acetyl group of Sia displayed the strongest STD intensity and other STD enhancements corresponded to H5 – H9 protons of the Sia moiety. Furthermore, STD effects for H4 and H6 of the galactose unit, in the case of the α(2,6) undecasaccharide, and for H3-H6, in the case of the α(2,3)-

linked sialoglycan, were observed. In addition, the protons belonging to the other sugar residues of the branches and the N-glycan pentasaccharide core did not appear in the STD spectra, clearly suggesting that the recognition is mainly driven by the Sia-Gal moieties. The other sugar moieties were clearly solvent exposed and far from the protein surface. These results were further confirmed by docking calculations of ligands **3** and **4** at Siglec-10 binding site (Figure 4.17).

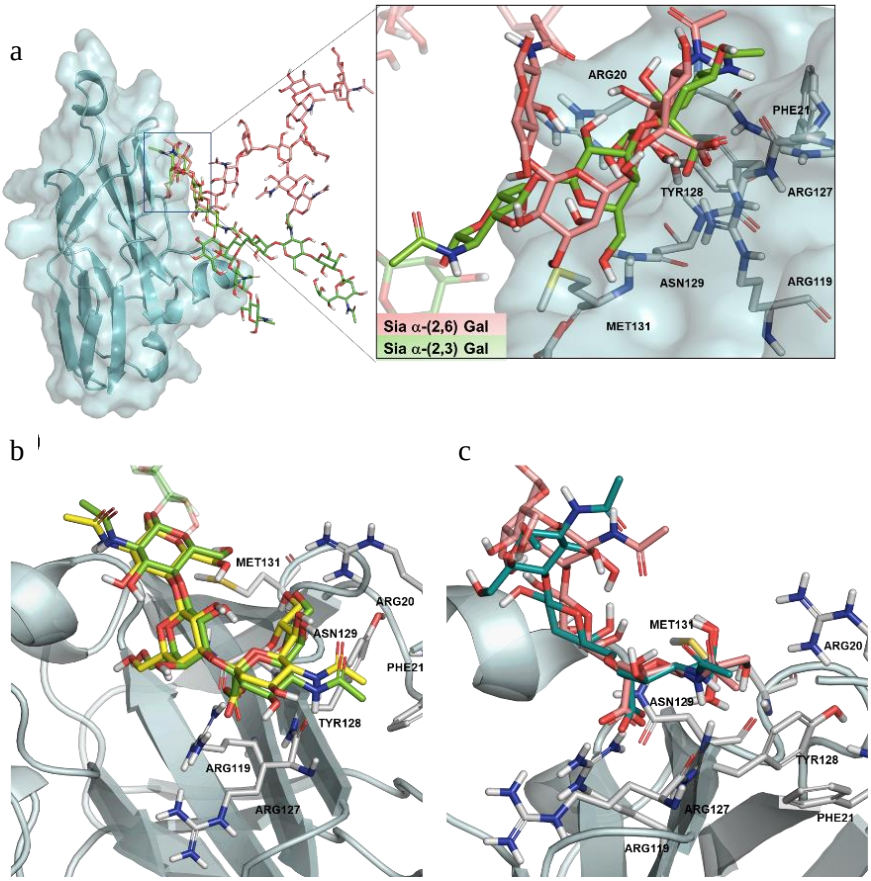


Figure 4.17 Molecular recognition of longer *N*-glycans structures by Siglec-10. a) Full view of the 3D complexes for the docking of ligand **3** (green) and ligand **4** (pink) using Siglec-10 (model IV) structural template. **b)** Superimposition of the Sia-α–(2,6)-Gal-β-

(1,4)-GlcNAc moieties from ligand **3** and ligand **4** into Siglec-10 (model IV) binding pocket. The carbohydrate and interacting amino acids are shown as sticks.

As shown in the three-dimensional structures of the corresponding complexes, longer complex type glycans can be accommodated in the protein binding pocket, with one glycan antenna establishing polar contacts with Arg119, Arg127, Asn129 and hydrophobic interactions involving Phe21 and Tyr128 as previously reported for the trisaccharides **1** and **2**, whereas the rest of the oligosaccharide is solvent exposed and not directly involved in the binding, in accordance with STD NMR results.

4.4 Recombinant expression of Siglec-10 in *Escherichia coli*

With the aim to perform structural studies by NMR on Siglec-10, the recombinant expression of the CRD of Siglec-10 in *Esecherichia coli* (*E. coli*) was attempted in collaboration with Giotto Biotech. First, the sequence encoding for Siglec-10 carbohydrate recognition domain (aa 18-140) containing a C36S mutation and a 6X-His tag at the C-terminus was subcloned in PET29b plasmid (Figure 4.18). The C36S mutation was introduced to prevent the mismatch of the disulphide bridge between Cys41 and Cys101, crucial for the appropriate folding of Siglec-10 V-set domain, with the aim to avoid refolding procedures.

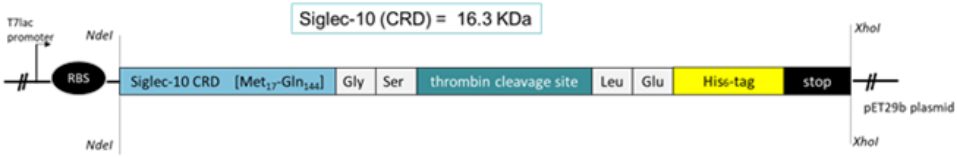


Figure 4.18 Representation of the employed Siglec-10 expression construct.
The nucleotide sequence encoding for Siglec-10 CRD (light blue box) was cloned into NdeI and XhoI restriction sites of the pET29b plasmid. Following the Siglec-10, Gly-Ser spacers for the insertion of the thrombine cleavage site encoding nucleotides (deep teal box), to cleave the terminal 6X His tag after the purification. Leu-Glu spacers were

included and then followed by 6X His tag (yellow box) and ending with a stop codon (black box).

Then, the expression level of Siglec-10 was tested in LB broth using several *E. coli* strains (BL21(DE3) Gold, BL21 (DE3) CodonPlus RIPL, C41(DE3), Rosetta (DE3) pLysS, Origami-2 pLysS). In detail, BL21(DE3) Gold were chosen as they showed optimized expression levels than the standard BL21(DE3), the BL21 (DE3) CodonPlus RIPL for the capability to for this strain to rescues expression of heterologous proteins from organisms that have either AT- or GC-rich genomes. C41 (DE3) were used for the possibility to avoid cell death associated with expression of recombinant toxic proteins. Rosetta (DE3) pLysS were considered for the ability to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* along with T7 lysozyme activity to suppress basal T7 expression. Last, Origami-2 pLysS was considered for the ability to facilitate proper disulfide bond formation together with T7 lysozyme expression.
The tested induction temperatures were 15°C- 25°C- 37°C and induction times of 0 h/ 4 h/overnight (o. n.), IPTG concentration of 1 mM. The results were assessed by SDS-PAGE (Figure 4.19). In the tested conditions, no expression of Siglec-10 (theoretical MW=16,3 KDa) was observed.

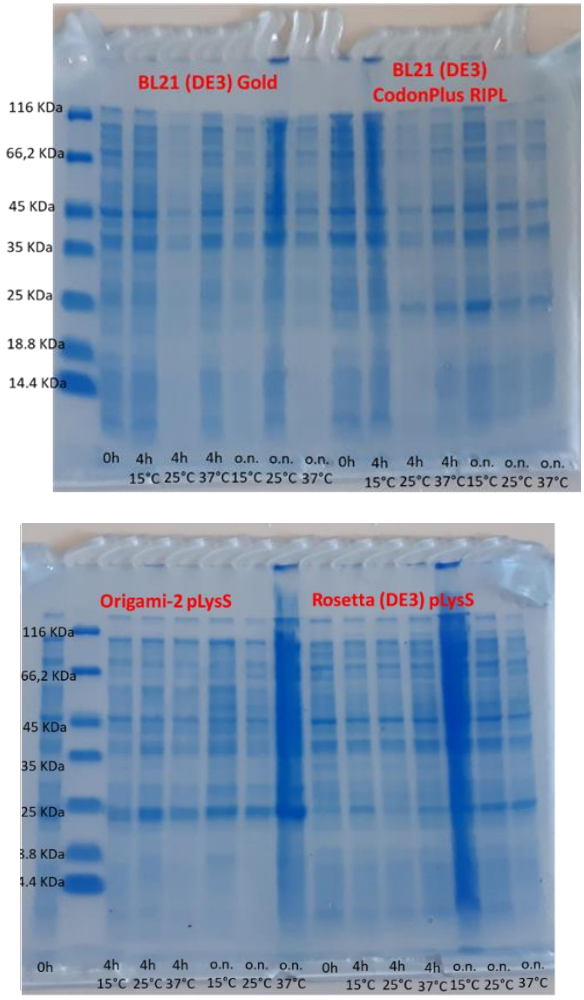


Figure 4.19 Coomassie blue-stained reducing SDS–PAGE (20%) gel of cell extracts obtained from different induction conditions by small-scale expression. For better comparability, approximately equal amounts of samples were loaded on each lane. The expression profiles showed that Siglec-10 was not expressed in any condition.

Therefore, based on the expression protocol of Siglecs in *E. Coli* developed by Pröpster et al.,[247] preparative expression of Siglec-10 was performed in Origami-2 and Rosetta strains, with induction temperature 15°C and induction

time 38 h, inducer concentration 1 mM. The solubility of the protein was checked by SDS-PAGE (Figure 4.20).

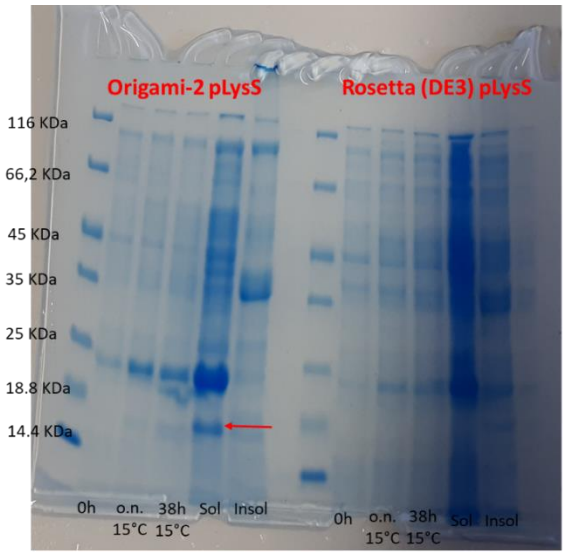


Figure 4.20 Large scale expression of Siglec-10 in Rosetta and Origami-2. Expression profiles of Siglec-10 in Rosetta (DE3) pLysS and Origami-2 pLysS strains at 15 °C induced with 1 mM IPTG for 38 h. Siglec-10 was expressed in the soluble fraction only in Origami-2 strains.

As result, no expression of Siglec-10 was observed in Rosetta strain, whereas a band corresponding to Siglec-10 MW was observed in the soluble fraction for Origami-2 strain. Despite this, the protein was not obtained in its pure form, as it did not show affinity for Ni^{2+}–NTA column. Other techniques including ammonium sulphate precipitation, Cation Exchange and Gel filtration chromatography, were unsuccessful.

For the expression of uniformly (^{15}N, ^{13}C) isotope-labeled proteins, M9 minimal medium supplemented with 1 g/L of $^{15}NH_4Cl$ and 3 g/L glucose was used, following the same expression conditions of LB cultures (Origami-2 strain, IPTG

1 mM, induction time 15°C and induction temperature 38h). However, Siglec-10 was not expressed in M9 medium.

So, gateway cloning[248] of pENTR_D_TOPO (Siglec-10) plasmid in pDEST plasmids containing HGB1, GST and MBP fusion tags was performed with the aim to produce the fusion constructs that may increase the expression yield and solubility (Figure 4.21).

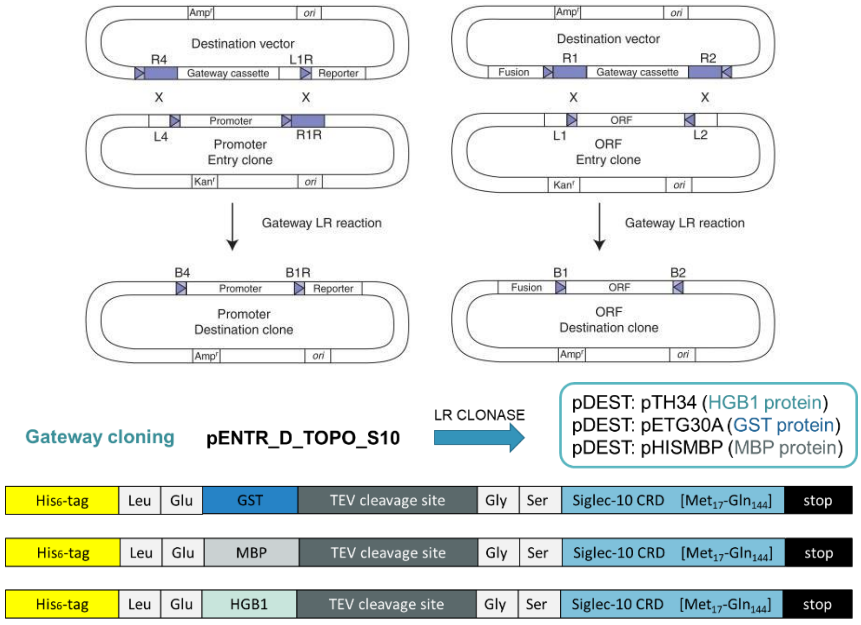


Figure 4.21 Expression constructs for Siglec-10 fusion proteins obtained by gateway cloning. a) The destination clone is generated using a Gateway LR reaction. In a Gateway LR reaction, an entry clone (pENTR_D_TOPO_S10) is mixed with a destination vector, and the LR clonase enzymes recombine the *att*L and *att*R sites of the matching subtype (i.e., *att*R4 with *att*L4, *att*R1 with *att*L1), exchanging the Gateway cassette with the cloned insert. For instance, a reporter construct with an open reading frame (ORF) and/or a regulatory DNA fragment (e.g., a promoter)—into a plasmid that encodes the reporter protein, ORF cloned in-frame with an amino-terminal reporter such as GST. Figure from ref. 246. **b)** Expression constructs obtained in this Thesis.

The expression of each plasmid was tested using BL21(DE3) Gold, BL21 (DE3) CodonPlus RIPL, C41 (DE3) strains, induction temperatures (15°C, 25°C, 37°C) and induction times (0 h, 4 h, o. n.), induced with 1 mM IPTG. The SDS-PAGE analysis revealed good levels of expression for most of the tested conditions. In particular, the best expression yeld in M9 was observed for MBP fused Siglec-10 CRD in BL21 (DE3) Gold, 18°C, 0.5 mM IPTG, as shown in Figure 4.22.

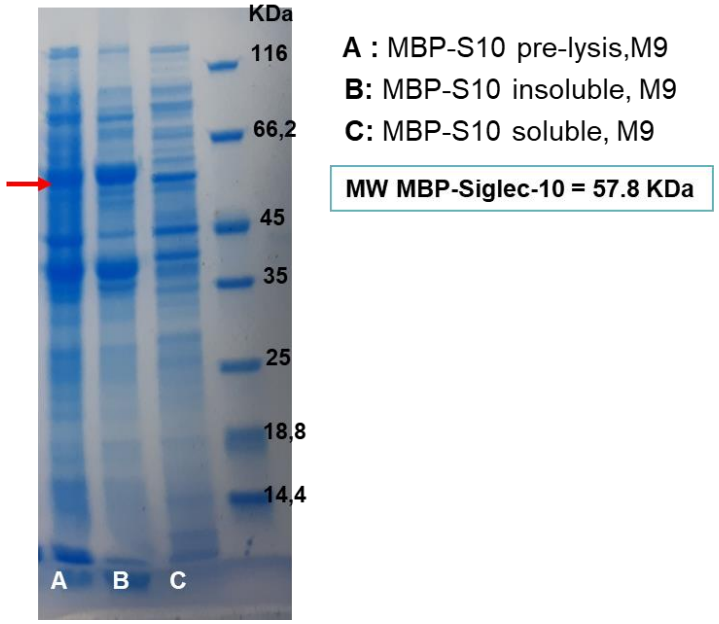


Figure 4.22 SDS-PAGE of the best conditions for the expression of MBP fused Siglec-10 in M9 medium. Expression profiles of Siglec-10 in BL21(DE3) at 15 °C induced with 0.5 mM IPTG for 4h. MBP fused to Siglec-10 was expressed in the insoluble fraction.

However, solubility tests showed that the Siglec-10 fusion proteins formed inclusion bodies in all the tested conditions, making difficult refolding and purification of the fusion proteins. All the results are summarized in table 4.5. For the solubility tests, either 500 mL/ 1L of cell cultures were carried out in the conditions listed in the table below, then cells were harvested by centrifugation an

subjected to lysis by sonication (see methods for further details). Then, the soluble (supernatant) and insoluble (pellet) fractions were separated by ultracentrifugation of the lysate and samples were dissolved in Laemmli buffer. The fractions were subjected to SDS-PAGE to check the presence of the protein in the soluble/ insoluble lane.

Table 4.5 List of solubility tests performed for Siglec-10 fusion proteins.

Plasmid	Media	Quantity	Antibiotics	Additives	C IPTG	Induction T	Induction time	Solubility
GST-Siglec10	LB	1 L	Amp	/	1 mM	37°C	o.n.	Not soluble
GST-Siglec10	M9	1 L	Amp	$MgCl_2$,$CaCl_2$, Glu, $(NH_4)_2SO_4$	1 mM	37°C	o.n.	Not soluble
GST-Siglec10	LB	1 L	Amp	/	1 mM	25°C	o.n.	Not soluble
GST-Siglec10	M9	1L	Amp	$MgCl_2$,$CaCl_2$, Glu, $(NH_4)_2SO_4$	1mM	25°C	o.n.	Not soluble
GST-Siglec10	M9	500 mL	Amp	$MgCl_2$, $CaCl_2$, Glu, $(NH_4)_2SO_4$	1 mM	18° C	4h	Not soluble
GST-Siglec10	M9	500 mL	Amp	$MgCl_2$, $CaCl_2$, Glu, $(NH_4)_2SO_4$	1 mM	18°C	o.n.	Not soluble
GB1-Siglec10	M9	500 mL	Amp	$MgCl_2$, $CaCl_2$, Glu, $(NH_4)_2SO_4$	1 mM	18° C	4h	Not soluble
GB1-Siglec10	M9	500 mL	Amp	$MgCl_2$, $CaCl_2$, Glu, $(NH_4)_2SO_4$	1 mM	18°C	o.n.	Not soluble
MBP-Siglec10	LB	1L	Amp	/	0.5 mM	18°C	3h	Not soluble
MBP-Siglec10	M9	1L	Amp	$MgCl_2$, $CaCl_2$, Glu, $(NH_4)_2SO_4$	0.5 mM	18°C	3h	Not soluble

4.5 Discussion

In the present study, the interaction of Siglec-10 with different sialoglycans which constitute the terminal end of the canonical sialylated complex type *N*-glycans exposed on mammalian cells was investigated by an integrated approach based on NMR, docking and molecular modelling. The orthogonal methods here applied allowed to describe, at the molecular level, the recognition and binding processes affecting Siglec-10 and sialylated glycans. The ligands' epitope in their bound conformations was profiled and consistent 3D-models of the interaction were provided. By the comparison of the epitope maps obtained for ligands **1** and **2**, it was possible to affirm that in both substrates the binding process to Siglec-10 was driven by the sialic acid moiety. In each system, sialic acid and galactose portions were recognized by the protein, while the *N*-acetylglucosamine pointed out from Siglec-10 binding pocket. The sialic acid residue was the most involved unit in the binding with Siglec-10. Nonetheless, it was shown that the sialic acid unit was the main residue recognized by Siglec-10 receptor, independently from the linkage connecting to the Gal unit.On the other hand, some differences were revealed in the STD epitope maps indicated that the ligand **2** established less contacts within the residues of the protein binding pocket, as indicated by the lower involvement in the interaction of protons H5, H8 and H9 of sialic acid unit.

Next, the bound conformation of both sialoglycans was defined using NOE based data in combination with a computational approach including MM and MD simulations. Such analysis demonstrated that the ligand **1** adopted a preferential conformation when bound to Siglec-10; in detail, the selection of the *-g* conformer upon binding was revealed. On the other hand, it was deduced that the ligand **2** adopted a bent *gt* conformation both in the free state and in the bound

state. As the crystal structure of Siglec-10 has not been resolved to date, the extensive homology modelling studies here reported provided novel structural insights into the molecular basis of *N*-glycans recognition by human Siglec-10. Then, docking and MD studies, together with CORCEMA-ST protocol, were used to achieve key information into the fine structural characteristics of the complexes. Taken together, the results indicated the likely reliability of the simulated structure of the human Siglec-10. The three-dimensional structure of the Siglec-10-sialoglycan complexes highlighted the architecture of the binding pocked essentially composed by hydrophobic (Phe21), polar (Tyr128, Asn129), basic (Arg119, Arg127) and acid (Glu66) residues. In detail, the Arg119, Arg128, Asn129 residues were directly involved in the interaction with Sia, suggesting they shaped a region relevant for the molecular recognition, which matches the previously studied members of Siglec family.[89]Moreover, it was observed that the CC' loop belonging to the Siglec-10 V-set domain established key interactions with the adjacent galactose moiety that appeared to participate in the interaction process, while the GlcNAc residue showed no relevant contribution to the binding.
However, it is worth to remark that, similarly to other Siglecs, including Siglec-2, Siglec-3, Siglec-5, Siglec-7 and Siglec-8, the Gal moiety has an important role to the recognition, besides the Sia moiety.[9] This observation allowed to further shape Siglecs molecular recognition.

Nonetheless, the ability of Siglec-10 to recognize both α(2,3) and α(2,6) linkages indicate a lower specificity[195]of the protein with respect to other homologous Siglecs and it is probably due to a more open binding site. The CC' loop shape predicted for Siglec-10, indeed, points outward to the binding residues, in contrast to other CD33-like Siglecs (Siglec-3, Siglec-8, Siglec-5). For this reason, the protein could easily accommodate the trisaccharide **1** in its extended topology

and in parallel may enable the conformational flexibility of the trisaccharide **2** (Figure 4.22).

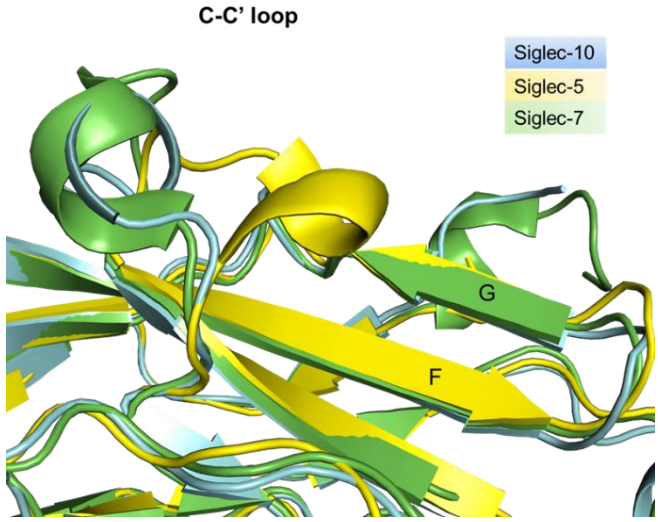


Figure 4.22 Superimposition of Siglec-10 model with Siglec-5 and Siglec-10 crystal structure to highlight the open loop conformation. Siglec-10 CRD is shown in cyan, Siglec-5 CRD (PDB ID: 2ZG2) is represented in yellow and Siglec-7 CRD (PDB ID: 2HRL) is represented in green. The beta strands outlining Siglec binding site are indicated in capital letters.

The results here shown improved the knowledge of the molecular mechanisms occurring between Siglec-10 and sialylated glycans involved in several crucial patho-physiological processes.[89] Given that the Siglec-10 is an appealing target for the development of glyco-based immunotherapy,[205] the study of Siglec-10-sialylated glycans interactions could provide new ideas for the design and optimization of synthesis of high-affinity ligands in their bioactive conformation, hence decreasing the entropy costs upon binding to Siglec-10. To date few data regarding the development of high affinity ligands specific for Siglec-10, also due to the lack of structural information concerning the receptor. Moreover, most of

the available studies are directed toward α(2-3) sialosiodes.[249,250] Given the structural similarities between CD22 and Siglec-10 that raised from the present work, it may worth to probe the binding of the STnThra analogue (see Chapter 3) by Siglec-10 to compare the mode of recognition. Similarly, many glycomimetics developed for CD22 may be tested for Siglec-10 binding.

Concerning the optimization of the expression and purification of labeled Siglec-10, with the goal to further characterize the interaction between Siglec-10 and sialylated ligands by means of receptor-based NMR techniques, the trials performed along the period spent at Giotto Biotech were not successful. However, future plans include the test of other expression strategies, such different host systems and using a longer construct including the C-2 set domain adjacent to the CRD domain of Siglec-10. These studies may allow for the Siglec-10 3D structure determination, thus ultimately validating the proposed interaction models.

Chapter 5- Molecular basis for the recognition of *Fusobacterium nucleatum* lipopolysaccharide by Siglec-7

5.1 Introduction

5.1.1 Siglec-7

Siglec-7 is an inhibitory receptor that is predominantly expressed on innate lymphoid NKs, but also found on eosinophils, monocytes, T cells and DCs.[251,252,253,254] Siglec-7 belongs to the CD33-related family and, as all Siglecs members, features a sialic acid binding N-terminal V-set Ig domain at the extracellular domain/portion/region, that is linked to two C2-set Ig spacers. At the cytosolic region, it contains an ITIM and an ITIM-like motif, that confer inhibitory properties to the receptor.

The 3D structure of Siglec-7 V-set domain was the first to be solved among CD33-related Siglec and has been widely studied in complex with several sialylated ligands (Figure 5.1).[254]

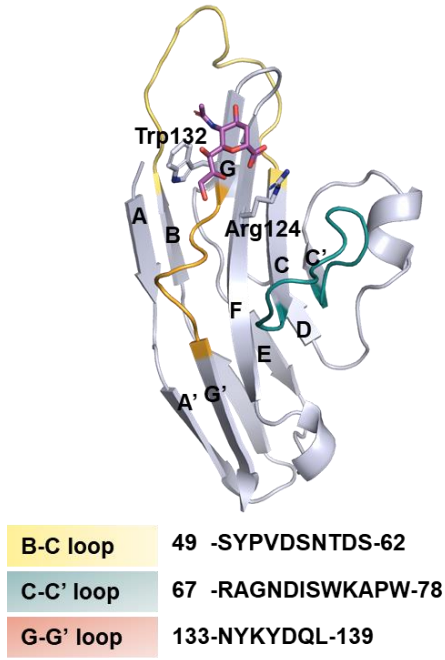


Figure 5.1 Structure of the V-set domain of Siglec-7 in complex with sialic acid. Side chains of amino acids that make direct contact with sialic acid are shown as stick. The C–C′ loop that is implicated in fine specificity of Siglec binding is shown in green. The B-C and the C-C' loop shaping the binding site are shown in yellow and orange, respectively (PDB ID: 2HRL).

The crystal structure highlighted the crucial sialic acid recognition and binding features including the key Arg124 residue located on the β strand F and a CC′ loop, which unveiled/brought a notable contribution to sialylated ligands specificity.[254] Interestingly, Siglec-7 showed a preferred binding specificity towards the Siaα2,8Sia epitope which is typical of some gangliosides like GD3, a disialoganglioside, even if some branched α(2,6) sialyl residues from other gangliosides such as DSGb5 and DSLc4, are also known as preferred binding molecules (Figure 5.2).[255] Notably, it was demonstrated that Siglec-7 undergoes

conformational changes in the association with the GT1b α2,8- disialyl ganglioside, allowing for the establishment of interactions with the branch point of the glycan.[256]

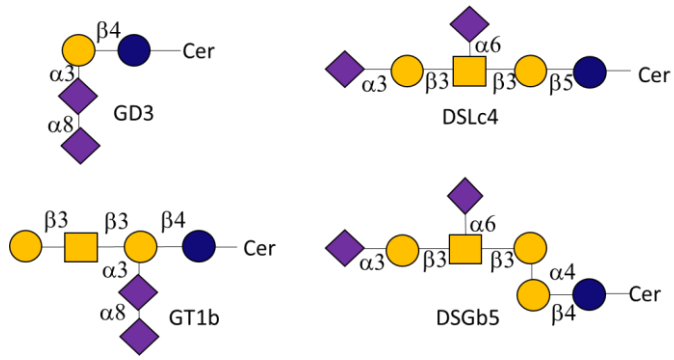


Figure 5.2 Structure of some of the ganglioside structures that are known to bind Siglec-7. The SNFG nomenclature for glycans was used.

Recently, the existence of an additional sialic acid-binding region, containing Arg residue, in addition to the well-known primary ligand-binding region was discovered, implying a mutual allosteric modulation upon sialylated glycans binding and possibly mediated by the CC' loop flexibility.[257]

In general, Siglec-7 is able to sense *self*, thus functioning to impede damage to normal cells and tissues.[258] Being majorly expressed on NK cells, it is associated to the negative regulation of NK cell-mediated functions, which are crucial within tumor immunosurveillance.[126] Furthermore, cancer cells that usually display Sias on their surface are able to engage Siglec-7, and this binding leads to its translocation to the activating receptor, consequently dampening NK cell–mediated cytotoxicity.[126,127,251] Thus, Siglec-7 has recently emerged as target molecule for cancer immunotherapy, as the hypersialylation on cancer cell surface contributes to escape NK cell activity (Figure 5.3). In detail, when the trans binding of cognate ligands to Siglec-7 occurs, its phosphorylated ITIM sites recruit

phosphatases SHP1/2 which impede the NK cell activating pathways such as the NKG2D pathway, allowing the tumor cell to evade immunity and endure the migration within the circulatory system.[127,252,259]

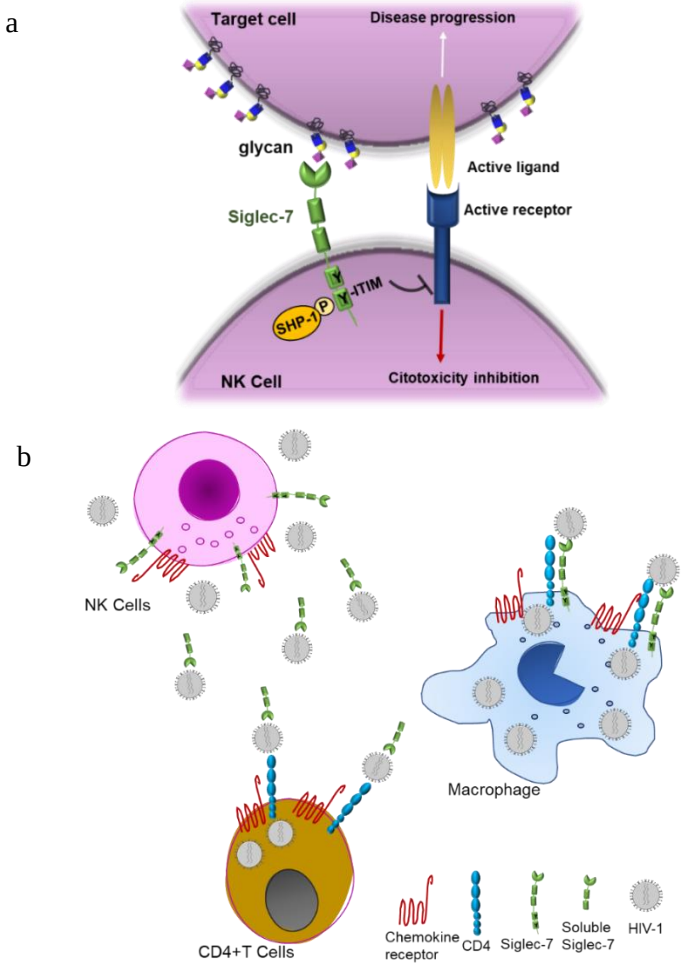


Figure 5.3 Mechanisms of immunomodulation by Siglec-7. a) Inhibition of NK cell cytotoxicity against tumor cells mediated by Siglec-7. *Trans* binding of Siglec-7 to its ligand results in the Src kinase-mediated phosphorylation of the ITIM motif of Siglec-7. **b)** Siglec-7 and sensitivity to HIV-1 infection on CD4+ T cells and macrophages**.**

Notably, besides the central role in cancer, the involvement of Siglec-7 in relevant pathologies including obesity, hepatitis, HIV-1 has emerged over the last years.[260,261,262] As for HIV-1 infection, it has been shown that the high levels of HIV-1 replication in patients affected by AIDS causes the decreased expression of Siglec-7 on NK cells surface and this increases the levels of soluble Siglec-7 in the infected plasma. The soluble form of Siglec-7 increased the sensitivity of CD4+ T cells towards HIV-1 infection. This effect of increased susceptibility to HIV infection was also demonstrated for CD4 expressing macrophages, even though the decrease of constitutive expression on monocytes and macrophage of Siglec-7 was not observed. This observation suggested that Siglec-7 spares from infection NK cells that, although expressing chemokine receptors, do not express CD4 receptor (Figure 5.3).[262,263,264]

Siglec-7 is also found on T cells subsets (CD8+/CD3+) where it participates to the regulation of the T cell receptor, thus contributing to set the activation threshold of T cells and consequently limiting activation-induced cell death and supporting the survival of memory T cells.[265] Interestingly, it has been revealed that Siglec-7 can promote the production of several pro-inflammatory cytokines and chemokines without the involvement of Sia, but by prompting the stimulation of inflammatory responses selectively on monocytes (not in T and NK cells) upon antibody crosslinking.[266]

Moreover, the function exerted by Siglec-7 on human eosinophils was addressed, illustrating the capability of the receptor to downregulate eosinophil activation, with potential therapeutic applications for the efficient treatment of eosinophil-derived illnesses such as asthma and allergies.[267] Importantly, Siglec-7 negatively regulates pathogen-associated molecular pattern receptors on dendritic cells in a sialic acid-dependent fashion.[268] In this context, the binding of Siglec-7

to *Pseudomonas aeruginosa* and *C. jejuni* was demonstrated, even though the interested signalling pathways and/or possible mechanisms of endocytosis still need to be fully disclosed.[269,270]

5.1.2 *F. nucleatum*

The anaerobic Gram-negative bacterium *F. nucleatum* is commonly found within the human oral cavity, and it is a prominent component of the oral microbiota.[271,272]Originally *F. nucleatum* was considered a mutualist, being in association with the host cells and tissues of the oral cavity and, likewise, with the other bacteria of the oral microbiota. Here, *F. nucleatum*, act as mediator between the oral mucosa colonizing bacteria, hence performing essential roles for the biofilm formation, accounting for the dental plaque (Figure 5.4).[273]
F. nucleatum also expresses numerous adhesins that are responsible for the communication with host cells and tissues.[274]

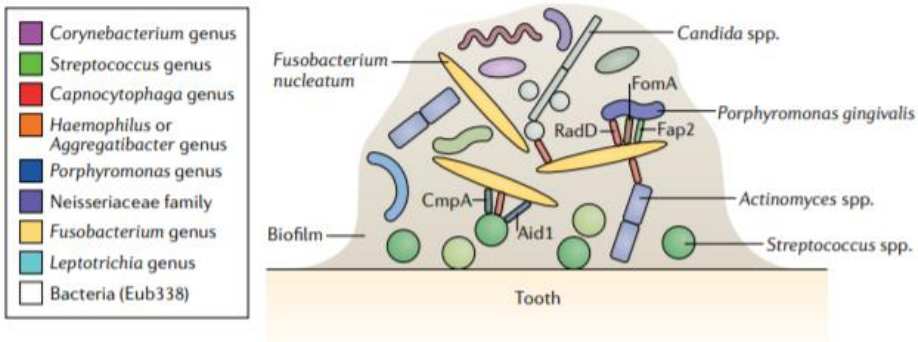


Figure 5.4 The organizing role of *F. nucleatum* in oral biofilm (Figure from ref. 274)

Nevertheless, the pathogenic disorders connected to its presence both inside the and outside the human oral cavity, recently highlighted *F. nucleatum* as a potential pathosymbiont.[274]

Specifically, in the oral epithelium, *F. nucleatum* is involved in the progression of periodontitis infection as it contributes to increase the virulence of other pathogens by prompting the production of antimicrobial peptides and pro-inflammatory citokynes.[274]

Outside the oral cavity *F. nucleatum* contributes to the etiology of other infections and diseases, including intrauterine infections, leading to preterm birth, miscarriage, neonatal sepsis, as well as gastrointestinal diseases such as inflammatory intestine disease, appendicitis and colorectal cancer.[275]

Hence, the *F. nucleatum* commensal/pathogen duality highlighted the importance to determine the molecular basis of this behavior. This led to the identification of possible virulence factors, although the molecular mechanisms underlying *F. nucleatum* morbidity still must be disclosed.[274,276]

Within this framework, the LPSs surrounding the outer membrane of *F. nucleatum* are considered possible element of virulence.[277] Indeed, as explained in the first chapter, LPSs are PAMPs and perform key functions in the bacteria-host interplay, as they are recognized by host PRRs, that consequently elicit the immune response. Generally, LPSs activates the dimerization of the Toll-like receptor 4 (TLR4) and the myeloid differentiation factor 2 (MD-2) complex, that occurs because of the recognition of the LPS lipid A. The dimerization event in turn triggers a signalling cascade resulting in the production of pro-inflammatory cytokines.[278] In this context, it was demonstrated that *F. nucleatum* LPS is able to trigger B lymphocytes[279] as well as to induce the expression of tumors necrosis factor alpha (TNF-α) and interleukin 8.[274]

Consequently, many efforts have been made toward the structural characterization of LPS from several *F. nucleatum* strains,[280,281,282,283 ,284,285,286] carried out with the intent of establish structure-activity relationships. Recently, the structure of the O-antigen of *F. nucleatum* strain 10953 was determined, leading to the

identification of sialic acid as a component of the O-antigen repeat.[287] The structure of the O-antigen repeating unit is reported below:

-[→4)-α-Neu*p*5Ac-(2→4)-β-D-Gal*p*-(1→3)-α-D-Fuc*p*NAc4N*Ac*-(1-]-
N B A

Thus, given that *F. nucleatum* cloack its cell surface with sialylated structures, it has been evaluated, and is recently under investigation, if and how this oral pathogen recognizes and interacts with host Siglecs, demonstrating that Siglec-7 is exploited to escape immune defense (Prof. Nathalie Juge group, Quadram Institute, Norwirch). Indeed, they established that the immune response elicited by some LPS from *F. nucleatum* is mediated by Siglec-7. In this regard, the goal of the here reported work has been to dissect the molecular basis of the recognition of the sialylated LPS from *F. nucleatum* 10953 by Siglec-7, through ligand-based NMR techniques, biophysical assays and *in silico* methodologies, to propose a structural model of the interaction. Globally, the acquired results allowed to determine the elements of *F. nucleatum* LPS binding by Siglec-7, thus improving the knowledge on the mechanism of action of this opportunistic oncopathogen.

5.3 STD NMR and fluorescence studies of *F. Nucleatum* LPS by Siglec-7

First, the LPS of *F. nucleatum* was extracted from 10953 strain by hot aqueous phenol method[288] and purified as described in the experimental section (see chapter 9). Then, the O-polisaccharide (OPS) was isolated from the lipid A portion upon mild acid hydrolysis.[2] The chemical structure of the OPS, recently established[287] showed the presence of α-N-acetylneuraminic acid (Sia, **N**),

[2] Performed by PhD Ferran Nieto- Fabregat

β–galactopyranose (Gal, **B**), and 2-acetamido-4-amino-2,4,6-trideoxy-α-galactopyranose (FucNAcN, **A**). Moreover, partially acetylated forms (above 30%) of **A** unit at position 4 of the FucNAcN were observed.

Once isolated and purified, the recognition and binding events between the *F. nucleatum* 10953 and Siglec-7 was explored by STD NMR experiments (Figure 5.5).

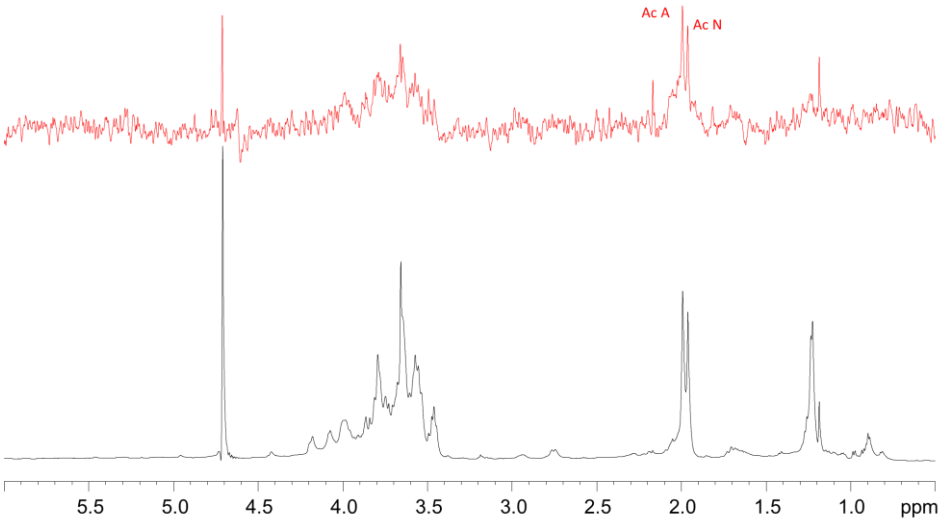


Figure 5.5 STD-NMR experiment of the Siglec-7/ *F. nucleatum* O-chain mixture. The reference ^{1}H-NMR spectrum (black) and STD 1D NMR spectrum (red) are shown (T=298K, 1:20 ratio). .

From the STD NMR it was possible to qualitatively assess the recognition of *F. nucleatum* OPS by Siglec-7. Particularly, the involvement of **N** and **A** residues was detected, with their STD signals showing a rather similar relative intensity, as observed for the acetyl signals of sialic acid (Sia, **N**) and fucosamine (FucNAcN, **A**) at about 2 ppm.

Next, with the aim of exactly identifying the regions of the polysaccharide more involved in the recognition and binding process, the OPS was depolymerized and purified by gel filtration chromatography (see experimental section); the isolated

oligomers were tested in the interaction with Siglec-7. A partially acetylated (at position 4 of the FucNAcN, above 30%) reducing trisaccharide was isolated (Figure 5.6), and interestingly no interaction was detected when performing the STD NMR experiment, suggesting that a higher number of OPS repeating units were required to establish a significant interaction with Siglec-7 receptor (data not shown). This was also confirmed by performing a fluorescence quenching titration of Siglec-7 using increasing concentrations of a solution containing the reducing trisaccharide fraction. As result, no decrease of the fluorescence intensity was observed even at higher protein ligand ratio, suggesting that no fluorescence quenching occurred, and thus no interaction was detectable by intrinsic fluorescence (Figure 5.6).

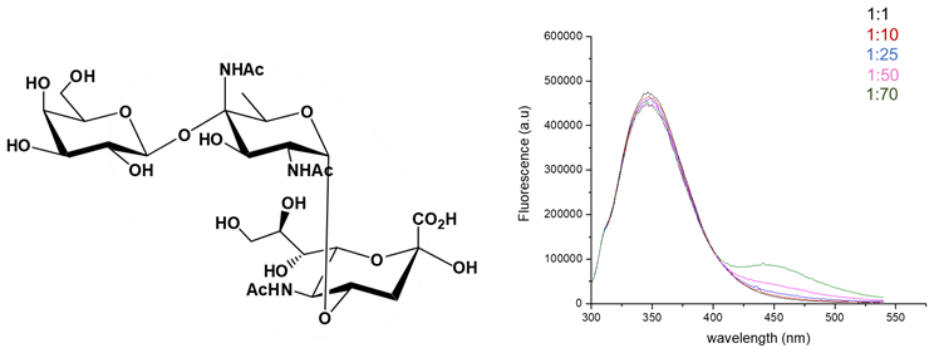


Figure 5.6 a) Fluorescence spectra overlay of Siglec-7 at increasing trisaccharide concentrations.

The second isolated fraction consisted in an oligosaccharide region composed by two repeating units, Figure 5.7. The STD NMR showed signals belonging to the oligosaccharide repeating unit, highlighting Siglec-7 ability to recognize constituting *F. nucleatum* OPS. Interestingly, no STD signals relative to the core region were observed in the STD spectrum, showing that Siglec-7 did not bind this LPS region.

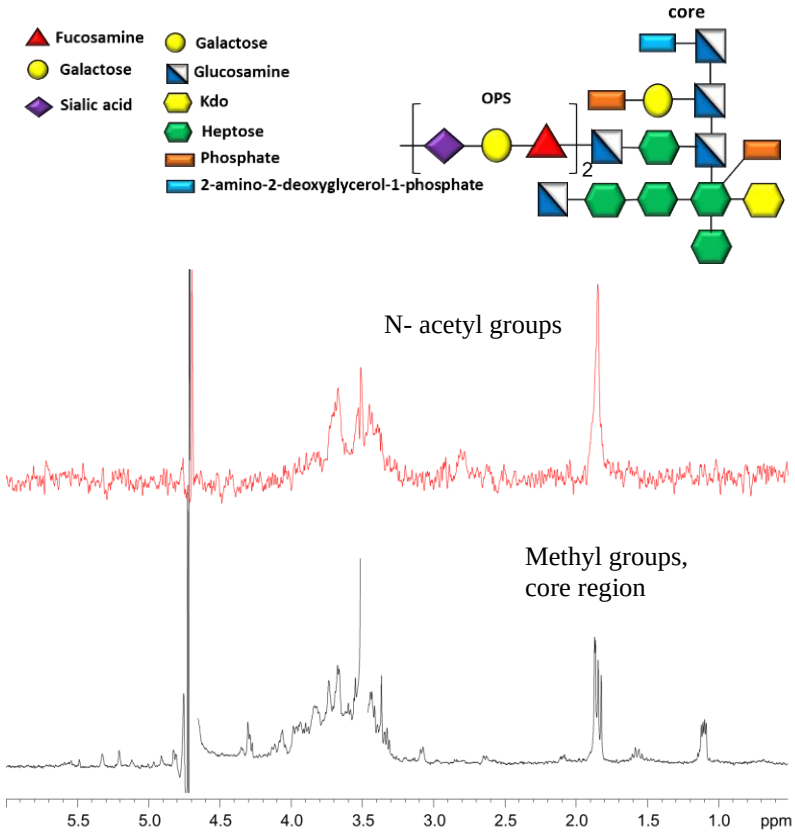


Figure 5.7 STD-NMR experiment of Siglec-7 and *F. nucleatum* oligosaccharide fraction containing two repeating units and the core oligosaccharide region. Superimposition of the reference ^{1}H-NMR spectrum (black) and STD NMR spectrum (red) (1:20 ratio, T=298K).

A third isolated fraction contained four OPS repeating units, and thus was a dodecasaccharide. The STD NMR experiment on this fraction allowed to confirm the interaction of the O-polysaccharide chain isolated from *F. nucleatum* LPS with Siglec-7, and to map the interacting epitope (Figure 5.8).

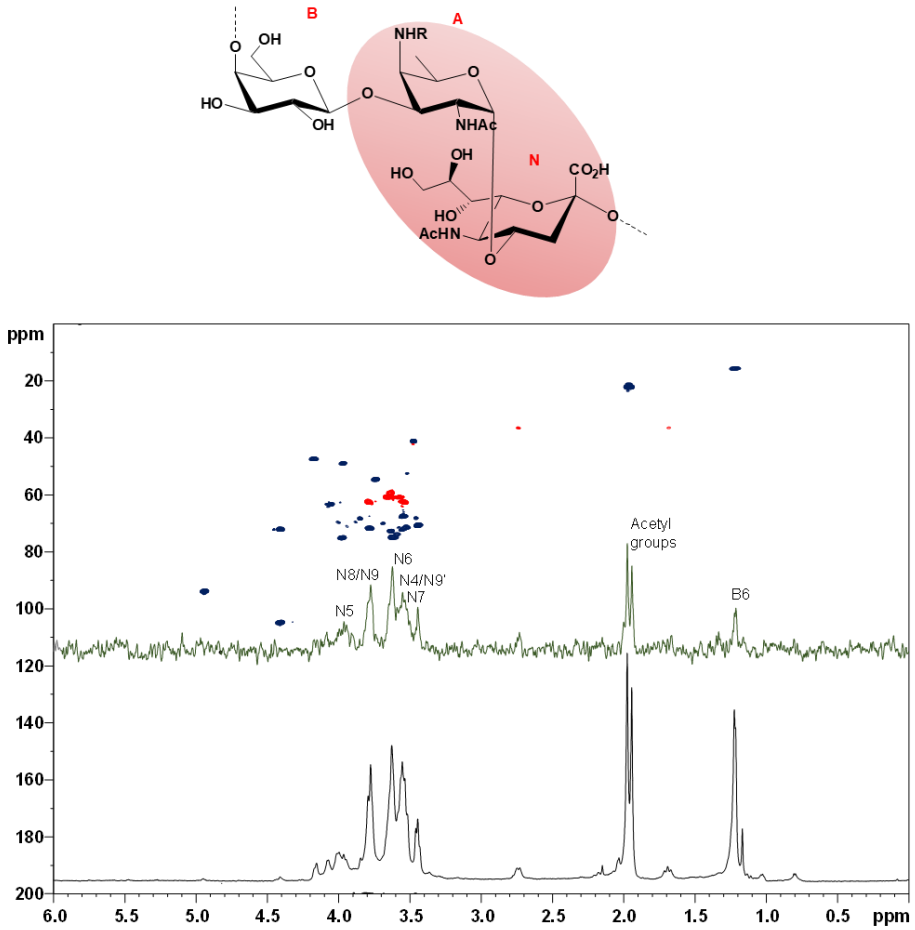


Figure 5.8 STD NMR analysis of *F. nucleatum* OPS recognition by Siglec-7. Superimposition of the reference ^{1}H-NMR spectrum (black) and STD 1D NMR spectrum (green) of Siglec-7 in the presence of partially depolymerized OPS from strain ATCC 10953. The ^{1}H-^{13}C HSQC spectrum (blue/red) and the chemical structure of the repeating unit of the OPS is reported. A preliminary exploration of the binding epitope allowed to highlight the sugar residues more involved in the interaction process (1:20 ratio).

In detail, from the STD spectrum it was assessed that the higher STD signal belonged to the N-acetyl group of Sia (**N**) unit. Significant STD signals belonged to the Sia protons of the glycerol chain. Interestingly, no significant STD

enhancements were detected for the protons of the Gal (**B**) unit, except for a weak STD signal of the H-2 protons. This suggested that the Gal unit was far from the Siglec-7 binding site. Concerning the FucNAcN unit (**A**), all the protons of the ring showed STD increases, in particular the highest signals derived from protons in position 3 as well as of the acetyl group.

Thereafter, to further characterize the interaction *F. nucleatum* LPS – Siglec-7, the dodecasaccharide fraction was used for titration experiments, following the quenching of Siglec-7 aromatic residues upon the addition of increasing concentrations of the dodecasaccharide fraction (Figure 5.9). The elaboration of the fluorescence data allowed for the estimation of the binding constant relative to the dodecasaccharide/Siglec-7 association.

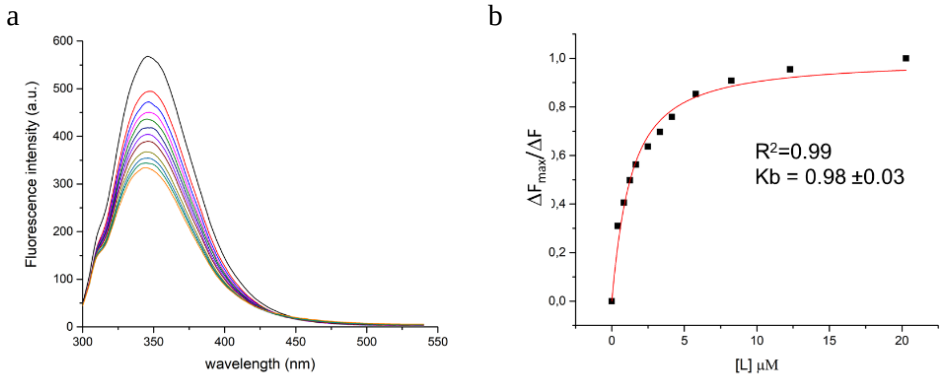
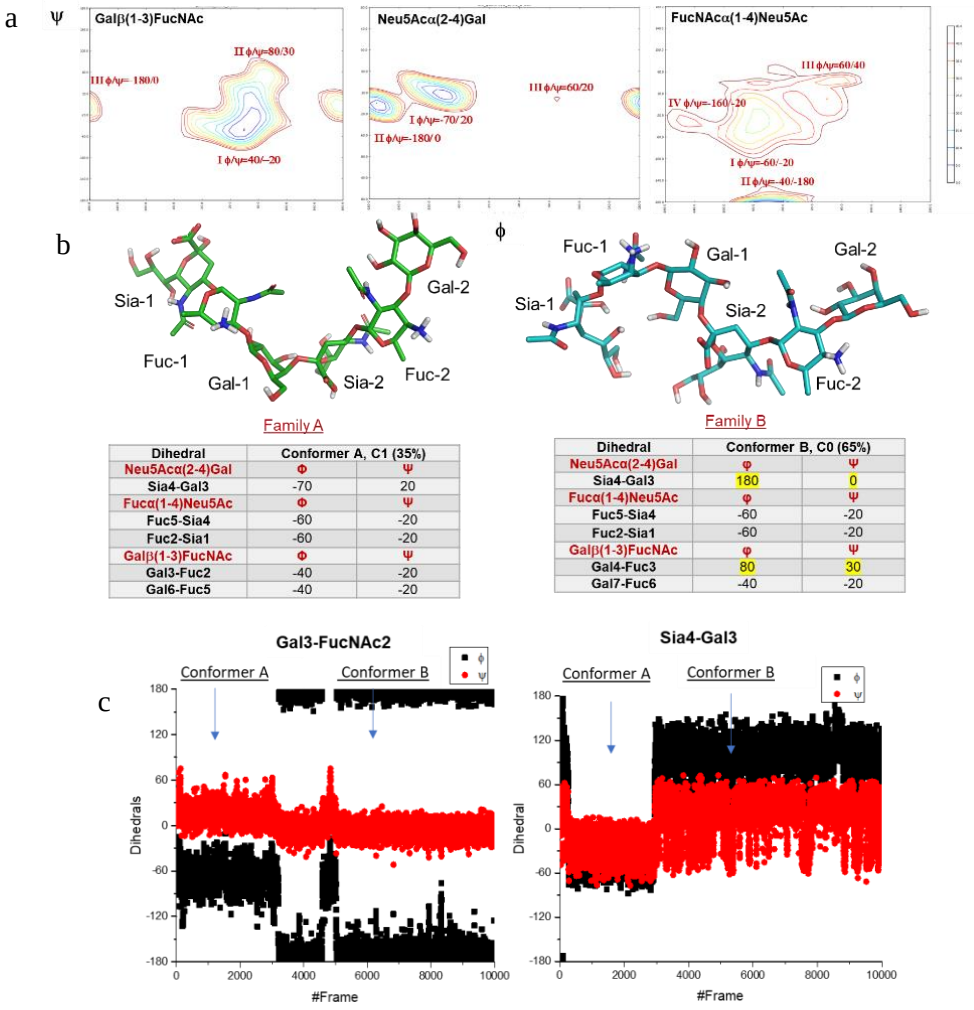


Figure 5.9 Binding affinity of Siglec-7 and OPS dodecasaccharide by steady state fluorescence. a) Fluorescence spectra overlay of Siglec-7 at increasing dodecasaccharide concentrations (from 1:1 to 1:100 ratio) **b)** Binding isotherm obtained by fitting the fluorescence data according to the equation described by Ribeiro et al.[289]

5.4 Molecular modelling of *F. nucleatum* O-chain and Siglec-7

5.4.1 *F. nucleatum* OPS hexasaccharide /Siglec-7 modelling

The investigation of the conformational features and the available conformational space for the glycosidic angles of the O-chain region was performed through MM and MD simulations. First, the three basic constituent disaccharides of the O-repeating unit (Siaα(2,4)Gal, Galβ(1,3)FucNAc4NAc, FucNAc4NAcα(1,4)Sia) were built and MM calculations were performed to analyze the energetically accessible conformational regions by means of adiabatic energy maps, using Maestro suite of programs. The obtained adiabatic energy maps for the glycosidic torsions φ (H1-C1-O-CX') and ψ (C1-O-CX'-HX') for Galβ(1,3)FucNAc and FucNAcα(1,4)Sia linkages, and φ (C1-C2-O-C4') and ψ (C2-O-C4'-H4'), for Siaα(2,4)Gal linkage are reported in Figure 5.10.



Dihedral	Conformer A, C1 (35%)	
Neu5Acα(2-4)Gal	Φ	Ψ
Sia4-Gal3	-70	20
Fucα(1-4)Neu5Ac	Φ	Ψ
Fuc5-Sia4	-60	-20
Fuc2-Sia1	-60	-20
Galβ(1-3)FucNAc	Φ	Ψ
Gal3-Fuc2	-40	-20
Gal6-Fuc5	-40	-20

Dihedral	Conformer B, C0 (65%)	
Neu5Acα(2-4)Gal	φ	Ψ
Sia4-Gal3	180	0
Fucα(1-4)Neu5Ac	φ	Ψ
Fuc5-Sia4	-60	-20
Fuc2-Sia1	-60	-20
Galβ(1-3)FucNAc	φ	Ψ
Gal4-Fuc3	80	30
Gal7-Fuc6	-40	-20

Figure 5.10 MM and MD study of *F. nucleatum* OPS hexasaccaride a) Adiabatic energy maps of the three disaccharides connected by a glycosidic linkage in the repeating units of O-chain from *F. nucleatum* ATTC 10953. **b)** Structure of the representative hexasaccharide conformations alongside the MD simulations, as obtained by cluster analysis using Kmean algorithm. [290]. The approximated average values of each glycosidic linkage dihedrals of the disaccharide in the two conformations are enlisted in the tables **c)** Evaluation of the ϕ/ψ dihedral fluctuations of the free hexasaccharide for the glycosidic linkages that shape the conformations A and B along the MD simulation.

Considering Galβ(1,3)FucNAcN and FucNAcNα(1,4)Sia units, the inspection of the maps permitted to assess that, for both disaccharides, the global minimum (indicated as I in the adiabatic maps in Figure 5.10a) value was in accordance with the *exo*-anomeric effect, as well as the second most populated one (indicated as II in the adiabatic maps in Figure 5.10a). In particular, concerning Galβ(1,3)FucNAcN the I minimum (φ = -40, ψ =-20), corresponded to *exo*-anti conformation and the II minimum (φ = 80, ψ =30), corresponded to *exo*-syn conformation.[291] Concerning the FucNAcNα(1-4)Sia disaccharide maps, the I minimum matched to the *exo* anomeric effect. As for Siaα(2,4)Gal linkage, two almost equally populated dihedral conformations, namely -g (I, φ = -60, ψ =-10) and t (φ = 180, ψ =-10) were found, (see Chapter 4).

Once the optimal values for φ and ψ dihedral angles had been estimated for each disaccharide, an hexasaccharide (corresponding to two OPS repeating units) was built and subjected to 100 ns MD simulations in explicit water with AMBER 18 (Figure 5.10c). It is worth to note that as the D-FucNAcN is not a standard sugar implemented in GLYCAM06-j forcefield, its parametrization was carried out as described in the experimental section (see Chapter 9, experimental section).

The MD simulation analysis in explicit water highlighted the presence of different populations for the hexasaccharide. In detail, two possible populated families, defined by a different set of dihedral angles were obtained from cluster analysis, designated A and B (Figure 5.10). The variation of the dihedral angles along each glycosidic linkage was monitored to characterize **A** and **B**, and it showed that the linkages that shaped the two conformational states were Gal3-

FucNAcN2 and Sia4-Gal3, the latter passed from the -g, favored in vacuum environment, to the t conformation in water.

In detail, the **A** family (corresponding to the set of dihedral energy minima most populated in vacuum) was stable for 30% (the initial 30 ns) of the simulation, then switched to conformation B, populated for 70% of time and that remained stable until the end of the simulation. **B** conformation was characterized by an extended topology, whereas the conformation **A** by a rather bent topology. The reason of the shift from one conformation to another could be ascribable to the greater exposure to the solvent, leading to more stabilizing interactions.

The hexasaccharide was docked into the crystal structure of Siglec-7 V-set domain (PDB ID: 2HRL). For the docking with AutoDock 4.2, it the set of mimima around each glycosidic linkage corresponding to the Family **B**, i. e. the most stable in water was considered as starting ligand geometry. As result, the internal sialic acid unit (Sia-2) was found to interact with the primary sialic acid binding site residues of Siglec-7, considering the pose belonging to the most populated and less energetic cluster (data not shown). The obtained complex was minimized using Maestro and then subjected to 100 ns MD simulation in explicit water, to evaluate the stability and the most relevant protein/ ligand interactions (Figure 5.10). As result, it was assessed that the complex was stable for the entire time of the simulation and that preferentially assumed the **B** conformation alongside the trajectory (Table 5.1, Figures 5.11 and 5.12).

Table 5.1 Siglec-7- hexasaccaride main hydrogen bonds fraction along the MD simulation. Fraction is referred to the number of frames in which the interaction occurs respect to the total simulation time (100 ns).

Acceptor	Donor	Fraction
SIA_2@COO-	ARG_124	0,9
SIA_2@O8	ASN_133	0,5
SIA_1@O5N	TRP_74	0,5
SIA_1@O8	ASN_133	0,4
LYS_131	SIA_2@N5	0,3
FUCNAc_2@O3	ASN_129	0,2
SIA_1@O9	LYS_135	0,5
ASN_129	FUCNAc_5@N4	0,3

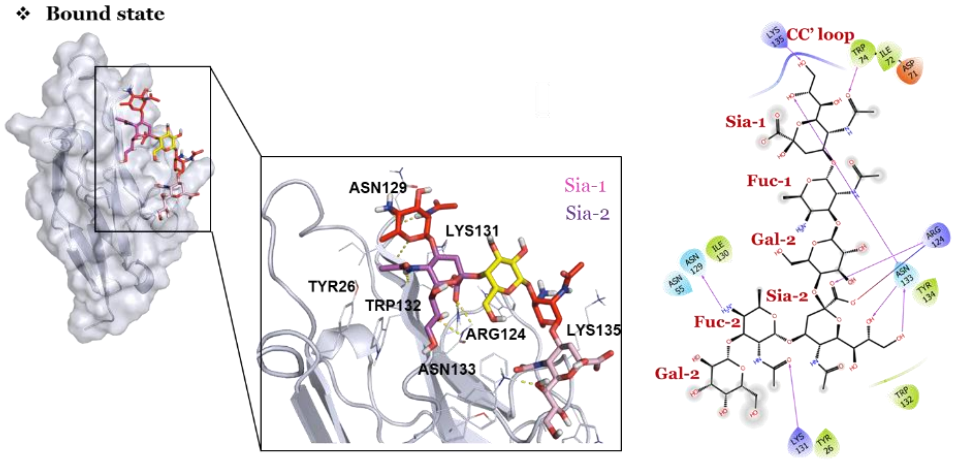
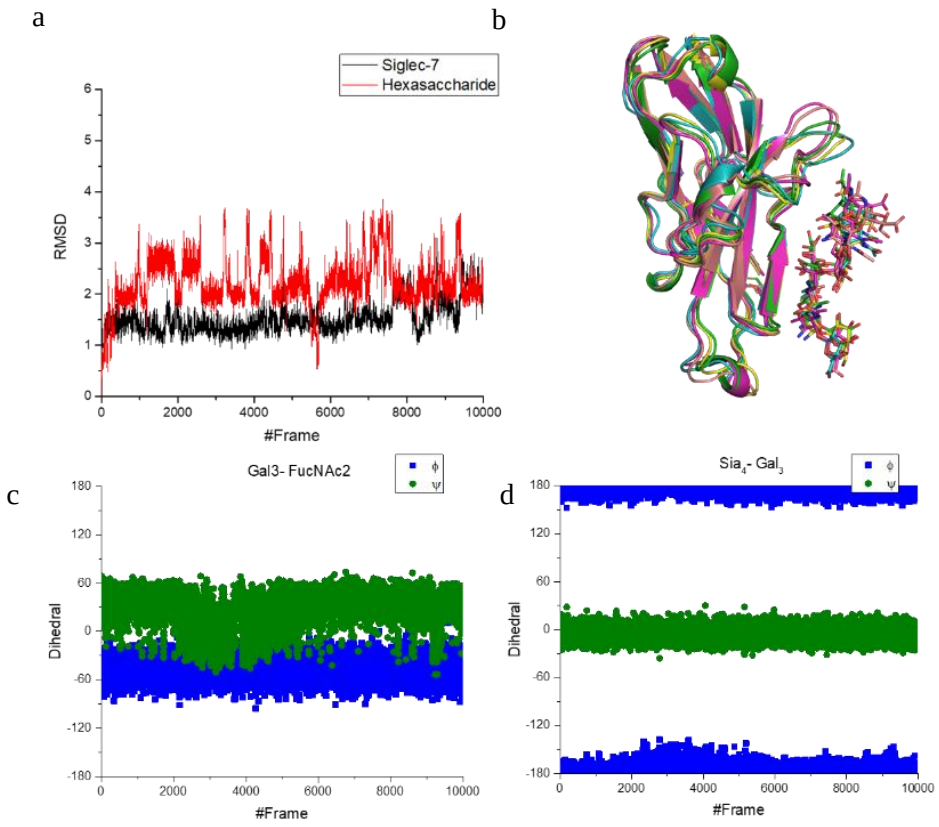


Figure 5.11 3D model of Siglec-7 V-set domain in complex with hexasaccaride from the *F. nucleatum* OPS a) Close up view of hexasaccharide binding mode at the Siglec-7 binding site. The main amino acid residues involved in the binding are represented in lines. The Sia, Gal and FucNAcN residues are depicted in magenta and yellow, respectively. **b)** Two-dimensional plot illustrating the interactions of the hexasaccaride OPS from *F. nucleatum* with the residues in the binding pocket of Siglec-7



Bound conformation

Dihedral	Conformer B, C0 (95%)	
Neu5Acα(2-4)Gal	Φ	Ψ
Sia5-Gal4	180	0
Fucα(1-4)Neu5Ac	Φ	Ψ
Fuc6-Sia5	-40	-20
Fuc3-Sia2	-40	-20
Galβ(1-3)FucNAc	Φ	Ψ
Gal4-Fuc3*	80	30
Gal7-Fuc6*	-40	-20

Figure 5.12 MD simulation analysis of Siglec-7/hexasaccharide complex a) RMSD analysis of the Siglec-7/hexasaccharide complex **b)** Superimposition of different snap-shots from the MD simulations of the hexasaccharide bound to Siglec-7. The non-reducing Sia and the CC' loop are evidenced. **c)** Hexasaccharide dihedral fluctuation along the main glycosidic linkages monitored during the MD simulation. The approximated average

values of each glycosidic linkage dihedrals of the disaccharide in the two conformations are listed in the table

Concerning the binding interactions (Figures 5.11 and Table 5.12), the Sia-2 unit remained anchored to the Siglec-7 primary binding site, through the salt bridge between Sia carboxylate and the Arg124, essential determinant of sialic acid binding. Furthermore, the lateral chain of Sia interacted via hydrogen bonds mainly with Asn133 on the GG' loop, furthermore the stacking of the ring protons with Trp132 further stabilized the interaction. The N-acetyl group of Sia formed a stable hydrogen bond with Lys126 residue. Overall, these interactions accounted for the high intensity of the STD signal belonging to Sia unit in the STD NMR experiment. Regarding the galactose units (Gal-1 and Gal-2), no relevant contracts were observed with the receptor as these residues were far off from its surface. Conversely, the FucNAcN residue adjacent to the Sia-4 unit (FucNAcN-2), was indeed involved in strong electrostatic interactions, specifically, a hydrogen bond between the sugar ammonium group at position 2 and Asn129 residue of Siglec-7 binding pocket. Additionally, the N-acetyl group of FucNAcN was involved in a hydrogen bond with Lys131 residue on Siglec-7 G strand. Likewise, Sia at the non-reducing end (Sia-1) of the hexasaccharide moiety was found to interact with the CC' loop region (Figure 5.11), in particular with Trp74 through a hydrogen bond between the amino acid and the carbonyl function of Sia N-acetyl group. Here, the acetyl apolar moiety formed hydrophobic interactions with Ile72. The lateral chain was also engaged in polar contacts with Lys135 residue. Interestingly, a conformational flexibility of the Siglec-7 CC' loop along the MD simulation was observed (Figure 5.10), for this reason the role of Siglec-7 loops in ligand binding was further explored by comparing the receptor loop flexibility in the free and bound states (Figure 5.13). To this end, MD simulation of unbound Siglec-7 was carried out and the relative alpha carbon root mean square fluctuation, to

assess the fluctuation of the backbone of the protein along the simulation (Cα-RMSF) and the backbone RMSD of the flexible loops were calculated (Figure 5.13). The Siglec-7 V-set domain showed good stability along the MD simulation, and the RMSF analysis showed larger fluctuations at the CC' and BC loop regions.

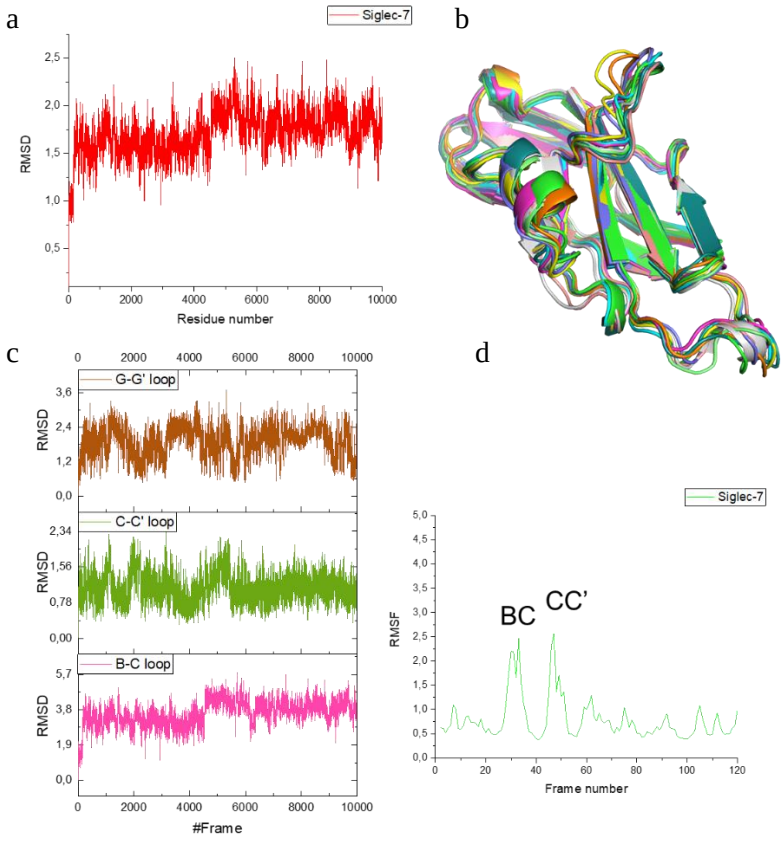


Figure 5.13 MD simulation of unbound Siglec-7. (PDB ID: 2HRL). **a)** Siglec-7 backbone RMSD. **b)** Superimposition of the top 5 clusters extracted from the MD simulation. **c)** By residue RMSD of Siglec-7. **d)** Backbone RMSD of CC', GG' loop and BC loops.

Such parameters were compared to those extracted from the bound state Siglec-7/hexasaccharide simulation, in order to assess possible conformational changes of the Siglec-7 loop regions upon oligosaccharide binding (Figure 5.14).

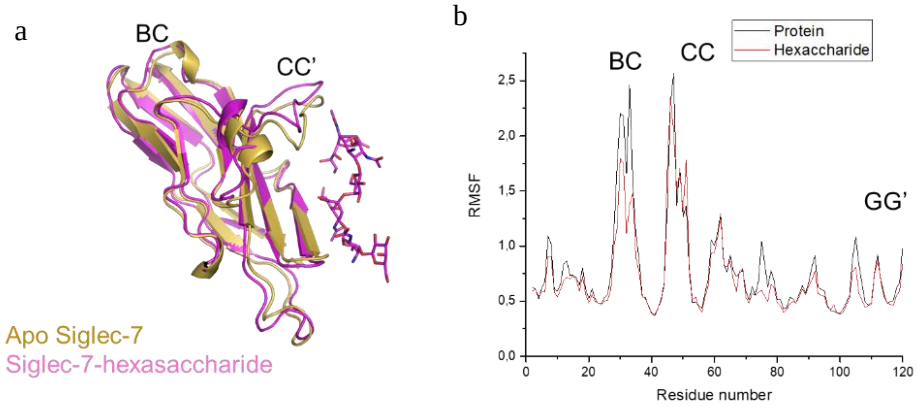


Figure 5.14 Comparison of the RMSF in the free and bound state Siglec-7. a) (dark yellow) apo- Siglec-7(purple) Siglec-7 bound to hexasaccharide model. **b**) (blue) RMSF of Siglec-7 (red) RMSF of Siglec-7/ hexasaccharide complex.

Interestingly, the overall stabilization of loops fluctuations upon binding was observed. A relevant decrease of the RMSD was noted at the region corresponding to the GG' loop region that was involved in the binding with the FucNAc-5 of the hexasaccharide. Furthermore, it was observed that the CC' loop conformation slightly opened with respect to the loop conformation of the apo protein to permit the accommodation of the long hexasaccharide moiety.

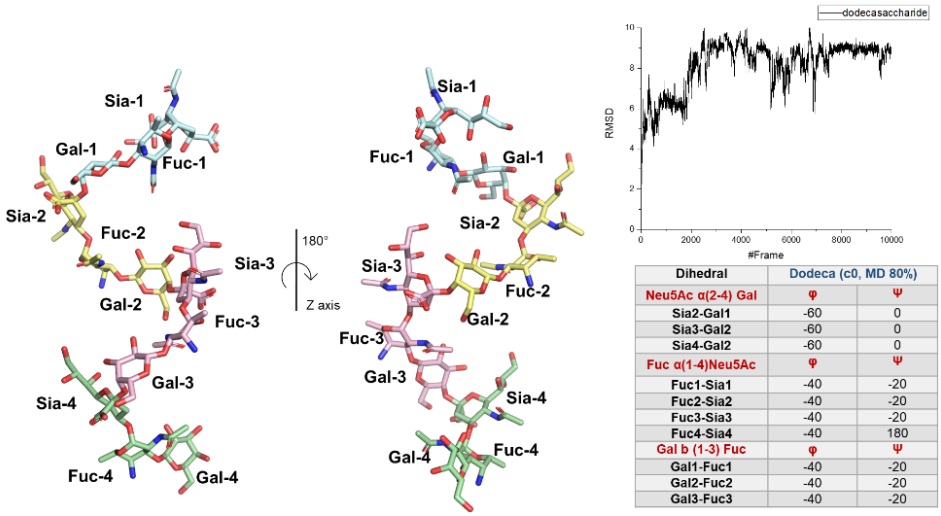
These results suggested a possible binding mode consistent with the STD NMR data, that indicated the involvement of the CC' loop in the *F. nucleatum* recognition by Siglec-7, thus further underlying the importance of this Siglec-7 region.

Additionally, docking and MD simulations studies were carried out also using conformer A, to observe if it allowed for a proper binding mode with Siglec-7. As result, a different stable conformation was observed upon binding. In this

complex, the fucose units were not involved in the binding as it was far from the receptor surface, whereas the Galactose (Gal-2) was interacting with the receptor, in complete disagreement with the experimental data and thus it was excluded as possible bioactive conformation (data not shown).

5.4.2 *F. nucleatum* OPS dodecasaccharide /Siglec-7 modelling

Next, a longer dodecasaccharide portion of the OPS (corresponding to four repeating units) from *F. nucleatum* from 10953 strain was considered, to further mimic the natural behavior of the pathogenic bacterium, and was investigated by *in silico* methods, in a similar fashion as the hexasaccharide. First, the dodesaccharide moiety was built using Amber Tools, using the same procedure carried out for the hexasaccharide (see chapter 9, experimental section). Again, when building the dodesaccharide, the energy minima from the adiabatic energy maps were considered and then the ligand was subjected to MD simulation in explicit water to investigate its conformational behavior in solution (Figure 5.15).



Dihedral	Dodeca (c0, MD 80%)	
Neu5Ac α(2-4) Gal	φ	Ψ
Sia2-Gal1	-60	0
Sia3-Gal2	-60	0
Sia4-Gal2	-60	0
Fuc α(1-4)Neu5Ac	φ	Ψ
Fuc1-Sia1	-40	-20
Fuc2-Sia2	-40	-20
Fuc3-Sia3	-40	-20
Fuc4-Sia4	-40	180
Gal b (1-3) Fuc	φ	Ψ
Gal1-Fuc1	-40	-20
Gal2-Fuc2	-40	-20
Gal3-Fuc3	-40	-20

Figure 5.15 MD simulation of the free dodecasaccharide from *F. nucleatum OPS*. The ligand RMSD along the MD simulation was also reported. The approximated average values of the dihedrals around each glycosidic linkage was reported for the most stable conformation

As result, the ligand changed conformation after 20 ns of MD simulation. From the dihedral fluctuations along the MD simulations, no variations of Siaα(2,4)Gal linkage, that remained stable around ϕ/ψ -60,0 for the entire simulation was noticed.

1Notably, the longer dodecasaccharide adopted a rather helicoidal structure. Interestingly, the carboxylate groups of Sia's, are exposed to the solvent and available to bind to a putative host immune receptor which recognize sialic acids, as Siglec-7. The helical structure of *F. nucleatum* OPS dodecasaccharide was used to carry out a manual docking within the Siglec-7 pocket to investigate the binding mode. To this end, Sia-2 was placed at the primary sialic acid binding site (at the Arg124) of Siglec-7 crystal structure, similarly to what observed for the docking

with the hexasaccharide. Thereafter, the resulting complex was subjected to minimization using Maestro suite and then subjected to MD simulations using AMBER 18 to assess its stability. The complex was shown to be stable along the MD simulation time, therefore the cluster analysis was performed and the hydrogen bond interactions between the protein and the ligand were monitored along the MD trajectory. (Figure 5.16, Table 5.2).

Table 5.2 Siglec-7- dodecasaccharide main hydrogen bonds fraction along the MD simulation. Fraction is referred to the number of frames in which the interaction occurs respect to the total simulation time (100 ns).

#Acceptor	Donor	Frac.	AvgDist.
Sia-2	Arg124	0,9813	2,7453
Sia-2	Asn133	0,533	2,8775
Lys131	Sia-2	0,2833	2,8901
Asp 59	Sia-4	0,883	2,678
Sia-4	Asn129	0,1556	2,9019
Fuc-2	Asn129	0,3239	2,8961
Asn129	Sia-2	0,1213	2,821

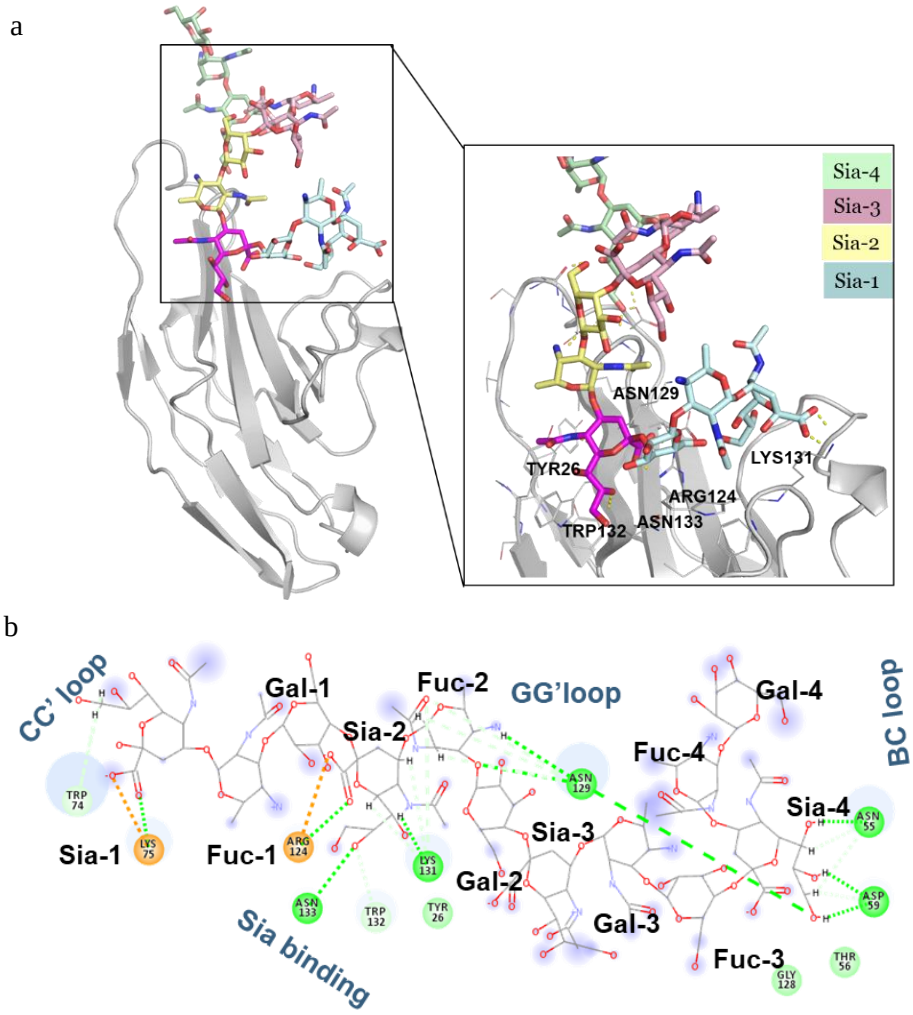


Figure 5.16 3D model of Siglec-7 V-set domain in complex with dodecasaccharide from the *F. nucleatum* OPS a) Close up view of dodesaccharide binding mode at the Siglec-7 binding site. The main amino acid residues involved in the binding are represented in lines. The Sia-6, binding at the primary sialic acid binding site is shown in magenta. **b)** Two-dimensional plot illustrating the interactions of the dodecasaccharide OPS from *F. nucleatum* with the residues in the binding pocket of Siglec-7.

From the analysis of the Siglec7-OPS interactions it was possible to assess that the Sia-2 residue of the dodecasaccharide was stable inside Siglec-7 where it formed steady contacts with the Arg124 (salt bridge), Asn133 on GG' loop (hydrogen bond), Lys131 (hydrogen bond) and Trp132 (stacking). Similarly to the hexasaccharide, Sia-1 was found to interact with the CC' loop Trp74 and Lys75. Moreover, FucNAcN-2 was again engaged in polar contacts with the Asn129. The binding interactions were further extended to Sia-4, that interacted through its lateral chain hydroxyls with Asn55 and Asp59. It is worth to note that the ligand maintained the helical shape upon binding.

As for the hexasaccharide/Siglec-7 complex, the Cα-RMSF was monitored to observe if the receptor binding site loops conformations were affected (namely CC', GG' and BC) by the dodecasaccharide interactions (Figure 5.17).

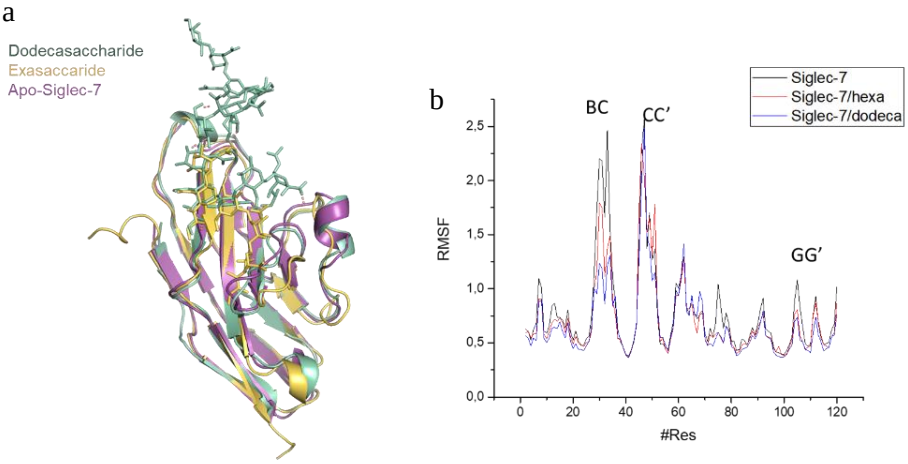


Figure 5.17 Analysis of Siglec-7 atomic fluctuations of free and bound states. a) Superimposition of the structures of Siglec-7 in the apo (pink), hexasaccharide bound (yellow), dodecasaccharide bound (green) **b)** Comparison of the RMSF in the free and bound state Siglec-7 (unbound= black line, hexasaccharide =red line, and dodecasaccharide= blue line).

Interestingly, from the RMSF comparison to the unbound Siglec-7, the presence of the long dodecasaccharide induced a greater stability of the fluctuations relative to the BC loop, and, to lesser extent, to the CC' and GG' loop regions, consistently with a strong and stable interaction network s formed between the receptor loops and the dodecasaccharide moiety from *F. nucleatum* OPS.

Last, it is worth to mention that MD simulations was performed starting from a complex in which Sia-3 was bound to the primary Siglec-7 binding site region, but the complex was not stable for the simulated time, further excluding this model of interaction (data not shown).

5.4 Discussion

Siglec-7 is well established as an inhibitory lectin-type receptor primarly located on NK cells where it acts as inhibitor of cancer cells cytotoxicity *via* sialylated ligands, such as GD3 or DSGb5.[256,257]

On the other hand, *F. nucleatum* emerged as an oncogenic pathogen in different human tissues, and multi-various pathways by which it can contribute to the development, growth, propagation of tumors were addressed.[261]

Recent studies highlighted the presence of sialylation of *F. nucletum* bacterial surface, and *de-novo* sialic acid biosynthetic pathways for this pathogen were described. Then, the isolation of sialylated LPS on certain *F. nucleatum* strains, as 10953, hinted that sialylated LPS may have a significant role at the immune interface. Indeed, sialylation of *F. nucleatum* may induce the activation of sialic acid binding immunoglobulin-like receptors, as Siglec-7, thus causes immuno-suppression that may promote its carcinogenic behavior.[274]However, the specific determinant of sialic acid in *F. nucleatum* LPS remains to be determined.

The recognition of *F. nucleatum* LPS with Siglec-7 receptor has been addressed by Prof. Nathalie Juge (in publication). Herein, the investigation of

interaction between the Fusobacterium sialylated LPS from 10953 strain and Siglec-7 was undertaken. To determine how *F. nucleatum* interacts with the sialic acid binding receptor, a combination of STD NMR, fluorescence assay and *in silico* docking and MD simulations was employed.

Firstly, the STD NMR experiments carried out using the O-antigen isolated from *F. nucleatum* 10953 strain, allowed to understand the residues of the OPS repeating unit directly involved in Siglec-7 recognition and binding. In particular, the strongest contribution arose from sialic acid unit, although a significant involvement was given by the FucNAcN sugar. Conversely the Gal unit showed a moderate contribution to the receptor binding. The association between O-antigen and Siglec-7 was further demonstrated by intrinsic fluorescence quenching experiments. Globally, these analyses also suggested that at least two O-antigen repeating units are needed to establish stable and longlasting interactions with the receptor.

Then, docking and MD simulations were carried out to define a 3D model of the O-antigen alone and in complex with Siglec-7 V-set domain. First, MM and MD calculations on the hexasaccharide permitted to identify a preferential ‘’extended’’ conformation in solution, that was maintained upon Siglec-7 binding, as demonstrated by docking and MD simulations. Interestingly, these analyses showed a preference for the binding of the internal sialic acid (Sia-4) residue, to the primary Sia binding side, composed by Arg124, Trp132, Lys131 amino acids. Furthermore, significant interactions were observed between FucNAcN-5 polar substituents mainly with Asn129 residue. Furthermore, the terminal Sia-1, interacted with Tyr74 and Lys75 on the CC’ loop.

Moreover, the construction of a longer saccharide epitope of *F. nucleatum* O-antigen from 10953 strain (dodecasaccharide) and the relative MD studies allowed for the identification of the helical topology, in the free and the bound state.

Interestingly, this helical conformation permitted to expose to the solvent the Sia units carboxylates, thus available for the interaction with Siglec-7, as well as other sialic acid binding receptors. A slight conformational change in the bound state allowed to form a less compact helical structure that accounted for an extensive interaction with the Siglec-7 shallow pocket, that involved Sia-6 in the primary sialic acid binding site and Sia-1 at the CC' loop, with a similar interaction network with respect to the hexasaccharide.

Surprisingly, it was shown that up to three Sia units can interact with the binding region of Siglec-7, that involved the B-C loop that interacted with Sia-9 by means of Asn55 and Asp59 residues. Interestingly, the RMSF, analysis carried out on the free and the bound state of both 3D complexes (with hexasaccharide and dodecasaccaride) allowed to further underlie the importance of the loops' flexibility, namely the CC' loop and the BC loop likewise that allows the accommodation of longer saccharide structures. Currently, 2D NOESY analysis is ongoing to compare the conformational data gathered by MD simulations.

In conclusion, these studies undertook novel structural insights on the recognition and binding of a sialylated LPS of *F. nucleatum* in the interplay with Siglec-7, both important actors in tumorigenesis. The obtained information will be valuable in the development of therapies effective against the opportunistic pathogen diseases, as well as for the modulation of Siglec-7 activity.

SECTION III-OTHER PROJECTS

Chapter 6- Human Macrophage Galactose-Type Lectin recognizes the Outer Core of *E. coli* lipooligosaccharide

6.1 Introduction

6.1.1 Bacterial LPS

As stated in the first chapter, bacterial cell surfaces are decorated with highly diverse glycoconjugates, in the form of capsular polysaccharides, peptidoglycans, lipopolysaccharides and other glycolipids,[292] which perform several functions ranging from structural to protective roles.[293] Bacterial glycans take part in many essential biological processes including pathogen recognition, receptor activation, cell adhesion and signal transduction. Besides, these structures often function as molecular patterns which are recognized by specific glycan-binding receptors of host immune system, in turn triggering a pathogen-specific immune response. It is well acknowledged that LPSs, with are the most prominent constituents of the outer membrane of Gram-negative bacteria,[294] are one of the principal virulence factors of a number of bacterial strains, including the enteropathogenic *E. coli*, that is associated to severe food and urinary tract infections.[295] Concerning the structural composition, LPS is constituted of three distinct motifs that are encoded by different gene clusters (Figure 6.1).

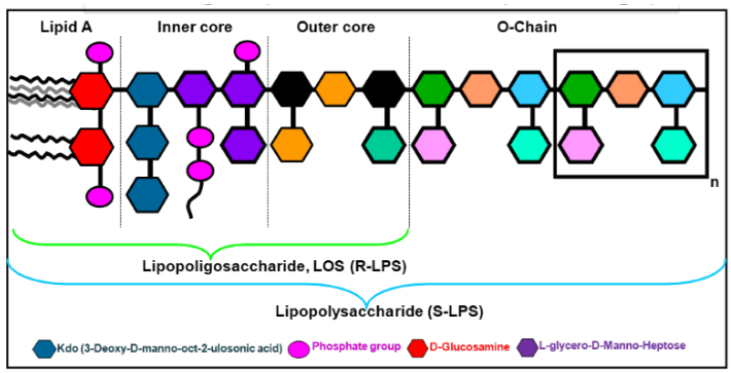


Figure 6.1 Schematic representation of the general chemical structure of bacterial lipopolysaccharide: Lipid A, internal oligosaccharide and specific O-chain. Rough type LPS (R-LPS) do not contain specific O-chain, and smooth type (S-LPS) contain two or more repetitive units of specific O-chain.

First, a glycolipid portion called Lipid A, consists of a phosphorylated glucosamine disaccharide moiety acetylated with fatty acid chains, and that anchors the LPS molecule to the outer membrane. The lipid A is covalently attached to a core oligosaccharide, that can be further distinguished into two different moieties: the more conserved inner region, characterized by the presence of peculiar sugar residues, such as 3-deoxy-D-manno-oct-2-ulopyranosonic acid (Kdo) and heptoses, and the more variable outer core. Last, the O-antigen is a polysaccharide composed of up to 40 oligosaccharide repeating units, which extends to the extracellular medium and serve as a hydrophilic coating surface.[296] On the basis of genome organisation and mutations, Gram-negative bacteria can also produce exclusively rough-type LPS (Figure 6.1), namely lipooligosaccharide (LOS), lacking the O-antigen.

To date, bacterial LPS recognition by host immune cells has been principally attributed to the Toll like receptor TLR4 in complex with MD-2,[297] however, given their ability to selectively recognize particular carbohydrate structures, human lectins have emerged as effective LPS receptors.[298]

Here, the LOS of *E. coli* R1 (Figure 6.2).[299] was extracted, purified and characterized by NMR spectroscopy and the LOS recognition MGL receptor has been investigated.

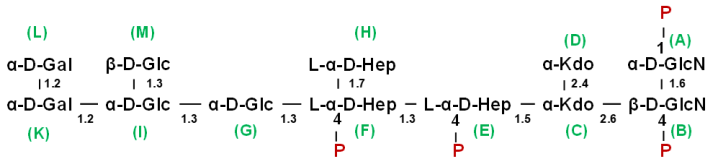


Figure 6.2 Structural assessment for the core OS from *E. coli* R1 (OS_{R1}).

6.1.2 MGL- a major member of C- type lectins

Recently, it has been underlined a possible role of MGL in bacterial infections as well.[306] As a remarkable case, it has been found that MGL binds to α–GlcNAc residues found on wall teichoic acid chains on the cell wall of *Staphylococcus aureus*.[307] From a structural viewpoint, MGL is a type II transmembrane protein composed of a cytoplasmic domain containing a YxxΦ endocytosis motif (where Y is a tyrosine, X represents any amino acid, and Φ is a hydrophobic aminoacid), followed by a transmembrane domain, and then by an extracellular domain (ECD). The ECD possess a coiled-coil neck that serve to stabilize the homotrimeric organization. At the ECD it is also located the C-terminal Calcium-dependent (C-type) CRD (Figure 6.3).

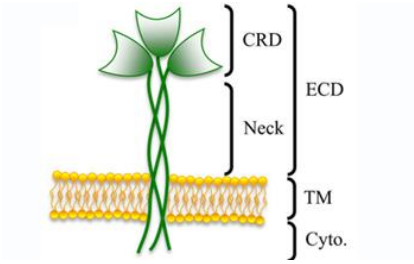


Figure 6.3 Domain organization of MGL. TM: transmembrane domain. Cyto.: cytoplasmic domain.

In contrast to several other C-type lectins, including DC-SIGN and the macrophage mannose receptor (MR), which exhibit a preference for mannose type sugars, human MGL is characterized by an exquisite specificity for terminal GalNAc/Gal residues, making it a suitable target for Gal containing LOS structure of *E. coli* R1.

In this regard, the recognition of exogenous antigens by MGL and thus the role in host defense was here explored, by here probing the ability of MGL to recognize bacterial LPS though the application of STD NMR analysis combined with computational studies. This work has been carried out in collaboration with Prof. Cedric Laguri, Prof. Franck Fieschi and Prof. Jean Pierre Simorre (University of Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale, Grenoble, France).

6.2 Molecular recognition of MGL

First, cloning and recombinant production of MGL ECD were carried out (Prof. Franck Fieschi group) using the sequence of the isoform-2 of human MGL which is shorter than the MGL isoform-1 (also called DC-ASGPR). The MGL extracellular domain, MGL-ECD (Figure 64a) has been cloned and overexpressed at high level in *E. coli* as inclusion bodies (Figure 6.4c).

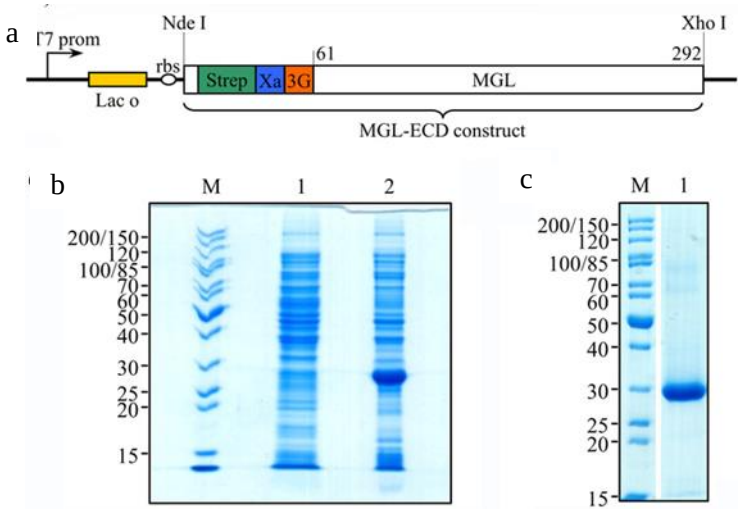


Figure 6.4 Cloning and purification of MGL-ECD. **a)** Construct for the overproduction of MGL-ECD. Strep: StrepTag II. Xa: factor Xa protease cleavage site. 3G: tri-glycine linker. **b)** SDS-PAGE analysis of MGL-ECD overexpression. Lane M: PageRuler unstained protein ladder (Fermentas); lane 1: total proteins before induction; lane 2: total proteins after induction. A band at about 28 kDa corresponding to MGL-ECD MW is overexpressed. **c)** SDS-PAGE analysis of purified MGL-ECD. Lane 1: A unique band at about 28 kDa corresponding to pure MGL-ECD is observed

Inclusion bodies were isolated and refolded as described in the method section. Folded functional MGL have a trimeric oligomeric state that allowed tight binding, through an avidity-based mechanism, on a GalNAc-Agarose column. Elution was performed by using EDTA, indicating that the binding was Ca^{2+} dependent.

The LOS of *E. coli* R1 was extracted, purified and de-acylated [3](see experimental section) to obtain a soluble oligosaccharide which was subsequently characterized by NMR spectroscopy. After the assignment of the NMR

[3] Performed by PhD Meriem Maalej

resonances (Figure 6.5), the ability of the MGL to recognize and bind the core oligosaccharide was assessed by STD NMR.[308]

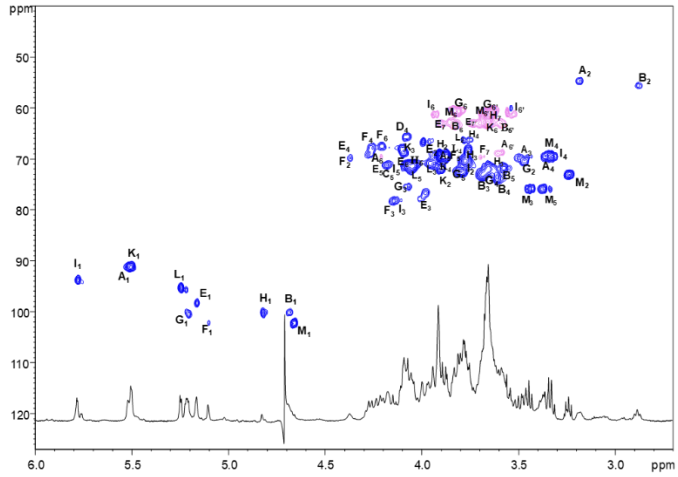


Figure 6.5 HSQC experiment. Zoom of the ^{13}C, ^{1}H HSQC spectrum of OS_{R1}. The most relevant heteronuclear correlations are reported.

The presence of STD enhancements in the STD spectrum of the mixture MGL: OS_{R1} evidently confirmed the protein-ligand interaction (Figure 6.6).

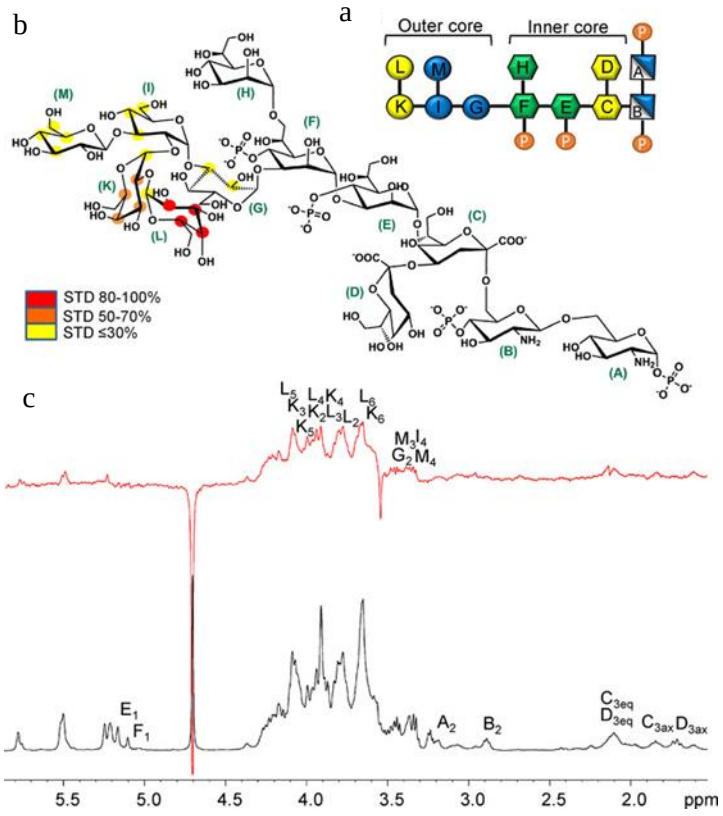


Figure 6.6 STD NMR spectroscopy analysis of the MGL–OSR1 mixture. a) Schematic structure of OSR1 derived from the LOS of E. coli R1. Symbol nomenclature for glycans has been used. Phosphate groups are depicted as white letter P in orange circles. **b)** STD-derived epitope mapping of the MGL: OSR1 interaction, with color coding from the highest (red) to lowest (yellow) observed STD effects. **c)** 1H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:100 mixture of MGL–OSR1. Some key proton resonances are labelled. STD NMR spectroscopy analysis indicated that the inner core pointed farther from the surface of the MGL protein, and it was mainly involved in the interaction with the terminal region of the OS.

The huge overlapping between different NMR resonances hampered an accurate quantitative study of the STD effects, however, the analysis of signals in isolated regions of the spectrum, together with the noticeable differences in the

multiplicity and relative intensities of STD signals in comparison to the corresponding reference (off resonance), implied that the outer core was more engaged in the binding with the lectin. Notably, the strongest STD signals belonged to the terminal disaccharide of galactose and in detail to the protons H3, H4 and H5 of the residues **L** and **K**. This observation agreed with the previously reported ability of the human MGL to accommodate in its binding site not only the N-acetylgalactosamine but also the galactose moiety.[309] Moreover, saturation transfer to the glucose residues, **I** and **G**, adjacent to the second galactose, was also observed although to a lesser amount. Additionally, some slight STD signals were detected for the protons H3, H4 and H5 of the residue **M**, suggesting it was somehow involved in the interaction with MGL. No significant STD effects were instead observed for protons at position 3 of Kdo residues, neither for protons at position 2 of the two glucosamine residues of the lipid A, suggesting that the inner core pointed farer from the surface of the protein that mainly interacted with the terminal region of the OS.

The interaction of MGL with LOS_{R1} was further illustrated by using molecular modelling and computational studies to get a three-dimensional model of the molecular recognition mechanism explaining the experimental STD data. Since the crystal structure of the protein was not elucidated previously to the publication of the present work, human MGL sequence was subjected to comparative homology modelling (Figure 6.7). The 3D model of the MGL CRD (Cys181-Leu308) was built by using as template the protein structure characterized by the highest sequence identity (73%), namely the CRD of the mammalian asialoglycoprotein receptor (PDB ID: 1DV8).

The obtained model retained the overall fold of the template structure, consisting in a CRD composed by core region of β– strands and by two α-helices

located on the core sides. The structure displayed three calcium ions coordinated by the protein residues, in three different calcium binding sites. The MGL CRD encompassed the characteristic glutamine-prolineaspartic acid (QPD) motif, located within a long loop region as well as the tryptophan-asparagine-aspartic acid (WND) motif that coordinates Ca^{2+} ion and it is known to be implicated in substrate recognition.[310] Precisely, the sugar-binding site was centred around the second calcium ion was formed by Gln243, Asp245, Glu256, Asn268, Asp269.

a

```
NP_006335.2*    MTRTYENFQYLENKV-KVQGFKNGPLPLQSLLQRLCSGPCHLLLSLGLGLLLLVIICVVG    59
BAB83508.1**    MTKEYQDLQHLDNEESDHHQLRKGPPPPQPLLQRLCSGPRLLLLSLGLSLLLLVVVCVIG    60
                **: *:::*:*:*:  . : :::** * * *********  *******.*****::**:*

NP_006335.2*    FQNSKFQRDLVTLRTDFSNFTSNTVAEIQALTSQGSSLEETIASLKAEVEGFKQERQAVH    119
BAB83508.1**    SQNSQLQEELRGLRETFSNFTASTEAQVKGLSTQGGNVGRKMKSLESQLEKQQKDLSEDH    120
                 ***::*.:*  **  *****:.* *:::.*::**..: ..: **::::*  ::: .  *

NP_006335.2*    SEMLLRVQQLVQDLKKLTCQVATLNNNGEEASTEGTCCPVNWVEHQDSCYWFSHSGMSWA    179
BAB83508.1**    SSLLLHVKQFVSDLRSLSCQMAALQGNGS----ERTCCPVNWVEHERSCYWFSRSGKAWA    176
                *.:**:*:*:*.**:.*:**:*:*:.**.    * **********: ******:** :**

NP_006335.2*    EAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYTWMGLSDPEGAWKWVDGTDYATGFQNW    239
BAB83508.1**    DADNYCRLEDAHLVVVTSWEEQKFVQHHIGPVNTWMGLHDQNGPWKWVDGTDYETGFKNW    236
                :*::**:*::*****:.* ***:***:::* . ***** * :* ********* ***:**

NP_006335.2*    KPGQPDDWQGHGLGGGEDCAHFHPDGRWNDDVCQRPYHWVCEAGLGQTSQESH--    292
BAB83508.1**    RPEQPDDWYGHGLGGGEDCAHFTDDGRWNDDVCQRPYRWVCETELDKASQEPPLL    291
                :* ***** ************  **************:****: *.::***
```

* C-Type lectin domain family 10 member A,isoform 2

** Asialoglycoprotein receptor 1

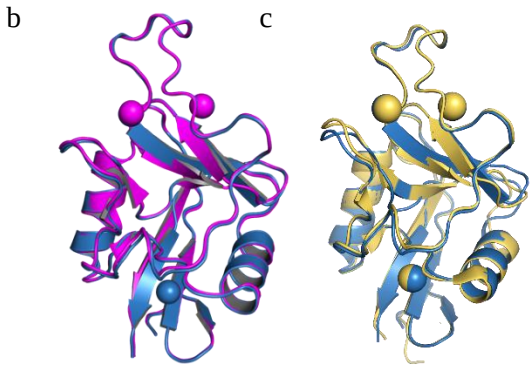


Figure 6.7 Homology modelling of MGL CRD.a) Multiple alignment of human MGL with the homologous protein ASGR. Alignment was performed with the CLUSTAL Omega program[311] and minimally adjusted manually. The sequence intervals of the CRD domain of human-MGL (Cys157-Leu284) and of the template used for the homology modelling, namely ASGR (Cys154-Leu181) are highlighted in grey. The residues important for galactoside binding are highlighted in cyan. **b)** Superimposition of MGL-CRD homology-based model (in blue) and the X-ray crystal structure from ASGR (in magenta) **c)** Superimposition of MGL-CRD homology-based model (in blue) and the recently solved X-ray crystal structure from MGL (in yellow), PDB ID: 6PY1.

These residues were also involved in the coordination of the calcium ion. The top edge of the binding was constituted by a curved loop that contained the conserved Trp247 residue, which represented a feature shared by most the Gal-specific proteins characterized so far.[19]

It is worth to mention that, shortly after the publication of the present work, the crystal structure of MGL was solved by Gabba et al.[312] Notably, the superimposition of the model with the crystal structure revealed a strongly similar fold, with an RMSD between the two structures of 0.429Å.

Then, docking calculations of OS_{R1} within the modeled structure were performed. As a first step toward the comprehension of the binding mode of the macrophage galactose-type lectin in the interaction with the bacterial OS, the

terminal disaccharide Galα(1,2)Gal was docked into the primary binding site of the MGL, allowing us to assess the binding specificity of the protein for the galactose residue. Two main different docking poses, with good theoretical binding energies and differing for a rotation of 180°, were predicted (named mode A and mode B, respectively) (Table 6.1 and Figure 6.8), as already observed by Marcelo et al. on the interaction of MGL with a single galactose unit.[309]

Table 6.1 Cluster rank, Cluster population, Computed binding energy and RMSD (Root Mean Square Deviation) for the molecular docking (AutoDock) of Galα(1,2)Gal disaccharide in the binding pocket of MGL. The values are referred to the lowest energy docked conformation from the highest ranked clusters 1 and 2, corresponding to mode A and mode B respectively.

Docking pose	Cluster Rank	Number of conformations in this cluster	Estimated Free Energy of Binding kcal/mol	RMSD from reference structure
Cluster 1 (Mode A)	1	97	- 1.23	15.934 Å
Cluster 2 (Mode B)	2	56	-1.08	13.401 Å

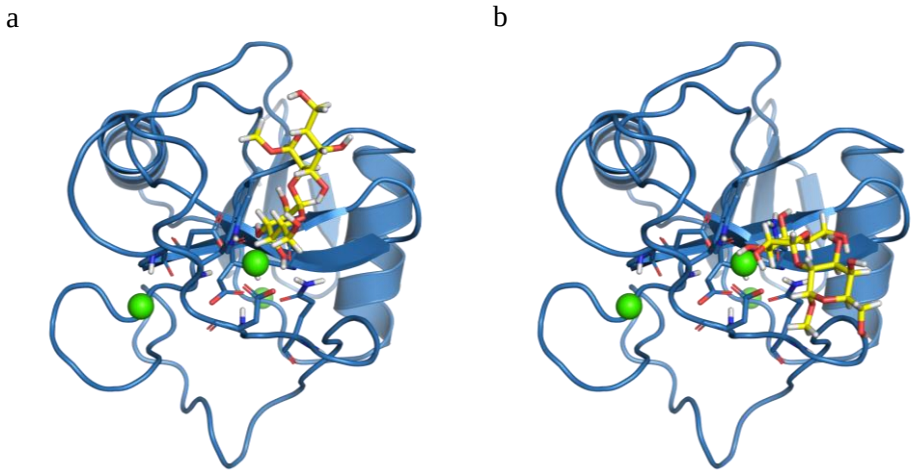


Figure 6.8 3D models of MGL - Gal-α-(1,2)-Gal complex. Complexes of MGL, obtained from docking calculations, showing the two main binding modes (**a** and **b**) of the galactose disaccharide when bound to MGL CRD. The two complexes mainly differ for the orientation of the terminal galactose moiety in the binding site of the protein.

In both cases, the terminal residue of galactose was coordinated by the calcium ion through the hydroxyl groups in position three (OH-3) and four (OH-4). OH-3 and OH-4 also established hydrogen bonds with Asp245, Glu256, Asn268 and Asp269. In full agreement with several X-ray structures of other C-type lectins bound to galactoside derivatives[313,314,315], depending on the orientation of the ligand in the binding pocket, a different face of the galactoside ring displayed hydrophobic interactions with the residue Trp247, engaging in CH-π interactions protons H3, H4, H5, and $H6_{a,b}$ in one binding mode and protons H1 and H2 in the other (Figure 6.9).

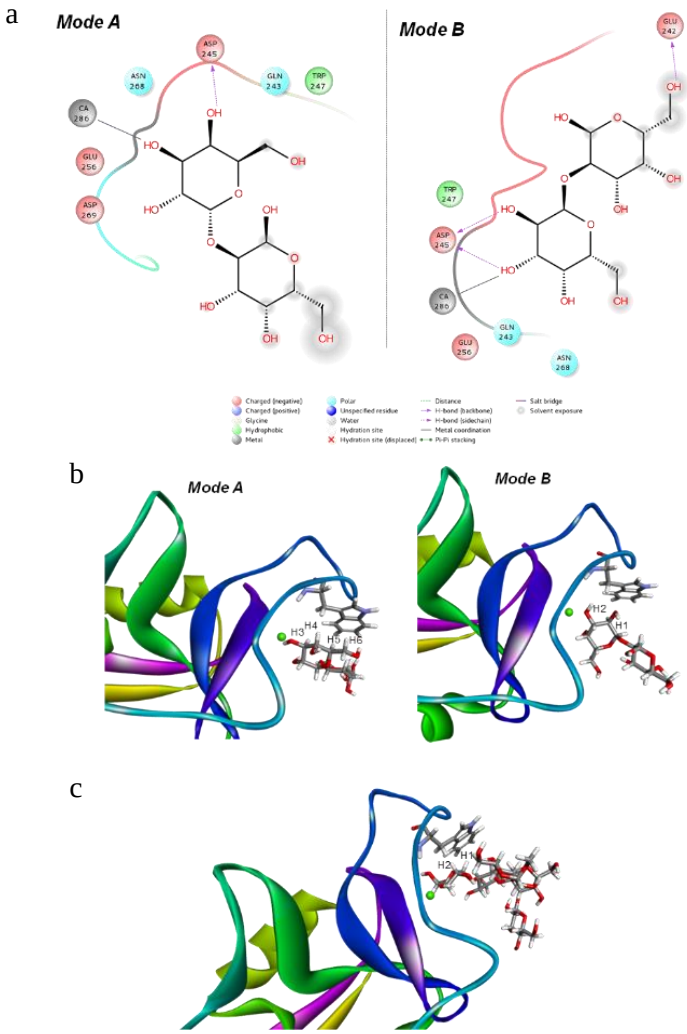


Figure 6.9 Docking studies of Galα(1,2)Gal disaccharide binding to MGL. a) Galα(1,2)Gal disaccharide binding at the MGL binding site. The main protein: ligand interactions of the two binding modes, A (on the left), B (on the right) are shown. **b)** Expansion of the protein binding site showing the stacking interaction between the residue Trp247 and the galactoside ring of Galα(1,2)Gal, according to the binding mode A (left) and B (right).**c)** Expansion of the protein binding site showing the stacking interaction between the tryptophan, Trp247, and the terminal galactose unit of the pentasaccharide.

Furthermore, the two binding modes also differed for the contacts formed between the protein and the second galactose unit, due to the different orientation of the sugar residue in the binding site of the protein, in mode a and b, respectively. In detail, besides the overall weak contacts for both binding modes, stacking interactions were observed between the H-3 proton of the second galactose and Trp247 only in the binding mode a. On the contrary, in mode b, a hydrogen bond between the OH in position 6 and Glu242 was observed.

Considering that from the STD NMR analysis the outer core of LOS resulted to be the moiety in closest proximity to the MGL binding surface, the galactose- containing branched pentasaccharide, composing the terminal part of the LOS, was then used for docking calculations (Figure 6.10).

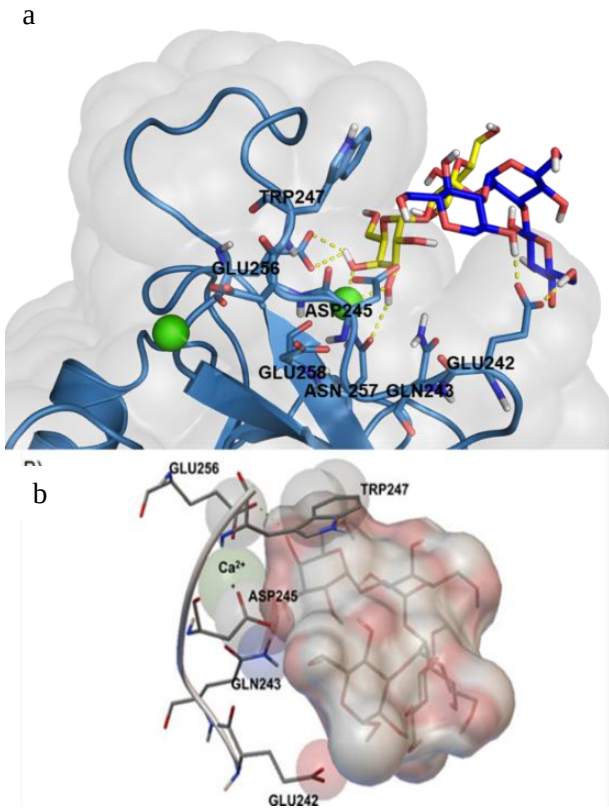


Figure 6.10 Docking studies of pentasaccharide binding to MGL. a) 3D model of the MGL–pentasaccharide complex. A) Pentasaccharide binding at the MGL binding site. The main amino acid residues involved in binding are labelled. Galactose and glucose residues are depicted in yellow and blue, respectively. **b)** The binding pocket of MGL in the presence of the pentasaccharide in the bound form, as derived from the AutoDock program. Only the amino acid residues involved in the binding process and the Ca^{2+} ion are depicted. Green dotted lines represent intermolecular hydrogen bonds.

As shown in the Figure of the best docked pose, the 3D structure of the complex showed that the terminal Gal moiety was placed into the Ca^{2+} containing sugar binding site of the MGL model. Crucial contacts between the hydroxyl groups at positions 3 and 4 of the galactoside ring and the Ca^{2+} ion were indeed observed and confirmed the specificity of the receptor for galactoside residues.

From the analysis of the receptor/ligand complex contact network, significant polar interactions were found between the terminal Gal hydroxyl groups at position 3 and 4 and the side-chains of Asp96, Asn119, Asn120 and Glu107 residues (Figure 6.10). In addition, hydrogen bond interactions with the carboxylate group of Glu93 were present, namely with OH-2 of the Glc residue (**G**) and OH-6 of the branched Glc moiety (**M**) respectively. Thus, the structural model further demonstrated that the terminal Gal of R1 outer core was directly involved in the molecular recognition process. Interestingly, in contrast with the results obtained on the disaccharide Galα(1,2)Gal, only one viable binding mode has been predicted for the pentasaccharide docked in the protein binding pocket. The selection of this binding mode, in which the polar contacts were markedly reinforced by the stacking interactions between the aromatic system of the conserved W98 and the protons H1 and H2 of the galactose moiety, could be ascribable to the steric hindrance of the pentasaccharide branched structure that favoured only one orientation of the galactose residue.

Last, the described MGL- pentasaccharide complex was used as starting point to manually dock the whole OS_{R1} structure onto the CRD of MGL, to depict the overall binding (Figure 6.11).

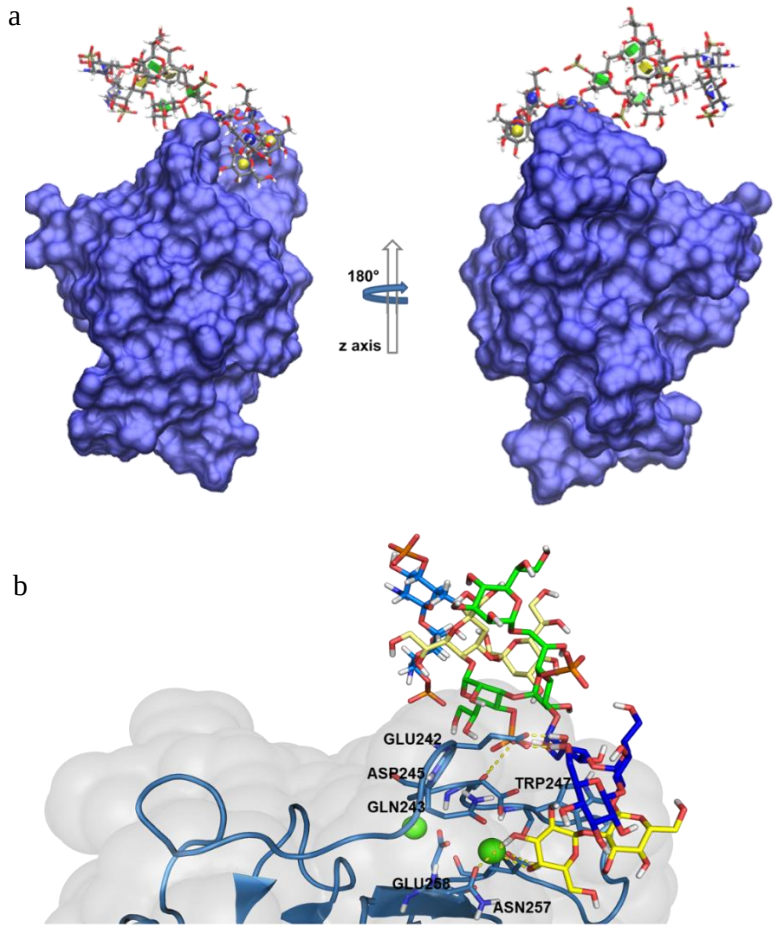


Figure 6.11 3D model of the MGL–OSR1 complex. a) Binding pose of the OS from E. *coli* R1, OSR1, docked into the binding pocket of the MGL starting from the 3D model of the MGL–pentasaccharide complex and based on NMR spectroscopy data. The 3D symbol nomenclature for glycans has been used. **b)** Specific interactions between binding-site residues and the OS are depicted. Sugar residues are colored according to the symbol nomenclature for glycans. Yellow dotted lines represent polar contacts.

From the resulting 3D complex, it was evident that the inner core of the LOS was not implicated in the interaction with MGL as it was directed far away from the binding pocket of the receptor model.
Overall, these observations were in good accordance with the previously discussed STD data, further confirming that significant interactions were established only between the protein and some residues of the outer core region, whereas the inner core did not play any role in the direct interaction with the MGL receptor.

6.4 Discussion

It was gathered evidence that the human C-type lectin MGL can detect not only tumour-related carbohydrate structures present in mucin-like glycopeptides, but also glycan structures displayed on the bacterial surface. Whereas a wide range of infective microorganisms were identified as pathogenic ligands of CTLs, such as DC-SIGN and MR,[316,317]glycan microarray profiling on MGL revealed its interaction merely with filoviruses and the helminth parasite *Schistosoma mansoni*.[318] Moreover, just a couple of papers have been published so far attempting to prove the potential role of the MGL in detecting LPSs structures isolated from *Bordetella pertussis* and *C. jejuni*.[319,320]

In this study, molecular insights into the structure of the MGL - LOS complex has been reported. It was demonstrated, by means of STD NMR, that the macrophage galactose-type lectin was able to recognize and bind the OS derived by the LOS of *E. coli* R1, whose major glycoform exhibits terminal Galα(1,2)Gal epitope, underlying the key role of the outer core in the binding process. Further details into the binding mode of MGL to LOS have been deduced by homology modelling and docking calculations. These results expanded the knowledge on the interaction between human CTLs and bacterial glycans and may aid to better

understand the role of MGL in affecting anti-bacterial immunity. MGL has been already reported as a good candidate receptor for DC-based cancer immunotherapy given its well-known ability to endocytose specific tumours-related antigens.[321] The discovery of the ability of MGL to detect endotoxin component of the outer membrane of the Gram-negative bacterium *E. coli* may provide a great potential for the future development of therapeutics for bacterial disease intervention.

As perspective, further studies are needed to understand if and how structural variations between different LOSs could affect recognition and binding process, thus tailoring the effect of the host-pathogen interactions.

Chapter 7-Structural basis for Glycan-receptor binding by mumps virus hemagglutinin-neuraminidase

7.1 Introduction

7.1.1 Mumps Virus

Mumps virus (MuV) is a human pathogen which belongs to the genus *Orthorubulavirus* of *Paramyxoviridae* family that is transmitted by respiratory-droplets. MuV is essentially constituted by enveloped, non-segmented, negative-strand RNA viruses[322,323.] Paramyxoviruses likewise comprise murine respirovirus (sendai virus), measles morbillivirus, avian orthoavulavirus 1 (newcastle disease virus), mammalian orthorubulavirus 5 (parainfluenza virus 5), and nipah henipavirus.

Generally, it triggers a systemic viral sickness characterized by the painful bulging of the salivary and parotid glands, referred to as parotitis. Other symptoms include fever, headache, muscle aches and tiredness. In acute phase, MuV infection can affect other tissues and organs, resulting in severe inflammations, including orchitis, myocarditis, pancreatitis and nephritis[324]. In rare situations, the infection can also propagate to the central nervous system, causing meningitis and encephalitis (Figure 7.1).

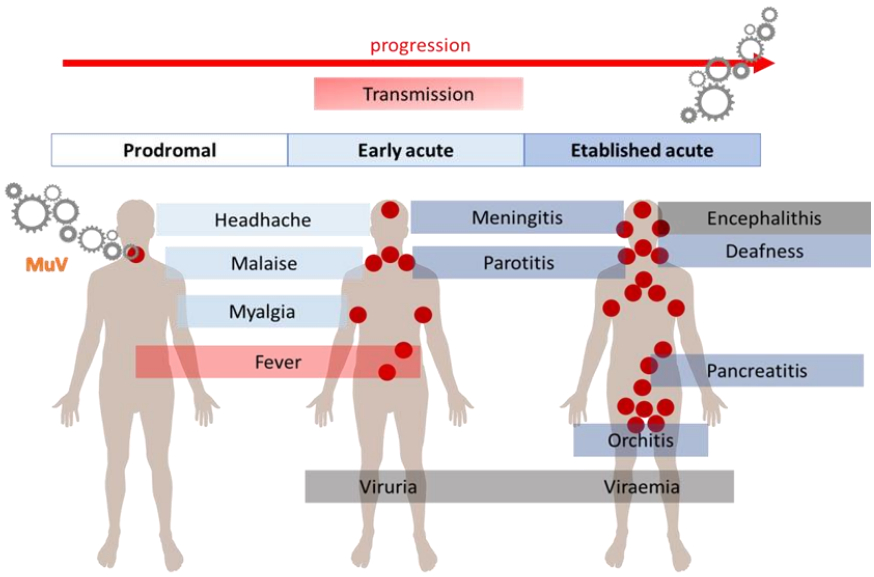


Figure 7.1 MuV Cinical presentation.

Despite the advent of an effective vaccine, mumps virus continues to spread throughout the world and although the majority of mumps infections are subclinical in vaccinated individuals, severe complications still arise mainly in underdeveloped countries. Moreover, in the last years, mumps outbreaks among vaccinated young adults have been also reported in different countries.[325,326]

In general, the MuV genome contains genes encoding for a set of 7 proteins, namely nucleocapsid (N), phospho (P), matrix (M), fusion (F), small hydrophobic (SH), hemagglutinin-neuraminidase (HN) and large (L) proteins[327,328],each performing roles that are fundamental for virus attachment, proliferation, assembly and budding (Figure 7.2).

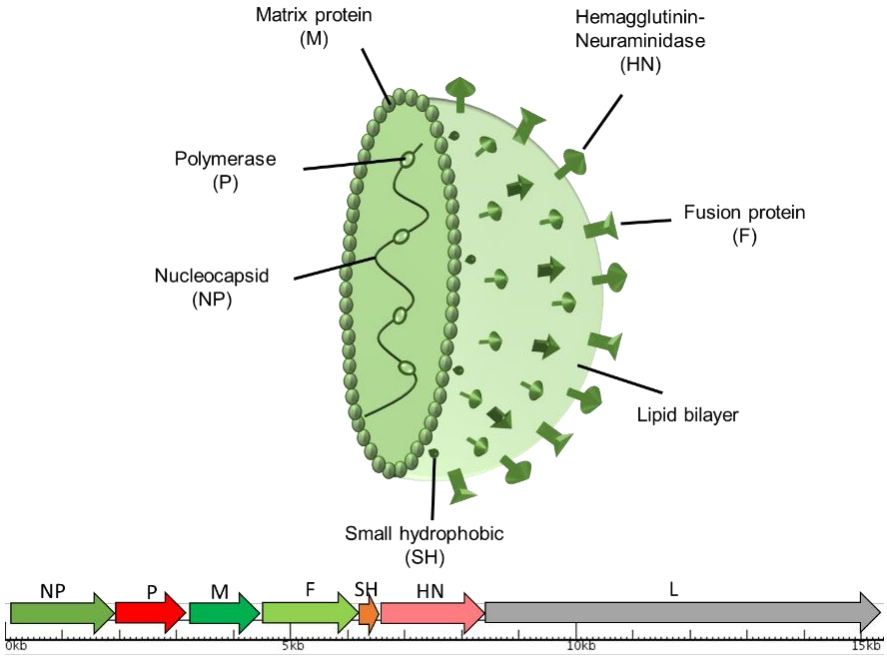


Figure 7.2 Viral particle composition and genome map of MuV-HN.

In brief, the N, P and L proteins are situated within MuV virion and are implicated in genome transcription and replication. The M protein is essential for virion assembly, propagation and is also involved the regulation of the transcription and replication phenomena. The SH protein is implicated in the evasion of the host immune response.[329] HN and F glycoproteins are mutually placed on the surface of the viral envelope and are responsible for adhesion, entry and fusion to the target cells.[330,331,332] SH, HN and F proteins represent the main target of neutralizing Abs produced by the host against MuV,[333] and also are responsible for the antigenic drift among different viral strains.

According to the current world health organization classification (WHO)[334], the MuVs strains are categorized into 12 genotypes, named as A to N, on the basis of nucleotide sequence analysis of the isolated strains genome.[335,336]

As for most of *Paramyxoviruses*, mumps virus HN protein (MuV-HN) possess both hemagglutinin and neuraminidase activity, hence playing a central role in the membrane fusion process during the virus entry as well as in the release and proliferation of the virus.[337] The MuV-HN neuraminidase belongs to the exo-α-sialidase enzyme class (EC number 3.2.1.18) which includes enzymes able to catalyze the hydrolysis of α(2,3), α(2,6), α(2,8) linked sialic acid residues.[338]

In detail, MuV-HN glycoprotein is a specific sialic acid-binding lectin able to recognize host cell-surface glycans which serve as attachment points for the virus, allowing not only its entry into the host cells but also the spreading of the infection. HN protein, indeed, also participates in the internalization of viral particles activating the fusion protein and permitting the fusion of viral membranes to the host cell. Finally, HN prevents self-agglutination of viral particles and favors the release of virions from the infected cells since it acts as a sialidase removing the sialic acid moiety from viral progeny (Figure 7.3).[339,340]

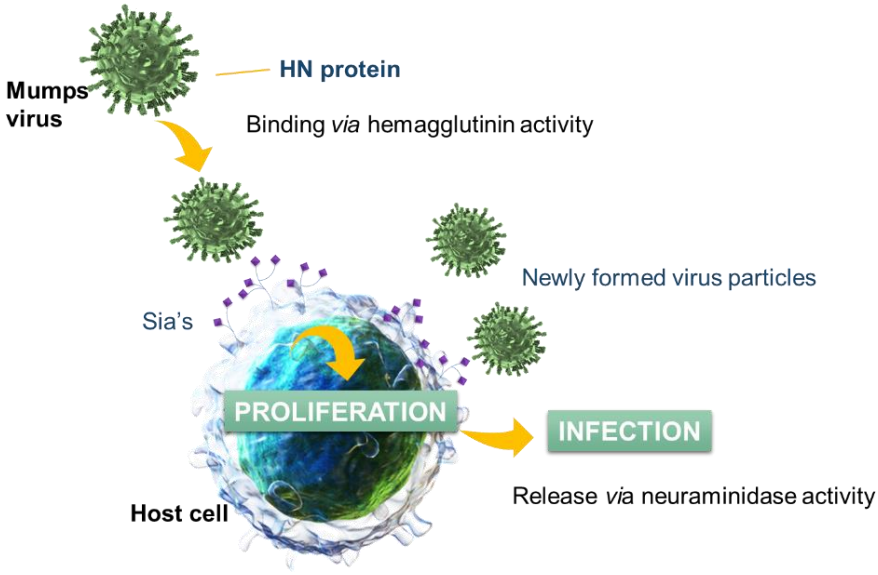


Figure 7.3 MuV-HN mechanism of action.

Chapter 7

The disruption of the interactions between virus-related HN glycoprotein and host glycans could prevent the virus attachment to host cells as well as its multiplication and release from the infected cells. Given the important roles in mediating the infectivity of the virus, MuV-HN represents an ideal target to impede either the viral infection and spreading, by hindering the virus attachment to host cell, the promotion of fusion activity and the neuraminidase function. Thus, MuV-HN has attracted attention as potential candidate for the development of more effective anti-viral pharmaceutics and vaccines.

Recently, the crystal structure of the MuV-HN, alone and in complex with different sialoglycans was solved, showing a net preference for α(2,3) -linked sialic acid in linear chains as glycan-receptors. Moreover, the co-crystals with 3'-sialyllactose revealed key interactions with a long saccharide chain.[341,342]

Here, with the aim to further describe the dynamics of recognition and binding processes at the basis of MuV infectivity, the molecular basis of sialylated ligands binding to MuV-HN were further investigated by a combination of NMR techniques and computational studies. In addition, the kinetic parameters of the sialoglycans hydrolysis catalyzed by the neuraminidase activity of HN were evaluated. The mechanisms of action of MuV-HN in the interplay with various sialosides was investigated using MuV-HN proteins derived from two different viral strains, namely Hoshino strain and SBL-1 strain, that belonged respectively to B and A genotypes. Receptor motifs that can be recognized by MuV-HN protein from these strains have been recently identified by glycan array screening as specific host-cell sialoglycans which terminate with Neu5Ac moiety. Among them, Neu5Acα2-3Galβ1-4GlcNAc (3'SLN), a sialoglycan broadly expressed in and exposed on various host tissues, was reported as one of the strongest binders; thus,

we decided to investigate its recognition by MuV-HN SBL-1 strain at molecular level by means of NMR spectroscopy, biophysical and computational approaches.

Furthermore, with the goal to provide the basis for the design and development of novel HN inhibitors, it has been carried out the characterization of the molecular interaction of MuV-HN protein from SBL-1 strain with a synthetic potential inhibitor of the neuraminidase activity by NMR, fluorescence analysis and computational studies. These results afforded novel insights into the molecular mechanisms of MuV infection, representing a step beyond into the development of more effective antiviral vaccines.

7.2 Molecular basis of sialoglycans binding to MuV-HN from Hoshino strain

Due to the ability of MuV-HN to recognize and hydrolyze sialoglycans exposed on host cell surface, we chose to investigate the binding by MuV-HN to different ligands (Figure 7.4), which are representative of glycosphingolipids, *O*-linked glycoproteins, as well as of *N*-linked glycans.

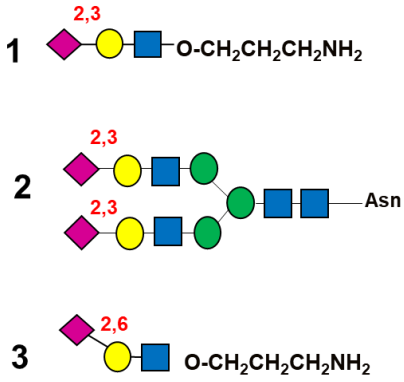
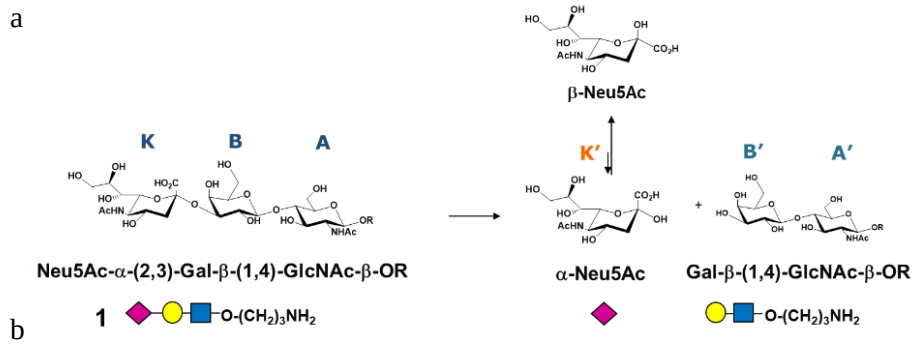


Figure 7.4 Substrates used in this study. They are representative of glycans exposed on glycolipids and glycoproteins of host cell surface

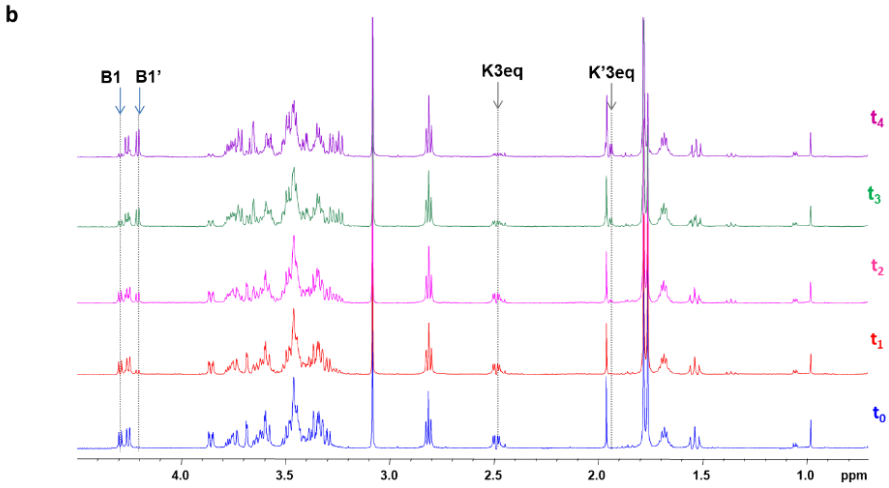
7.2.1 Kinetic analysis of the hydrolysis of sialoglycans by MuV-HN

At first, to get into the structural details on the hydrolytic activity of MuV-HN from Hoshino strain, we examined the kinetic mechanism of the hydrolysis. Indeed, the enzyme kinetic parameters were evaluated by progress curve analysis using NMR detection of products and substrates hydrolysis of substrates **1** and **2** (Figure 7.5)[343].

a



b



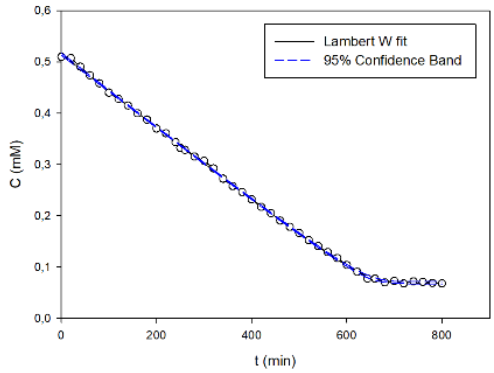


Figure 7.5 Kinetic analysis of MuV-HN hydrolysis of 3'SLN a) Scheme of the mechanism of hydrolysis of trisaccharide **1** catalyzed by MuV-HN neuraminidase. The SNFG nomenclature has been used. **b)** Section of 1H-NMR spectra at different time of the enzymatic reaction have been reported. 7.5 μM of MuV-HN protein and 550 μM of trisaccharide **1** were solved in PBS deuterate buffer (pH=7). The NMR quantification of the substrate concentration was performed by integration of the well-dispersed resonances of **1** (B1, K3eq). **c)** Analysis of the MuV-HN kinetics toward trisaccharide **1** by means of the explicit reformulation of the integrated form of MM equation with Labert W fit as solution. The plot of the concentration of trisaccharide **1** as function of time has been reported. The substrate concentration was evaluated from the $H3_{eq}$ resonance of Neu5Ac unit. The Lambert W fit (Equation 1) of the kinetic data by the Lambert-W fit afforded a KM value of 12 μM and Vmax of $7*10^{-3}$ (mM/min). The blue dashed line represented the confidence interval of the fit.

So, the hydrolysis of the sialylated trisaccharide **1** (Figure 7.5) was followed by ^{1}H NMR. As shown in the 1D NMR spectra reported in Figure 7.5, HN-MuV cleaved the terminal Neu5Ac moiety, producing *N*-acetyllactosammine (LacNAc, residues A' and B') and a residue of reducing Neu5Ac (*red*-Neu5Ac, K'). The progress of the hydrolysis was evident from the decrease of the signals of the substrate (as B1 and $K3_{eq}$) and the simultaneous increase of the intensity of the products resonances (as B'1 and $K'3_{eq}$), as indicated in Figure 7.5b. Expectedly, the hydrolysis reaction catalyzed by MuV-HN advanced with a retaining mechanism, i.e. with the net retention of configuration of the cleaved sugar [344,345]

(Figures 7.5 and 7.66); the consecutive and progressive anomerization process then caused the decrease of the α-Neu5Ac anomer in favor of the most stable β-Neu5Ac anomer.

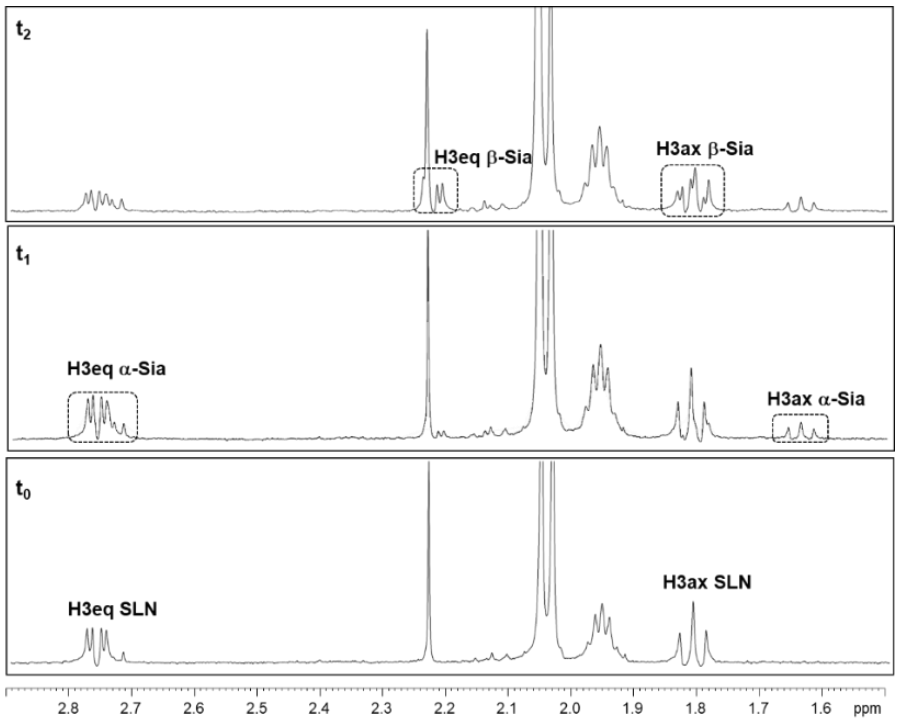


Figure 7.6 HN-MuV follows a retaining mechanism. 1H NMR analysis shows the production of α-Neu5Ac after incubation of MuV-HN with substrate **1**, demonstrating that MuV-HN follows a retaining mechanism. Indeed, MuV-HN initially produces α- Neu5Ac from 3' sialylactosamine; the α-Neu5Ac undergoes mutarotation to form the more stable anomer, β-Neu5Ac, over time.

To establish the kinetic parameters of the hydrolysis, selected isolated NMR resonances (as the signal of $H3_{eq}$ of the Neu5Ac unit in the substrate) were integrated and subsequently plotted as function of time. The variation of the substrate concentration with time can be efficiently described by the explicit

reformulation of the Michaelis–Menten kinetics[346] that provides the Lambert-W function as solution, developed by Goličnik (Equation 1).[347,348]

$$[S] = K_M W\left\{\frac{[S]_0}{K_M}\exp\left(\frac{[S]_0 - V_{max}t}{K_M}\right)\right\} \qquad \text{Equation 7.1}$$

Form the analysis of the kinetic data through a nonlinear least-square fit routine, it was possible to build the kinetic curve showed in Figure 7.5c and to evaluate K_M and V_{max} values. The fit of the experimental data showed a good quality, with a χ^2 value of 0.9998, and provided the kinetic parameters summarized in Table 7.1, including a K_M of 12 μM and a V_{max} of 7 $*10^{-3}$mM/min.

Table 7.1 Kinetic parameters obtained from the progress curve analysis for the hydrolysis of trisaccharide 1 and undecasaccharide 2 promoted by MuV-HN.

	K_M (μM)	**V_{max} (mM/min)**	**Kcat (min^{-1})**	**Kcat/K_M ($mM^{-1}min^{-1}$)**
trisaccharide 1	12	0.0007	0.1	8.55
undecasaccharide 2	3.5	0.000023	0.0025	0.65

Next, the same experimental procedure was carried out by using as substrate the longer sialylated *N*-glycan **2** (Figure 7.4). Similarly to ligand **1**, the NMR analysis supported the retaining mechanism but revealed a different kinetic activity of the enzyme toward the two different substrates, **1** and **2** (Table 7.1, Figure 7.7). In detail, the ^{1}H NMR spectra detected after incubation of **2** with MuV-HN showed the initial production of α-Neu5Ac, that was then followed by its anomerization (Figure 7.7b). Moreover, the comparison of the kinetic parameters calculated for the hydrolysis of **1** and **2** (Table 7.1) evidently demonstrated that the receptor was more active toward **1**, as advised by the 12-fold higher value of the kinetic efficiency (K_{cat}/K_M). [349]

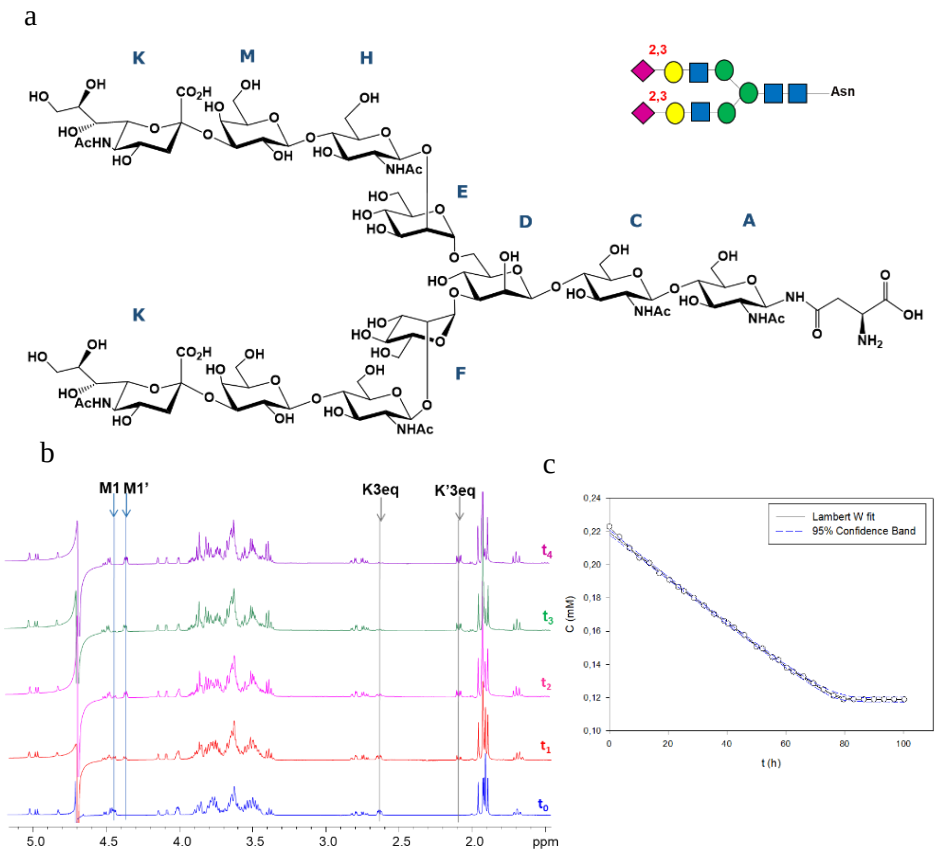


Figure 7.7 Kinetic analysis of the hydrolysis of undecasaccharide 2 by MuV-HN. a) Structure of the sialylated undecasaccharide **2** studied in the interaction with MuV-HN is reported by using the SNFG nomenclature. Chem Draw 2006 Software (http://www.cambridgesoft.com) was used to draw the undecasaccharide. **b**) Section of 1H-NMR spectra at different time of the enzymatic reaction of substrate **2** (T=298K, pH=7). The NMR quantification of the substrate concentration was performed by integration of the well-dispersed resonances of **1** (M1, K3eq). **c)** Analysis of the MuV-HN kinetics toward **2** by means of the Lambert W function. The fit of the kinetic data by the Lambert-W fit afforded a Km value of 3.5 μM and Vmax of $2.3*10^{-5}$ (mM/min). The blue dashed line represented the confidence interval of the fit.

7.2.2 Molecular recognition of sialoglycans by MuV-HN protein

The interaction between MuV-HN and different sialoglycans was described through ligand-based NMR techniques as STD NMR and transferred-NOESY. Due to the neuraminidase activity of MuV-HN, the NMR experiments were performed at 283 K to slow down MuV-HN hydrolysis rate.

The STD NMR analysis of the trisaccharide **1** mapped the glycan receptor interacting epitope, allowing for the identification of the ligand moieties participating in the interaction with MuV-HN (Figure 7.8).

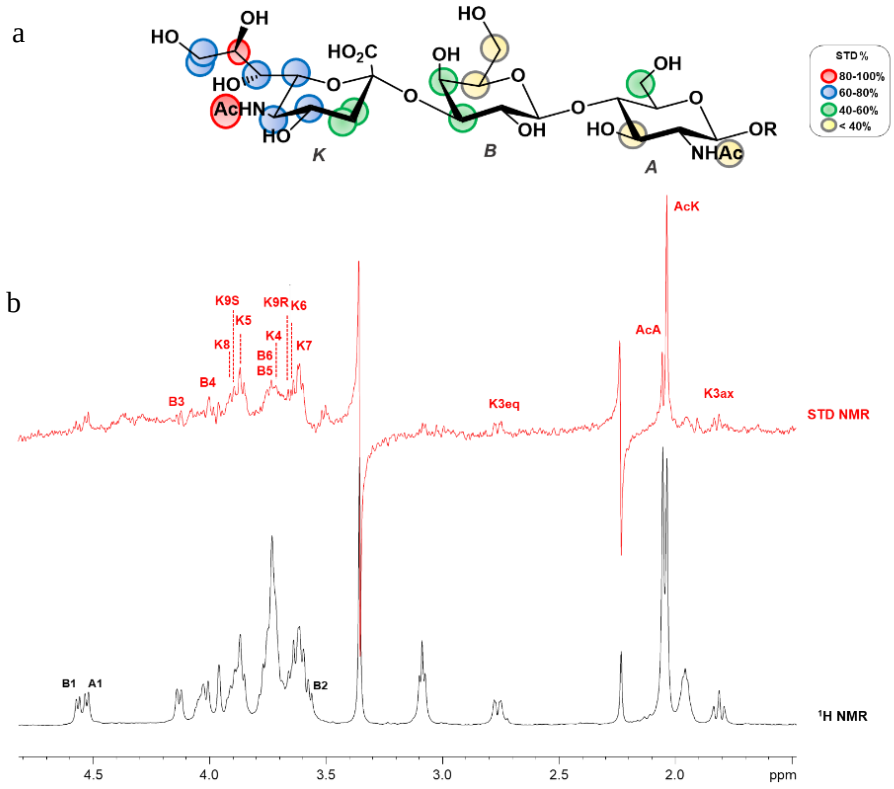


Figure 7.8 STD NMR analysis of trisaccharide 1 in the interaction with MuV-HN. a) Epitope map of **1**. STD percentages were calculated, and normalized with respect to the highest STD signal of the acetyl group belonging to sialic acid **b)** The off-resonance

spectrum as reference (black) and the STD (red) of MuV-HN−trisaccharide-**1** mixture with a molecular ratio of 1:70, at 283 K.

The comparison of the off-resonance and STD NMR spectra (Figure 7.8) highlighted changes in the intensity and multiplicity of specific signals. In detail, the Neu5Ac (**K**) was the portion of **1** mostly involved in the interaction with MuV-HN, with the highest STD effect belonging to the *N*-acetyl group and strong involvement in the interaction also detected for H8, (saturation transfer above 90%); also, protons H4, H5, H6, H7 and H9, exhibited a considerable participation to the binding process (above 70%). Lower STD signals were observed for the H3 axial and equatorial protons of **K**. These data confirmed the key role of the sialic acid residue in the interaction process, additionally a noticeable transfer of magnetization from MuV-HN to **1** was also noted for the Gal (**B**), especially to its H3 and H4 proton signals, that exhibited STD enhancements above 50%. Finally, proton signals H3, H6 and the *N*-acetyl group of A gave also rise to STD enhancements, although to a lesser degree, suggesting a minor contribution of the *N*-acetylglucosamine unit to the interaction with MuV-HN than K and B units. The different participation of the saccharide units of **1** in the recognition and binding process was also confirmed by the substantial difference in terms of STD intensity of the two singlets around 1.8 ppm respectively belonging to the acetyl groups of the GlcNAc (**A**) and Neu5Ac (**K**) residues (Figure 7.8). Thus, from the STD NMR analysis it was obtained the interacting epitope of the trisaccharide **1** (Figure 7.8) with MuV-HN, that was able to differently accommodate not only the Neu5Ac and Gal units, that established the most relevant contacts with the protein, but also the adjacent GlcNAc unit, which moderately contributed to the recognition process.

Interestingly, it was revealed the protein was able to bind the reducing sialic acid produced during the hydrolysis reaction. Indeed, STD NMR analysis conducted on the reaction mixture upon the hydrolysis of the trisaccharide **1** (Figure 7.9) displayed STD signals belonging to the more stable β-anomeric form (β-Neu5Ac) of the reducing Neu5Ac, as further confirmed by the STD NMR of free reducing sialic acid and the MuV-HN (Figure 7.10).

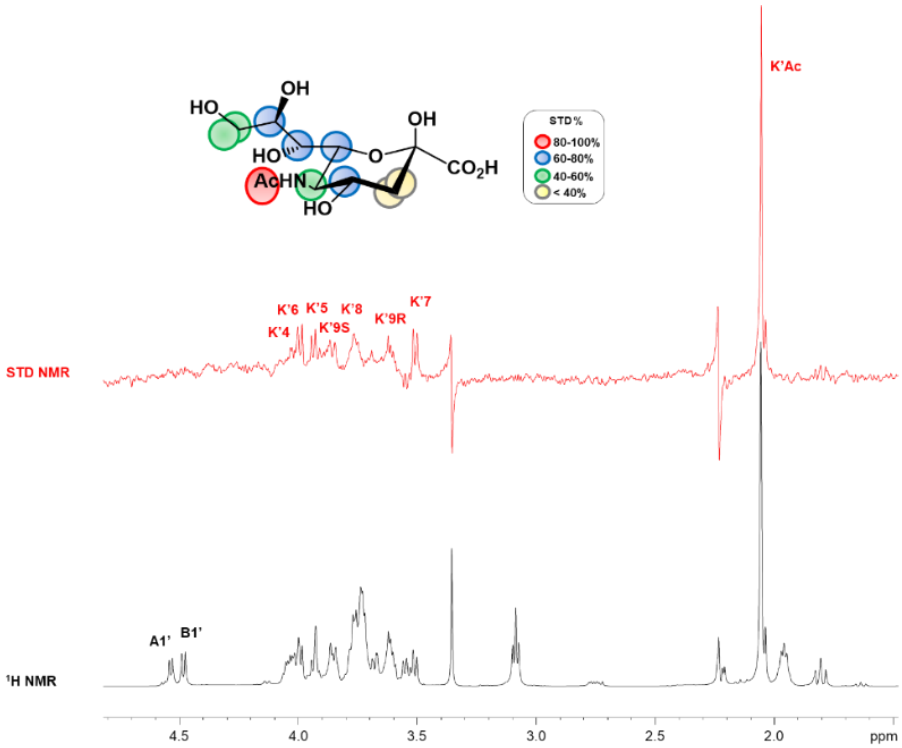


Figure 7.9 STD NMR analysis of reducing Neu5Ac produced during the hydrolysis of 1 by MuV-HN.

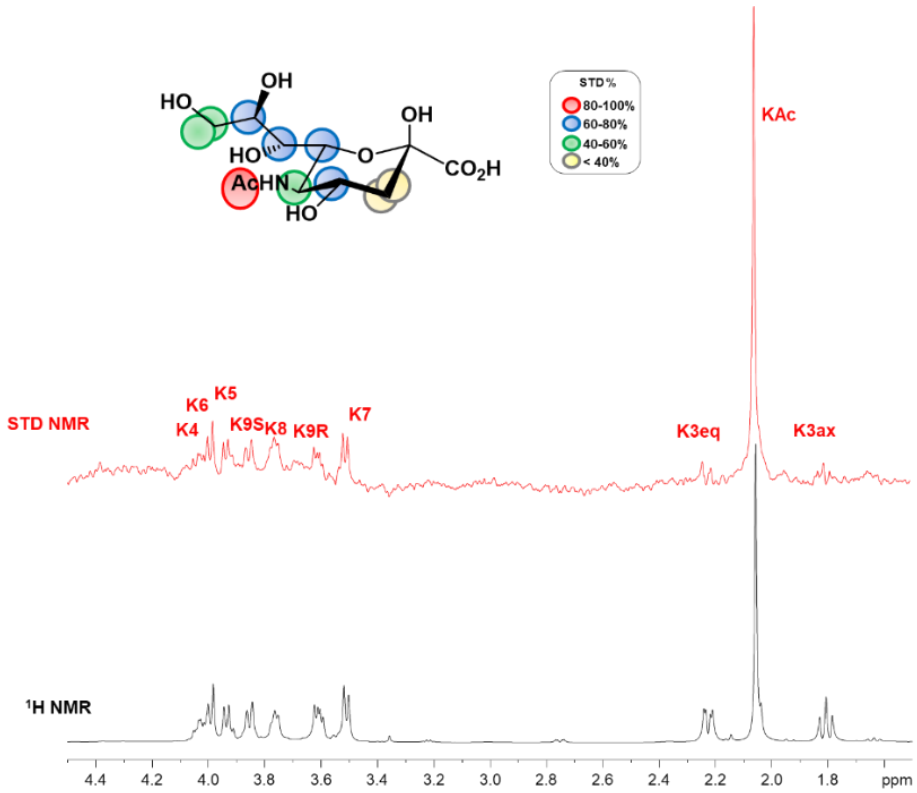


Figure 7.10 STD NMR analysis of reducing Neu5Ac in the interaction with by MuV-HN. The spectrum was recorded with a protein/ligand ratio 1:70 at 283K.

Furthermore, it was demonstrated the specificity of the hydrolysis by MuV-HN, as the trisaccharide **3**, containing α(2,6) -linked sialic acid, was neither recognized nor hydrolyzed by MuV-HN (see Figure 7.11, the lack of STD signals is diagnostic of absence of interaction).

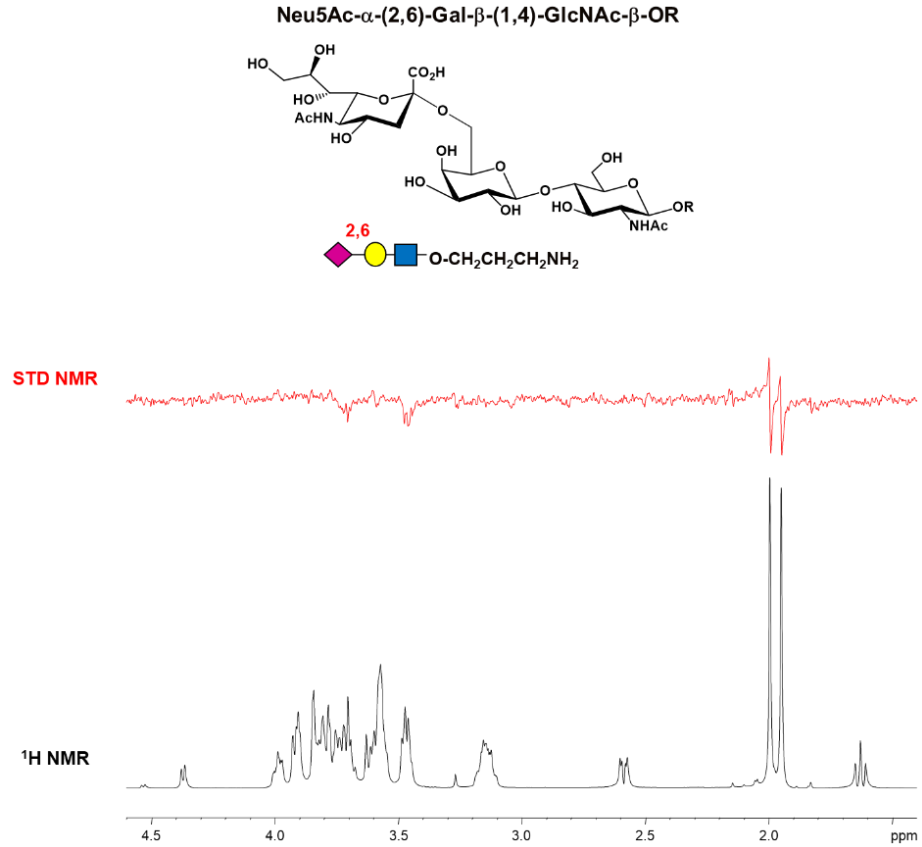
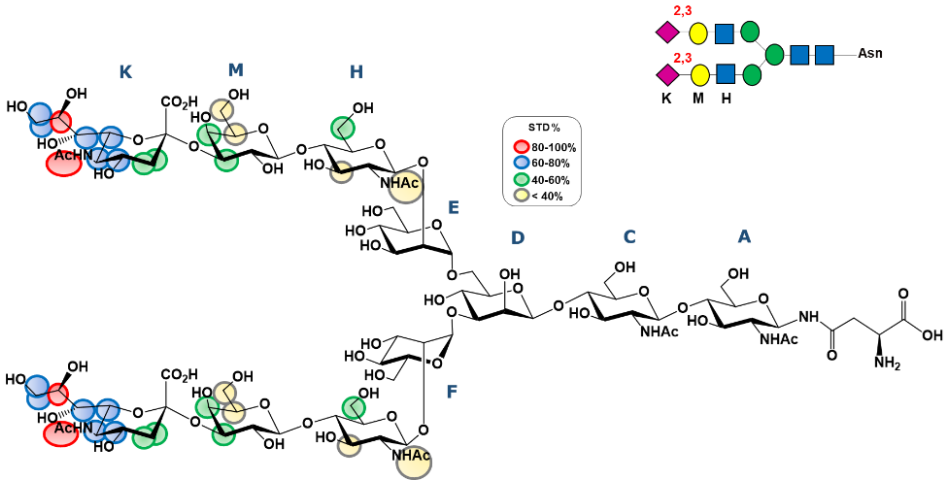


Figure 7.11 Structure of the Neu5Ac-α(2,6)-Galβ(1,4)-GalNAc (trisaccharide **3**) studied in the interaction with MuV-HN and the STD NMR experiment of trisaccharide **3** in the interaction with MuV-HN at 283K.

The ability of MuV-HN to recognize and interact with common cell surface sialylated glycoconjugates was then confirmed by investigating the interaction of MuV-HN with a longer sialoglycan, the undecasaccharide **2** (Figures 3 and 12), an acknowledged sialylated complex type glycan exposed on mammalian cells. In agreement with the epitope map of **1**, the STD analysis of **2** revealed that the sialic acid moiety received the largest fraction of saturation transfer. STD enhancements were also observed for protons belonging to galactose and *N*-

acetylglucosamine residues, while no STD contributions came from the other sugar units, thus suggesting that the accommodation of complex *N*-glycans in the MuV-HN binding site is not influenced by the length of the glycan chain. Nevertheless, the lower enzymatic activity of the enzyme toward the substrate **2** showed by the results of the kinetic analysis (see above) highlighted how the recognition, accommodation and subsequent enzymatic hydrolysis was partially hampered by the steric hindrance of the glycan chains.

a



b

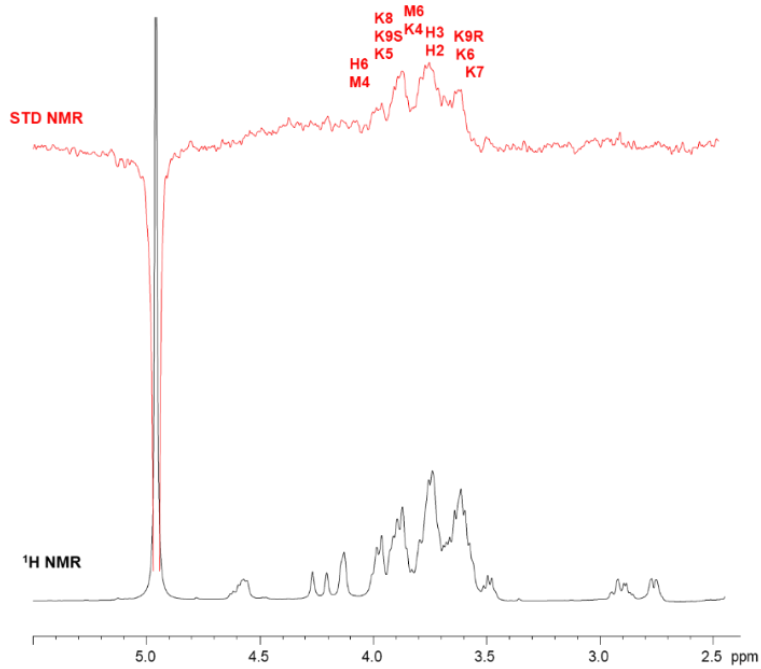


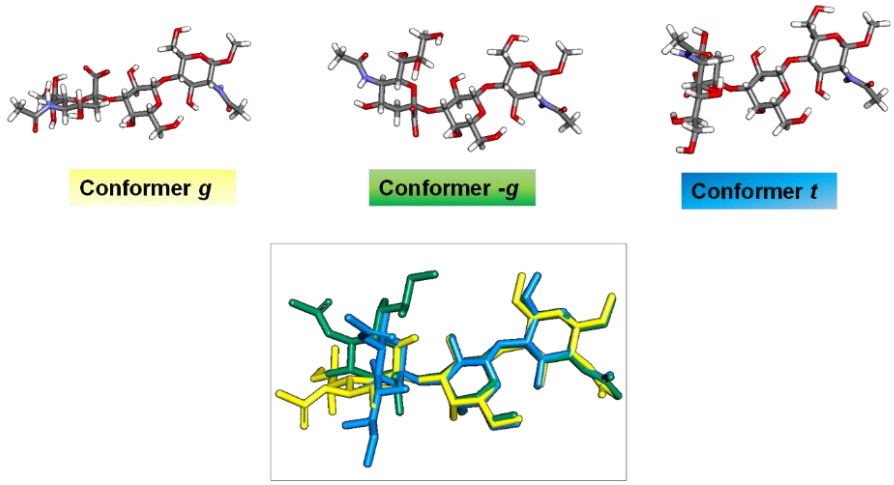
Figure 7.12 STD-derived Epitope mapping of undecasaccharide **2** and STD NMR experiment of **2** in the interaction with MuV-HN at 283K are shown.

Further details on sialoglycans recognition by MuV-HN were collected by the assessment of the bioactive conformation. The evaluation of the conformational behavior of **1** was performed by analyzing NOE and tr-NOE crosspeaks obtained in 2D NOESY spectra acquired both in free and in bound states (Table 7.2).

Table 7.2 Conformational analysis of trisaccharide 1 in the free and bound states compared with respect to the conformer populations described by the ϕ torsion angle values, according to the literature.[350] The cross-peak intensities, correlated to the ^{1}H-^{1}H *inter*-molecular distances, are indicated as strong (S), medium (M) or low (L).

Distance	Conformer			Exp. Free state	Exp. Bound state
	g **Φ = -60° Ψ = 0°**	**+*g*** **Φ = 60° Ψ = 0°**	***t*** **Φ = 180° Ψ = 0°**		
H3 Gal-H3$_{ax}$ Neu5Ac	4.19	3.11	**2.18**	2.74	2.35
H3 Gal-H3$_{eq}$ Neu5Ac	4.20	2.03	**3.40**	3.36	3.43
H4 Gal-H3$_{eq}$ Neu5Ac	4.15	2.25	**5.15**	4.24	/
H4 Gal-H3$_{ax}$ Neu5Ac	5.04	2.13	**4.45**	3.94	4.31

The NOESY experiment revealed that, in the free state, the trisaccharide **1** exhibited an equilibrium between different conformational states, primarily described by the so known *–g*, *g*, *t* conformers and that were distinguished by the ϕ (H1-C1-O-CX') torsion angle around the Neu5Ac-α–(2,3)-Gal glycosidic linkage (-60°/60°/180° respectively)[351] (Figure 7.13 and 7.14).



Neu5Ac-(2,3)-Gal	*g*	*-g*	*t*
ϕ (C1-C2-O1-C3')	60	-60	180

Figure 7.13 Conformational behavior of trisaccharide 1. The three major conformations in solution of the trisaccharide **1** that differ from the value of the phi torsion of the Neu5Acα2-3Gal linkage: the g, -g and t. The glycosidic torsion angles definition was ϕ= O5-C1-O1-C6', ψ= C1-O1-C6'-C5'.

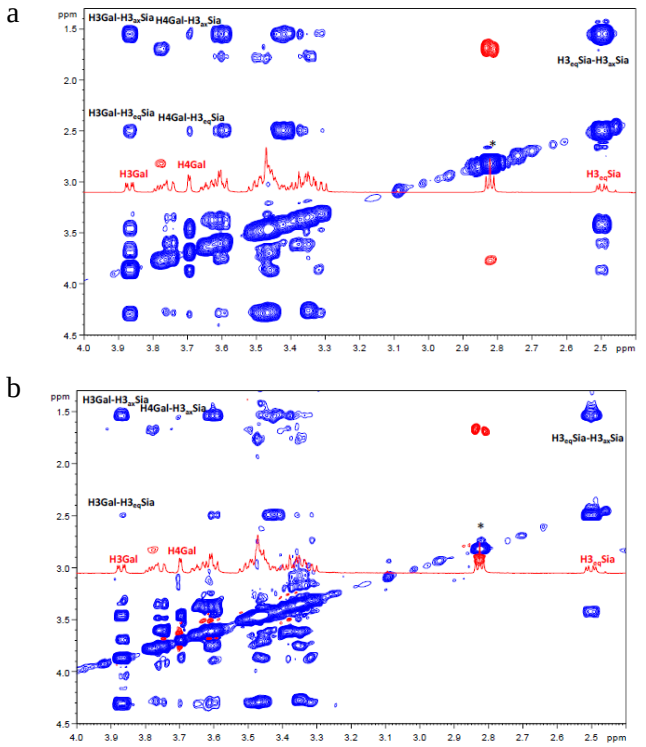
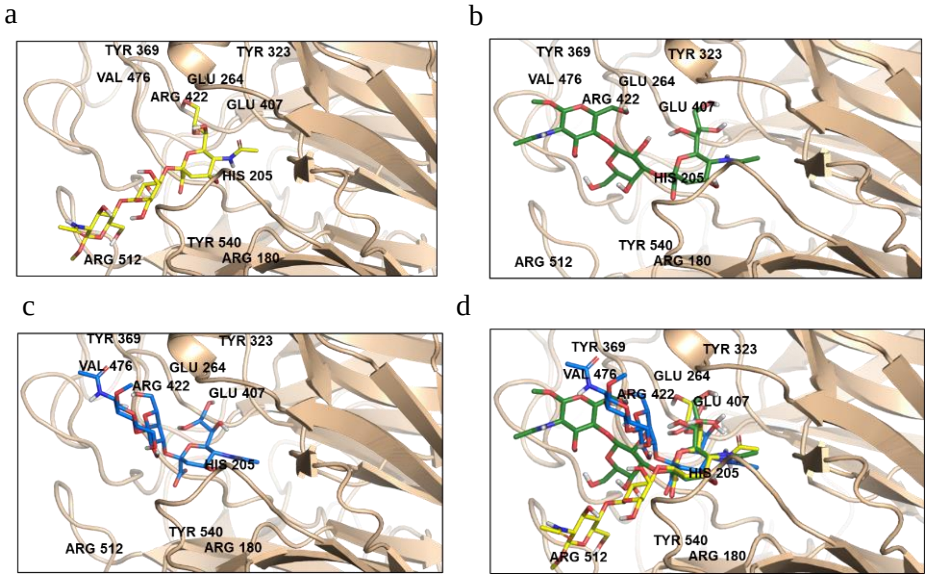


Figure 7.14 NOESY and tr-NOESY analysis of substrate 1. NOESY spectrum **a)** of **1** in its free state (mixing time of 600ms) and tr-NOESY spectrum **b**) of **1** bound to Muv-HN (mixing time of 400ms) at 283K. The protein-ligand molar ration was set at 1:30. Some differences in terms of signal intensities were observed comparing the free and bound state. The signal at 2.8 ppm marked with the asterisk belongs to the aglycon moiety at reducing end of **1**, which exhibits a high degree of flexibility.

Among the others, the key distances between the H-3 methylene protons of the sialic acid and H3 and H4 protons of the galactose residue suggested that trisaccharide **1** exhibited a preference for the *t* conformer in the bound state (Table 7.2).

These data were confirmed by employing a combined approach based on docking calculations and the use of CORCEMA-ST. First, docking calculations of the MuV-HN/ 1 complex were performed using the recently published crystal structure (PDB ID: 5B2C) of MuV-HN receptor-binding head domain.[340]The trisaccharide **1** was built and optimized by means of Maestro suite of programs. All the three main conformers, namely *-g*, *g*, *t*, were considered in the calculations using AutoDock. The docking analysis underlined that, for each conformer, a consistent binding mode with good theoretical energy was possible (Figure 7.15).



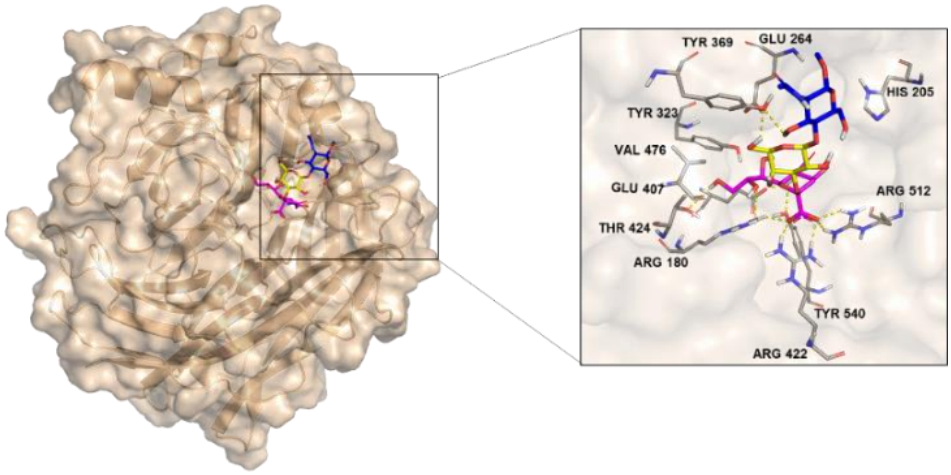
Conformer	Cluster rank	Number of conformations in this cluster	Number of conformations in this cluster (Å)	Estimated Free Energy of Binding (kcal/mol)	Estimated Inhibition Constant, Ki (mM)
t	1	124	55.996	-3.71	1.91 mM
g	1	104	56.729	-3.52	2.63 mM
-g	2	98	57.457	-2.75	9.72 mM

Figure 7.15 3D model of MuV-HN in complex with substrate 1 derived by docking calculations. a-d) Binding pocket of MuV-HN in the presence of the substrate **1** in the *g*, *-g* and *t* conformation, as derived by Autodock program. Only the amino-acid residues involved in the binding process are depicted. Green dotted lines stand for *inter*-molecular

hydrogen bonds. Pymol 2.3 Software (https://pymol.org/2/) was used to draw the figures. Docking results for the three possible conformers of substrate **1** obtained are also shown.

Thus, three representative complexes were submitted to CORCEMA-ST program to compare calculated and experimental STD effects for each conformer. As result, the complex built with the *t* conformer (Figure 7.16) exhibited the best agreement between theoretical and experimental STD data (R-NOE of 0.23, Figures 7.16 and 7.17), among the three examined structures, thus confirming that the ligand preferentially assumed a *t* conformation around Neu5Ac-Gal linkage upon the interaction with MuV-HN.

a



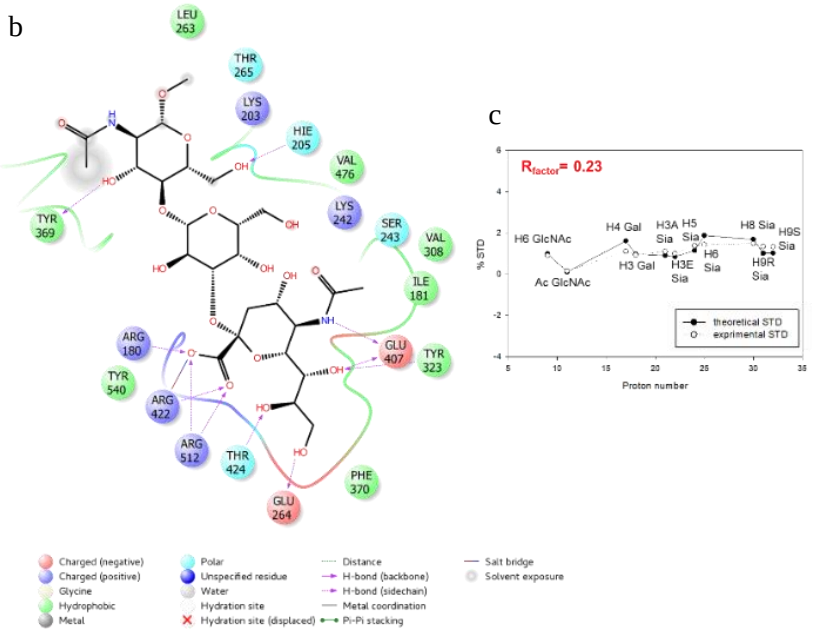


Figure 7.16 3D model of MuV-HN in complex with trisaccharide 1 and CORCEMA analysis. a) Close up view of ligand **1** binding mode at the MuV-HN active site. The main amino acid residues involved in the binding are represented in stick. Galactose and N-acetyl glucosamine residues are depicted in yellow and blue, respectively. **b)** Two-dimensional plot illustrating the interactions of the sialylated trisaccharide **1** with the residues in the binding pocket of MuV-HN. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows such with functional groups of the backbone. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions. **c)** Comparison between experimental (dashed line) and theoretical (solid line) STD data for the selected 3D complex of the interaction between trisaccharide **1** and MuV-HN head domain.

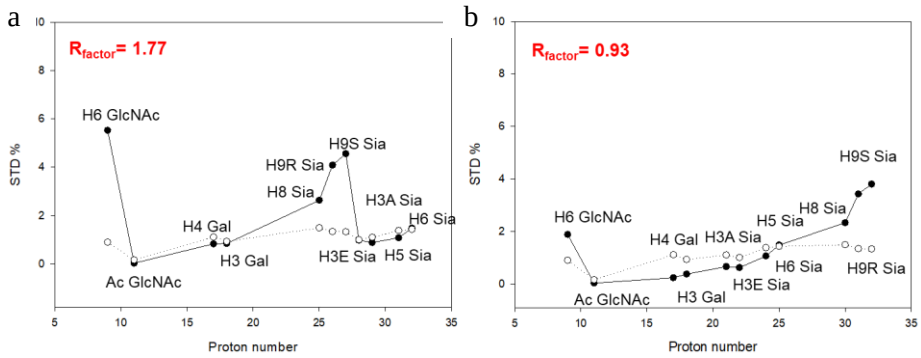


Figure 7.17 CORCEMA-ST analysis of other 3D complexes obtained by docking calculations using different conformers. a) -g conformer **b)** g conformer.

From the CORCEMA-ST prediction of the *t* conformer complex, protons from all the trisaccharide units exhibited STD effects, thus assessing their participation in the interaction with MuV-HN. In detail, the highest predicted STD effects belonged to the acetyl group of Neu5Ac, due to the strong hydrogen bond between the *N*-acetamide moiety of Neu5Ac and the Glu407 backbone, as well as hydrophobic contacts involving the acetyl group located in a hydrophobic cleft shaped by Val and Ile residues. Also, H8 proton of Neu5Ac unit displayed more than 80% of saturation, consistent with the contacts occurring between the hydroxyl group in position 8 and Tyr323 and Thr424 residues of MuV-HN (Figure 7.16a-b). Significant STD effects were also predicted for H5, H6, H7, H9 of Neu5Ac, all involved in interactions with Tyr323 and Glu264 residues. Lower STD enhancements were instead predicted for $H3_{eq}$ and $H3_{ax}$ of Neu5Ac and H3, H4, H5, H6 of Gal in agreement with their proximity to the receptor surface and the weaker hydrophobic contacts observed in the complex. On the other hand, no theoretical STD NMR effects arose for H1 and H2 protons of Gal, as they pointed outside the glycan-receptor binding pocket. Finally, STD enhancements were

predicted for the *N*-acetamide moiety and the protons H-2, H-4, H-6, all belonging to the GlcNAc face that was straight directed toward the MuV-HN protein surface. It is worth to note that the profiles of the predicted STD effects, belonging to the other two possible conformers (*-g* and *g*, Figure 7.17), were very different with respect to the corresponding experimental STD enhancements and for that, resulted in higher R-NOE values. Therefore, CORCEMA-ST analysis allowed to further exclude that trisaccharide **1** adopted these conformations in the interplay with MuV-HN.

Overall, the combination of NMR, docking and CORCEMA analysis permitted to draw an accurate model of the interaction between trisaccharide **1** and the MuV-HN head domain (Figure 7.16a) and select the right conformer among three different possible available. According to the selected model, the *t* conformer of **1** was able to fully fit the binding pocket of the MuV-HN, characterized by a rather extended topology, which allowed to form various interactions with almost the entire sugar backbone of **1**, as clearly displayed in the 2D representation in Figure 7.16b. In particular, the trisaccharide **1** stretched in the binding pocket of the MuV-HN, established several polar contacts with the active site residues that are conserved among all MuV genotypes. The major determinant of the binding was the sialic acid moiety that established electrostatic interactions and hydrogen bonds with catalytic site MuV-HN polar residues mainly through its carboxylate, glycerol chain and *N*-acetamide group. The strong ionic interactions formed by the Neu5Ac carboxylate group and the guanidinium groups of Arg180, Arg422 and Arg512 of the MuV-HN were essential for viral binding.[352] The Neu5Ac unit was further involved in hydrogen bonds through the hydroxyl groups of its glycerol chain and Tyr323, Glu264 residues of the MuV-HN, together with the interaction between its *N*-acetamide moiety and Glu407 residue. Also, the

acetyl group of Neu5Ac contributed to the binding, being engaged in hydrophobic interactions with the receptor non-polar residues. Even the Gal moiety was in close contact with the protein, though stabilized by weaker interactions with the MuV-HN surface, such as the contacts involving Val476 and Tyr369 residues of the binding cavity of MuV-HN. Finally, the GlcNAc unit mainly contributed to the overall interaction with the MuV-HN highly conserved Tyr369 and His205 residues, though displayed a lower contribution to the recognition and binding processes. These results deepened the previously underlined[328,329] contribution of the third sugar unit from the reducing terminal in the interaction with MuV-HN, together with the crucial role of Tyr369 in the recognition of sialylated glycan receptors.

Interestingly, docking calculations carried out using the *red*-β–Neu5Ac, showed that it displayed a different orientation inside the receptor pocket with respect to the Neu5Ac unit of **1** (Figure 7.18), forming strong ionic interactions between its carboxylic group and the highly conserved active site arginine triad, but resulting in a slightly different set of interactions involving its lateral chain and *N*-acetamide moiety. Particularly, the *N*-acetyl group of Neu5Ac formed a hydrogen bond with Glu264 residue in place of Glu407, whereas the lateral chain engaged polar interactions mainly with Tyr323 and Glu407 residues. Furthermore, the hydroxyl group in position 2 of the sugar was involved in a hydrogen bond with the Tyr540 residue.

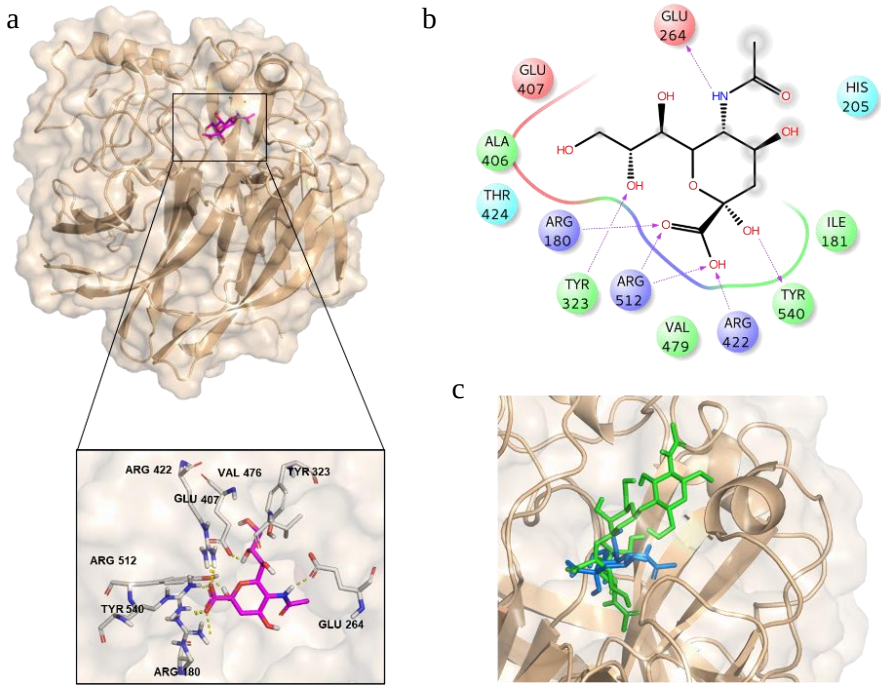
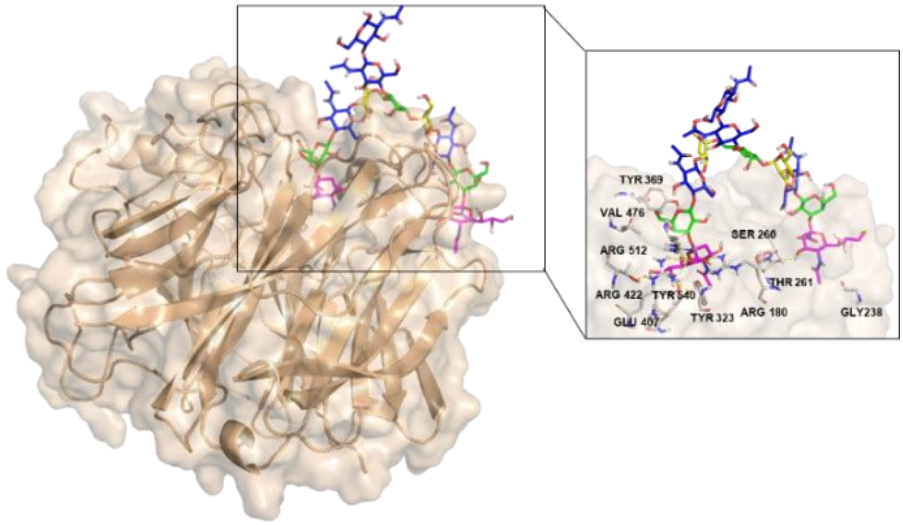


Figure 7.18 3D model of MuV-HN in complex with β-Neu5Ac. a) Close up view of β-Neu5Ac binding mode at the MuV-HN active site. The main amino acid residues involved in the binding are represented in stick. **b)** Two-dimensional plot illustrating the interactions of the β-Neu5Ac with the residues in the binding pocket of MuV-HN. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows such with functional groups of the backbone. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions. Although the different orientation inside MuV-HN binding site of the β-Neu5Ac with respect to the α-anomer in **1**, the triad of arginine residues of the MuV-HN binding pocket was again engaged in strong ionic interactions with the Neu5Ac carboxylate. **c)** Superimposition of β-Neu5Ac and trisaccharide **1** in the MuV-HN binding site.

On the other hand, docking calculations using the longer complex type N-glycan, highlighted that the pattern of interactions established in the undecasaccharide **2**/HN complex was comparable to that of the trisaccaride **1**/HN (Figure 7.19); the

triad of arginine residues of the MuV-HN binding pocket was again engaged in strong ionic interactions with the Neu5Ac's carboxylate; the Neu5Ac's lateral chain established hydrogen bonds with the polar residues of the receptor, however, a reduced polar network was observed (Figure 7.19).

a



b

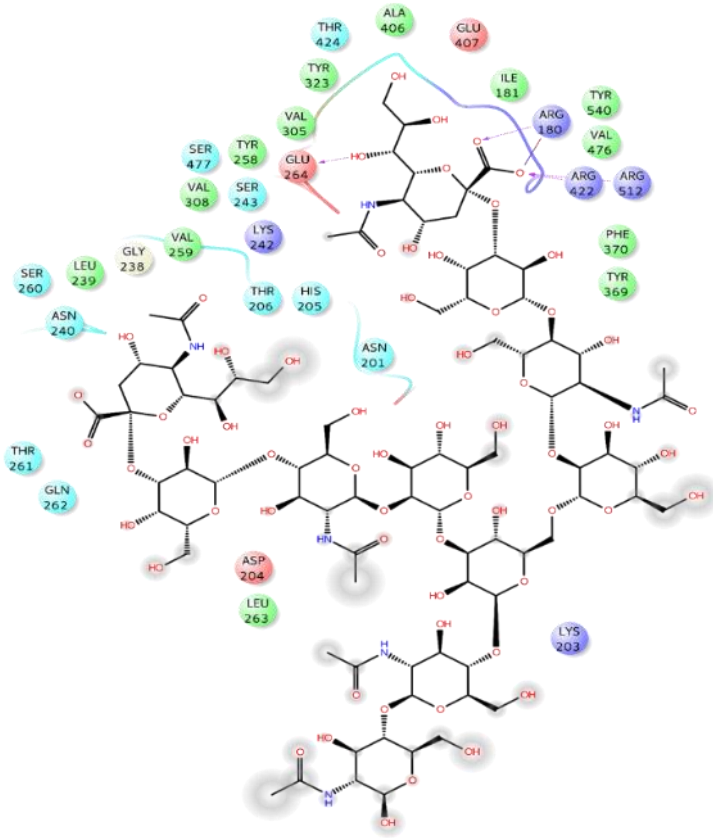


Figure 7.19 3D model of MuV-HN in complex with undecasaccharide 2. a) Close up view of ligand **2** binding mode at the MuV-HN active site (arm 1-6). The main amino acid residues involved in the binding are represented in stick. Galactose and *N*-acetyl Glucosamine residues are depicted in yellow and blue, respectively. **b)** Two-dimensional plot illustrating the interactions of the sialylated undecasaccharide **2** with the residues in the binding pocket of MuV-HN. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows such with functional groups of the backbone. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions.

MuV-HN was indeed able to accommodate either the α–1,6 arm or α–1,3 arm (data not shown) of the sialylated undecasaccharide in its binding pocket and,

in both cases, the highly conserved electrostatic interactions involving the Neu5Ac carboxylate and the Arg180, Arg422, Arg512 residues were observed. However, the MuV-HN/**2** complex did not show the hydrogen bonds formed by Glu407 and Tyr323 with Neu5Ac unit, along with the interactions between the third sugar unit and Tyr369 residue. This was likely due to the greater steric hindrance given by the length and branching of **2**, that hindered the optimal accommodation of its terminal moiety into the binding pocket of the protein. This is also in accordance with previous binding assay experiments which revealed a lower affinity of MuV-HN for branched complex-type branched glycan receptors.[328] Interestingly, the sialic acid unit at the terminus of the glycan antenna that is not accommodated into the binding pocket engaged polar interactions with the MuV-HN outer surface residues (Gly238, Ser260, Thr260). These non-specific interactions could account for the higher K_M value found with the NMR kinetic analysis of the undecasaccharide **2**.

7.3 Molecular basis of sialoglycan binding to MuV-HN from SBL-strain

First, the enzymatic activity of MuV-HN from SBL-1 strain was probed toward trisaccharide **1** (3'SLN). Upon addition of the substrate in solution, MuV-HN catalyzed the cleavage of the terminal Neu5Ac moiety, producing *N*-acetyllactosammine (LacNAc, residues A' and B') and reducing sialic acid (*red*-Neu5Ac, K') respectively (Figure 7.20).

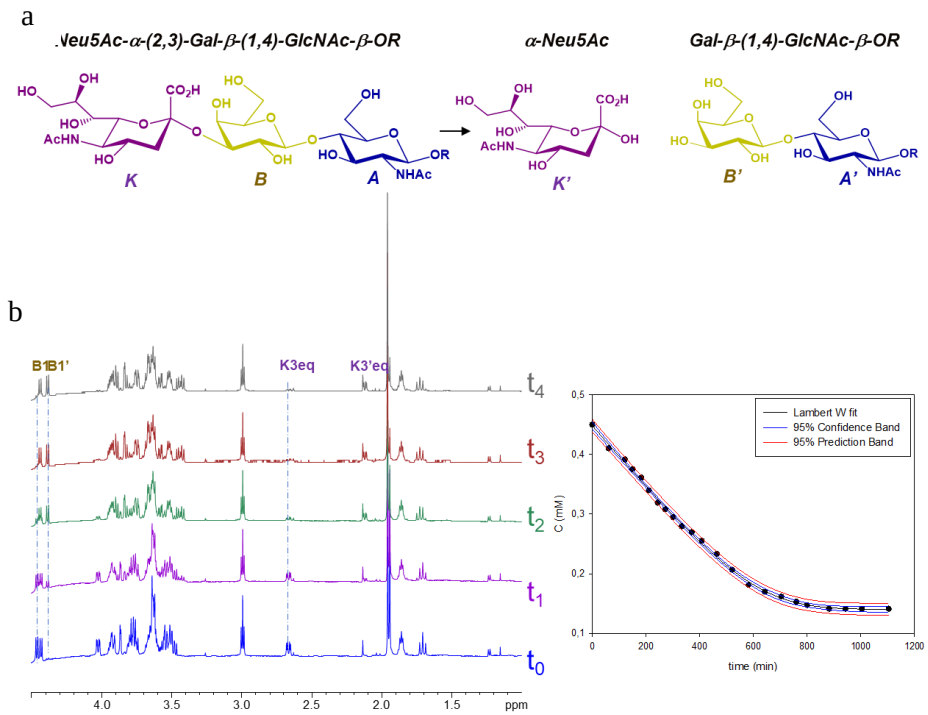


Figure 7.20 Kinetic analysis of the hydolysis of 3'SLN by MuV-HN (SBL-1 strain). **a)** Scheme of the mechanism of hydrolysis of 3'SLN catalyzed by MuV-HN (SBL-1 strain). **b**) (left panel) Section of ^{1}H-NMR spectra at different time of the enzymatic reaction have been reported. The spectra were recorded using 13 μM of MuV-HN protein and 700 μM of 3'SLN in PBS deuterate buffer (pH=7). The NMR quantification of the substrate concentration was performed by integration of the well-dispersed resonances of the product (B1, $K3_{eq}$). (right panel) Analysis of the MuV-HN kinetics toward 3'SLN by means of the explicit reformulation of the integrated form of MM equation with Lambert W fit as solution. The plot of the concentration of 3'SLN as function of time has been reported. The substrate concentration was evaluated from the $K3_{eq}$ resonance of Neu5Ac unit. The fit of the kinetic data by the Lambert-W fit produced a K_M value of 5 μM and V_{max} of $6*10^{-3}$ (mM/min). The blue dashed line represented the confidence interval of the fit and the red dashed lines are the prediction bands.

Therefore, the real time kinetics of 3'SLN hydrolysis by MuV-HN was followed by 1D ^{1}H-NMR, as reported in Figure 7.20b, where the overlapped NMR spectra

detected over time progression are shown. The progress of the hydrolysis was assessed by monitoring the decrease of the signals of the substrate (as B1 and $K3_{eq}$) and the concurrent appearance and following enhancement of the intensity of the products resonances (as B'1 and $K'3_{eq}$). Similarly to other viral sialidases,[353] MuV-HN acts as retaining glycosidase, with a net retention of substrates configuration. The $H3_{eq}$ resonance of Neu5Ac was monitored to evaluate the variation of the substrate concentration and, by fitting the kinetic data through the Lambert W function[354], the values of the enzyme K_M and V_{max}, were determined (Figure 7.20). Notably, a slightly increased affinity (K_M) for 3'SLN was observed for SBL-1 strain with respect to the Hoshino strain [355] suggesting a higher enzymatic affinity of the protein to the glycan, in agreement with the stronger binding observed by previous glycan array data.[329]Despite the neuraminidase activity of MuV-HN induces the cleavage of the sialic acid within NMR time scale, STD NMR allowed to characterize the ligand binding epitope (Figure 7.21).

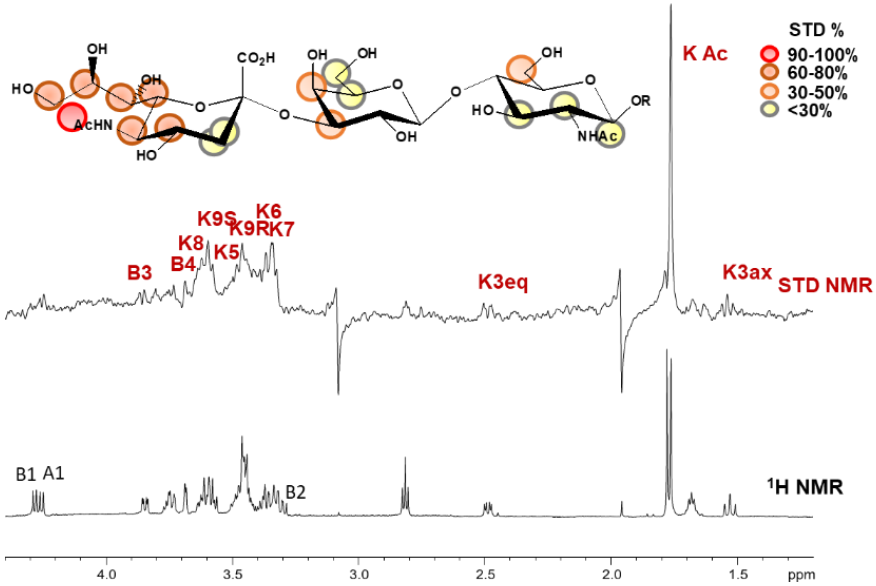


Figure 7.21 STD NMR analysis of 3'SLN in the interaction with MuV-HN from SBL-1 strain. (top panel) Interacting epitope map of 3'SLN as derived by STD NMR data. (bottom panel) The ^{1}H NMR spectrum and the STD NMR spectrum of MuV-SBL-1–3'SLN mixture with a molecular ratio of 1:70, at 283 K.

The qualitative analysis of the STD effects demonstrated that the entire trisaccharide was accommodated into the binding site of the protein. The moiety that contributed the most to the interaction was the Neu5Ac (**K**) which established several contacts with the protein, not only through its acetamide portion but also via the glycerol chain. The highest STD effect, set at 100%, arose from the N-acetyl group; protons H7 – H8 – H9 likewise showed significant STD enhancements; finally, STD signals were detected for certain protons of the Neu5Ac carbohydrate ring, as H4, H5 and H6. As already observed in the case of MuV-HN from Hoshino strain, also the other two residues (**A** and **B**) were involved in close contact with MuV-HN although to a lower extent. Furthermore, from the evaluation of the relative intensities of two N-acetyl groups of Neu5Ac **K** and GlcNAc

A, it was clear that the contribution from the acetamide of the GlcNAc was not significant.

To get structural insights into sialoglycans recognition by MuV-HN (SBL-1 strain, genotype A), comparative homology modelling was carried out considering as template the crystal structure of MuV-HN head domain (Hoshino strain, genotype B), PDB ID: 5B2C, which possessed above 95% of sequence similarity to the target sequence (Figure 7.22), in accordance with the little number of genomic mutations observed among different MuV genotypes at the gene coding for HN protein.

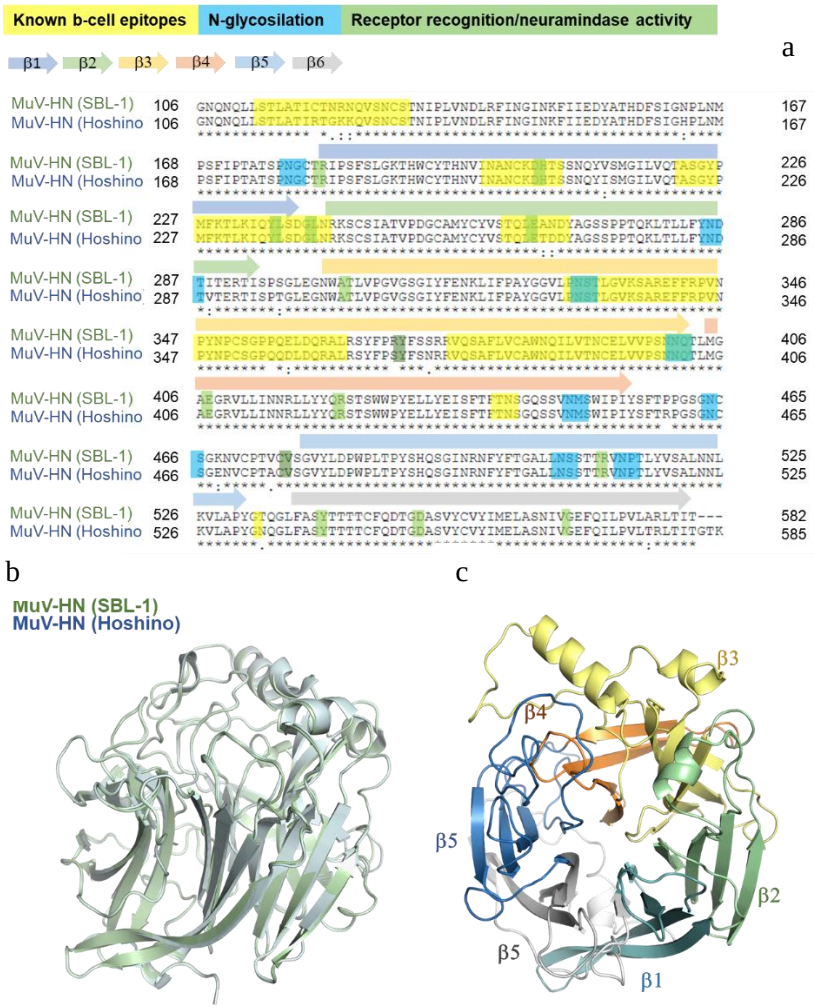


Figure 7.22 3D model of MuV-HN from SBL-1. a) CLUSTAL alignment of the head domain of MuV-HN from different strains. Amino acids involved in molecular recognition and neuraminidase activity are highlighted in green, N-glycosylation sites are highlighted in blue and known B cell epitopes in yellow. Sequences corresponding to the six four-stranded beta sheets are indicated with differently colored arrows **b)** Superimposition of the head domains of MuV-HN SBL-1 model (green), and MuV-HN Hoshino strain crystal structure, PDB ID: 5B2C (blue). **c)** MuV-HN monomer six bladed propeller structure, whose constituent structural elements (β1–β6 sheets) are differently colored.

Interestingly, from the sequence analysis of the hemagglutinin-neuraminidase from the two different strains (Figure 7.22), the most relevant modifications of the amminoacid sequences occurred at the well-established B-cell epitopes, particularly at 327–363; 375–403; 440–443; and 533 regions, as similarly detected when comparing other MuV-HN genotypes.[356]

The homology model of HN protein from SBL-1 strain was then built using Swiss Model server and its quality was assessed through SAVES resource. The corresponding Ramachandran plots generated by means of the PROCHECK tool showed that 98% of the residues' dihedrals belonged to the core/allowed regions, hence confirming the fitness of the model. Comparing the refined model with the target structure, a similar 3D topology was noted (Figure 7.22). As predicted by the sequence analysis, changes in the amminoacid sequence among the proteins from the two different strains did not affect neither the overall structure of the sialic acid binding motif nor the accessibility of the protein to the glycans on the host cell surface. In detail, MuV-HN (SBL-1 strain) model showed a globular structure, characterized by six four- stranded antiparallel β-sheets (β1-β6) arranged in the so-called bladed beta propeller fold, that is characteristic of many viral neuraminidases[357] (Figure 7.22c). At the center of the domain, it was located the dual binding/ sialic acid cleavage site of the protein, comprising the conserved sialidase active site residues, namely Arg180, Glu407, Arg422, Arg512, Tyr540, Glu561, and Asp204.[341] Furthermore, the binding site comprised the Tyr369 and Val476, highly conserved among the MuV-HN genotypes, that revealed to be relevant factors of Mumps glycans' recognition.[341,342]As several studies highlighted the importance of the flexibility of specific loops in other viral sialidases,[358] the MuV-HN conformational behavior was investigated by MD simulations. First, the MD simulation of the apo form was carried out using AMBER 18 (Figure 7.23).

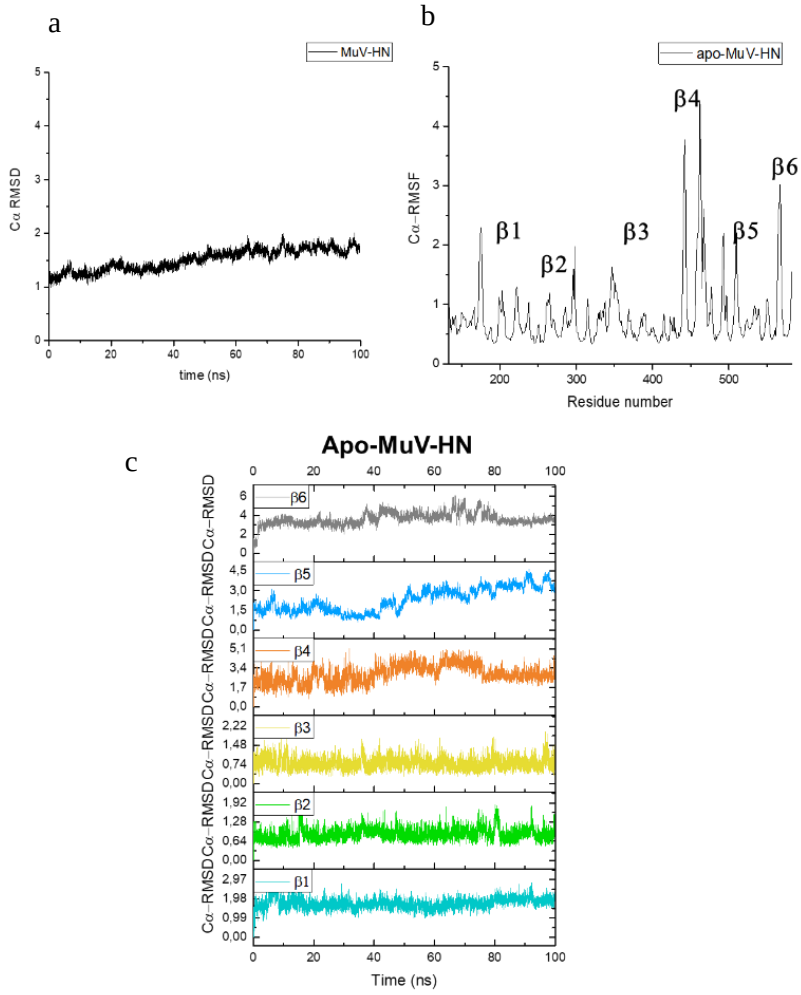


Figure 7.23 MD simulation analysis of MuV-HN homology model. a) Backbone RMSD of the protein along the trajectory. **b)** Per residue atomic fluctuation of the structure, calculated using the protein Cα atoms. The peaks in the RMSF plot corresponded to the mobile loops connecting the six ß-sheets and the β strands in each if the β-sheets .**c)** Backbone RMSD of the inter-sheet loops, namely ß1 (201-206), ß2 (260-265), ß3 (365-375), ß4 (438-446), ß5 (530-535), ß6 (560-570).

As result, MuV-HN showed a slight conformational rearrangement along the simulated time, with a small increase of the backbone RMSD which converged

to 1.5 Å (Figure 7.23a). Furthermore, the RMSFs relative to the Cα-atoms was monitored along the trajectory, denoting smaller values in the secondary structure regions, further confirming the homology model stability (Figure 7.23b). Interestingly, greater fluctuations occurred in the loop regions, especially at the positions corresponding to the β1 (Asn201-Ser207) and β4 (Thr438-Ser446) inter-strand loops. In particular, the latter showed the highest flexibility from the RMSF plot (Figure 7.23b). Moreover, the RMSD of the inter-sheets loops was calculated (Figure 7.23c), revealing a conformational rearrangement from the initial geometry that affected mostly the loops of β5 and, to lesser extent, the β6 sheets. Nonetheless, no relevant conformational changes were detected when calculating the RMSD of the β1 and β4 sheets' inter-strand loops along the trajectory (Figure 7.23c). The obtained homology model was subsequently employed for docking using AutoDock 4.2 program considering as ligand 3'SLN 3D coordinates. Based on the predicted binding energy and cluster population the best pose of 3'SLN/MuV-HN complex was selected (Figure 7.24) and it was submitted to 100 ns MD run using AMBER 18.

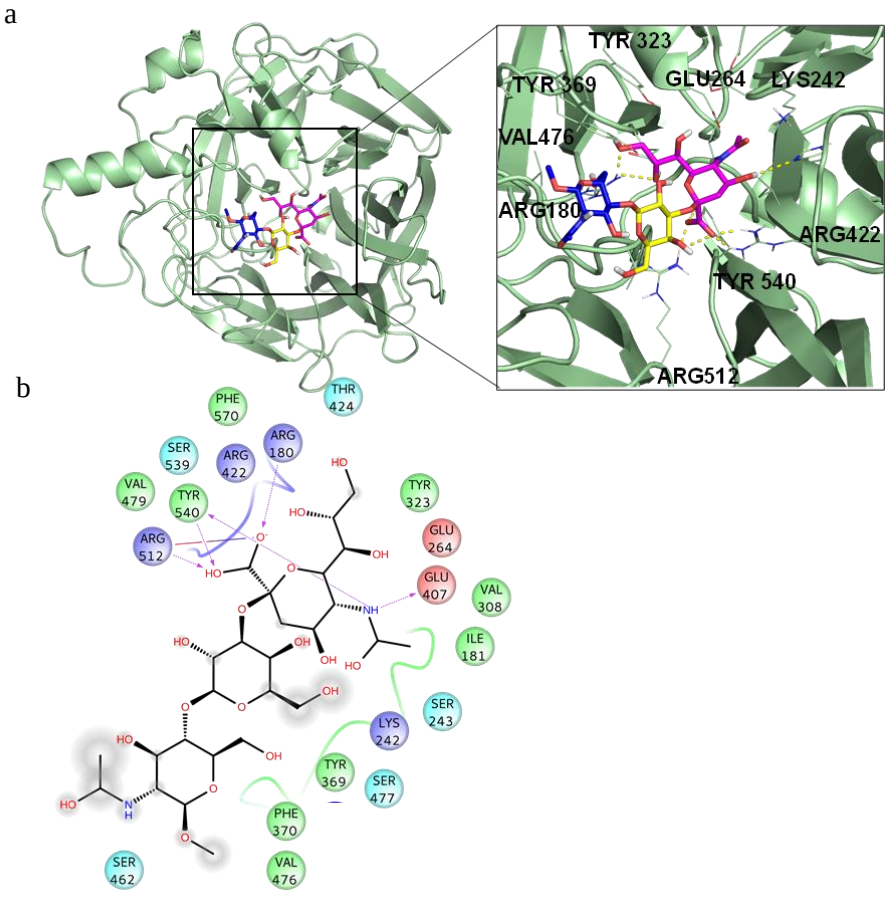


Figure 7.24 Interaction interface of MuV-HN and 3'SLN. a) 3D model of MuV-HN from SBL-1 in complex with 3'SLN. Close up view of 3'SLN binding pose at the MuV-HN active site. The main amino acid residues involved in the binding are highlighted as lines. Sialic acid, galactose and N-acetyl glucosamine residues are colored in magenta, yellow and blue, respectively according to the SNFG nomenclature. **b)** Two-dimensional plot illustrating the interactions of 3'SLN with the MuV-HN binding site residues. Dotted arrows indicate hydrogen bonds with functional groups from side chains and solid arrows those involving the backbone functional groups. The residues shown, close to the ligand, are involved into hydrophobic and polar interactions.

The stability of the complex was confirmed by the steady receptor and ligand RMSDs, and by the similarity between the most populated MD clusters.

Additionally, the analysis of the protein-ligand interactions demonstrated that the main contacts were preserved along the whole trajectory (Figure 7.25).

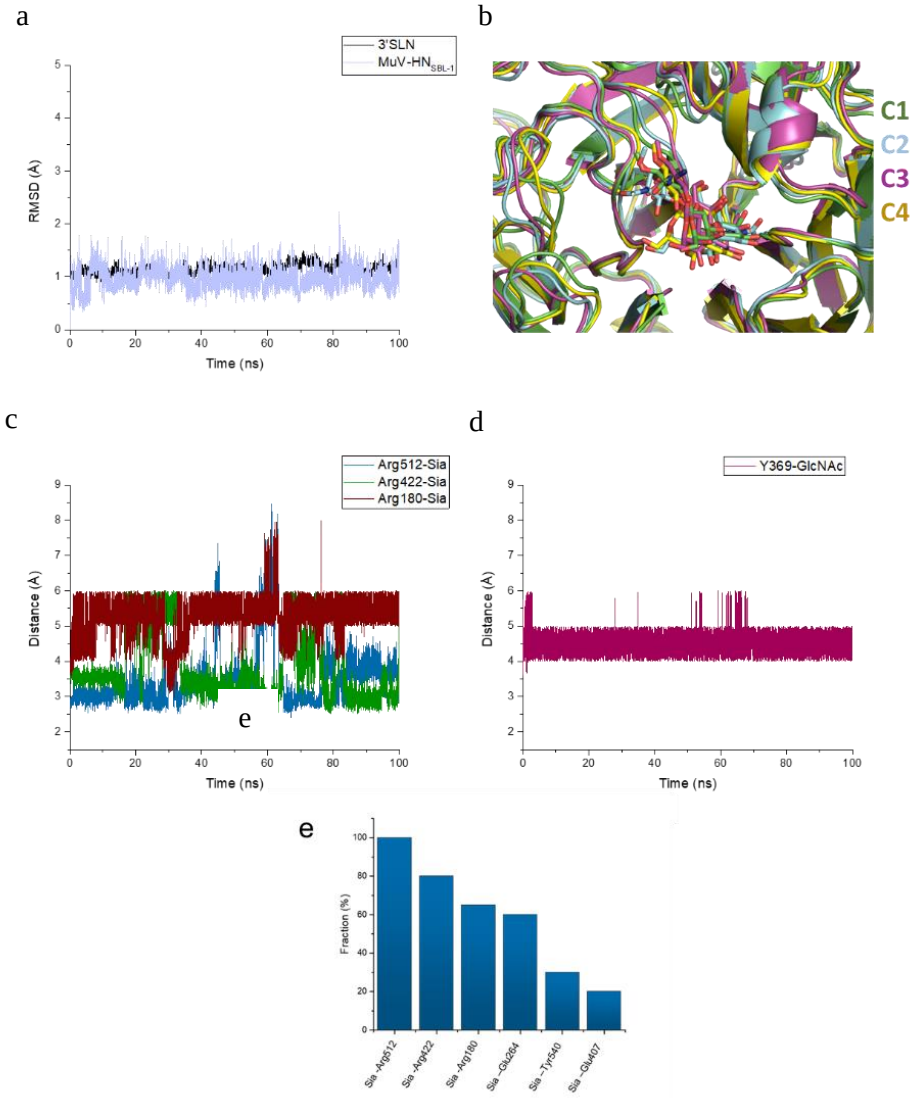


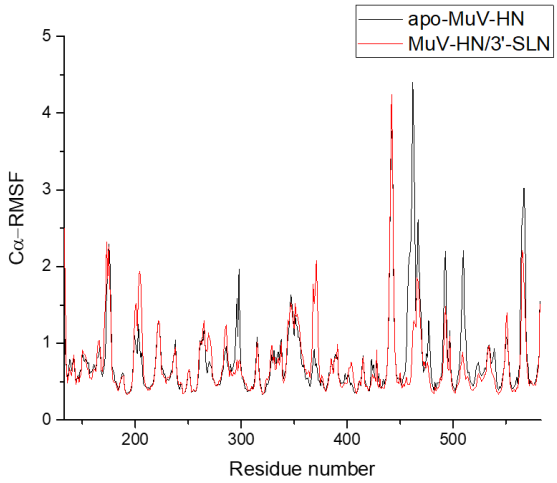
Figure 7.25 MD simulation analysis of MuV-HN/3'SLN complex. a) Protein and ligand backbone RMSD variation along the MD **b) S**uperimposition of the most populated clusters (C1-C4) derived according to k-mean algorithm estimation **c)** Distance between MuV-

HN triarginyl cluster: Arg180, Arg422 and Arg512 and Neu5Ac carboxylate. The average between the heavy atoms directly involved in the interaction has been considered **d)** Average distance between Tyr369 aromatic ring protons and H1-H3-H5 atoms of GlcNAc unit **e)** Frequency most representative protein/ligand distances.

In accordance with NMR data, computational analysis showed that all three sugar moieties were in contact with the receptor surface, as reported in Figure 7.25. The interaction pattern of the chosen binding model showed many similarities with that observed for MuV-HN/3'SLN complex from Hoshino strain.[341,342] In detail, the sialic acid unit formed strong interactions between the carboxylate moiety and three arginine residues of the receptor, namely Arg180, Arg422 and Arg512. The N-acetyl moiety formed the hydrogen bond with Glu407 and non-polar interactions with Ile181 and Val308. The GlcNAc sugar ring was engaged in CH-pi interactions with Tyr369. Moreover, the Gal residue was close to the receptor surface, but no significant polar interactions were observed, as its hydroxyl groups were directed toward the solvent. Even so, the Gal moiety was in proximity to the Val476 side chain thus entailing hydrophobic contacts with MuV-HN.

In order to verify if a particular loop stabilization occurred in the MuV-HN bound form, the RMSF and loop RMSD analysis of the 3'SLN/complex was performed. As a result (Figure 7.26), fewer fluctuations upon binding of the inter-strand loops of β2, β5 and β6 sheets were observed. Interestingly, a significant stabilization of the intra-strands loops connecting β2- β3 and β4- β5 sheets was also observed.

a



b

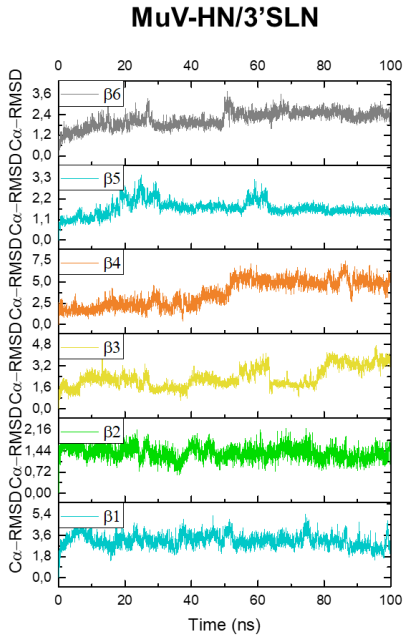
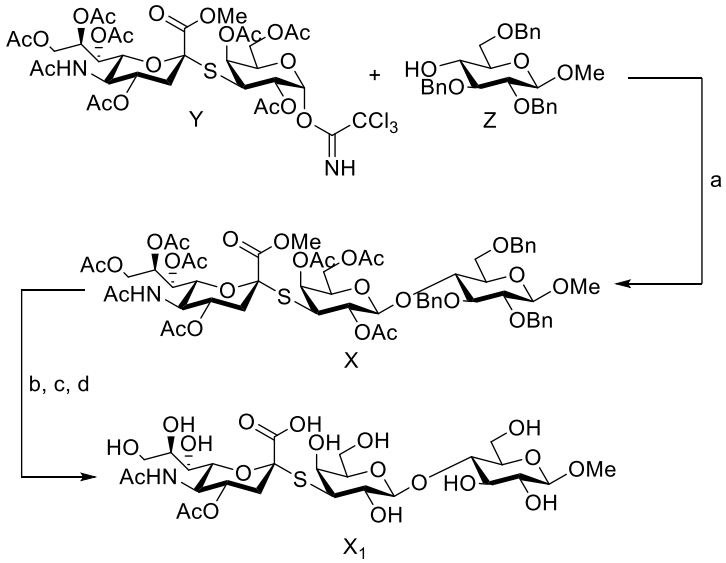


Figure 7.26 Flexibility of MuV-HN upon 3'SLN binding. a) Superimposition of the Ca RMSF plots of apo-MuV-HN (black) and MuV-HN/3'SLN complex (red) **b)** Backbone RMSD of MuV-HN inter-sheet loops in the complex, namely ß1 (201-206), ß2 (260-265), ß3 (365-375), ß4 (438-446), ß5 (530-535), ß6 (560-570).

Given that from NMR and computational analysis the entire trisaccharide seemed to be involved in the recognition and binding process, and being the terminal sialic acid moiety anchoring the glycan to the binding pocket of MuV-HN, a thio-derivative of 3'SLN, namely thio-3'SL (Scheme 7.1), synthesized by Dr. Oscar Francesconi group (University of Florence), and tested as a potential inhibitor of neuraminidase activity. Indeed, the incorporation of thio- functionality on the anomeric position of the sialic acid should indeed affect the lability of the Neu5Ac-Gal glycosidic linkage without hindering the recognition of the sialoglycan by the enzyme.



Scheme 7.1 Synthesis of thio-3'SL (compound **X1)**. Reagents and conditions: **a)** $BF_3{\cdot}Et_2O$, 3Å molecular sieves, dry CH_2Cl_2, 0°C, 3h; **b)** Pd/C, H_2, MeOH, room temperature, 24h; **c)** MeONa, MeOH, room temperature, 2h; **d)** NaOH 1M in H_2O, room temperature, 2h.

In detail, the thio-3'SL (**X1**) was prepared starting from methyl-2,3,6-tri-O-benzyl-β-D-glucopyranoside (**Z**)[359] and the thrichloroacetimidate **Y**,[360] both

prepared accordingly to literature methods. Glycosylation, catalyzed by $BF_3{\cdot}Et_2O$, gives the protected trisaccharide **X** with a 50% yield. Removal of benzyl protective groups by hydrogenation and of the acetylic groups with MeONa in MeOH was followed by hydrolysis of the methyl ester with aqueous NaOH to give the thiotrisaccharide **X1** with a 75% yield on three steps.

First, the hydrolytic stability of the thio-3'SL and its ability to interact with MuV-HN protein was examined by means of NMR spectroscopy and fluorescence analysis. Expectedly, the thio-3'SL was not hydrolyzed by MuV-HN, as it was detected by monitoring over time the NMR spectra of MuV-HN/thio-3'SL mixture (data not shown).

Though, the thio- sialoside was recognized by MuV-HN as demonstrated by fluorescence data. The binding interaction of thio-3'SL with MuV-HN from SBL-1 strain was indeed analyzed by fluorescence spectroscopy exploiting the presence of aromatic amino acids in the protein binding pocket. In detail, the changes in the fluorescence intensity of tryptophane residues of MuV-HN when adding increased amount of thio-3'SL were tracked (Figure 7.27).

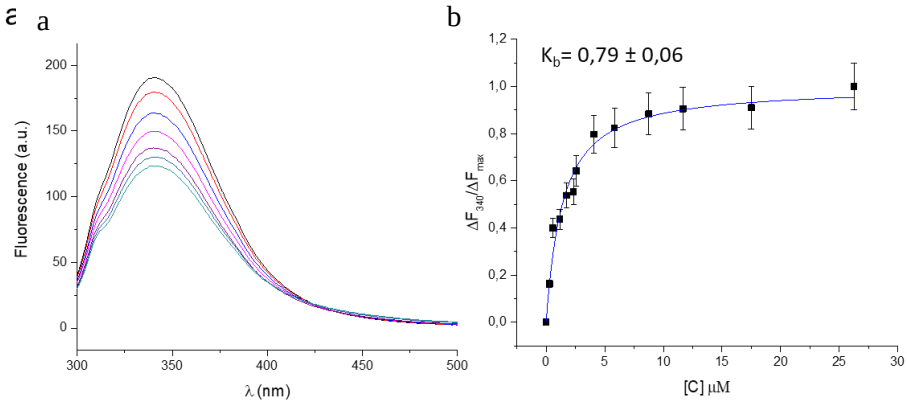


Figure 7.27 Binding analysis of MuV-HN/thio-3'SLN interaction by steady state fluorescence. a) Fluorescence analysis of MuV-HN/thio-3'SL interaction Fluorescence titration of MuV-HN upon the addition of thio-3'SL. **b)** The relative binding isotherm, the

value of the binding constant (K_b) is also reported. For each data point, 10% Y error bars are shown.

The interaction was quantitatively evaluated by fitting the resulting binding curve achieved with the application of the quenching data elaboration described by Ribeiro et al.[289]As a consequence, a binding constant (K_b) value of 0.79 ± 0.05 ($\mu M^{-1)}$ was obtained (Figure 7.27b).

Then, the molecular characterization of the complex has been then carried out by STD NMR and docking analysis. The ligand interacting epitope has been indeed described by STD NMR experiments (Figure 7.28).

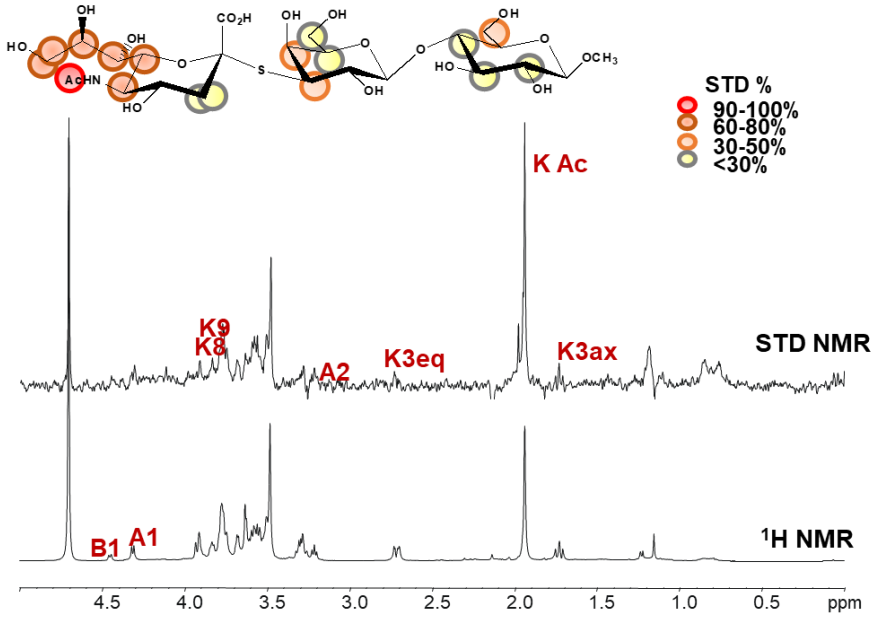
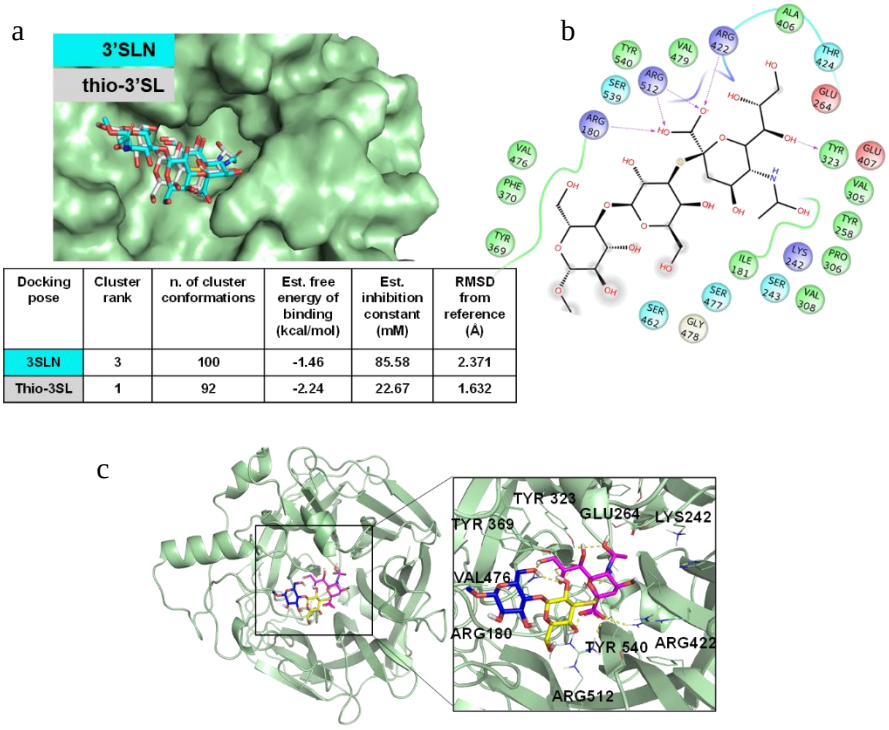


Figure 7.28 STD NMR analysis of thio- 3'SL in the interaction with MuV-HN from SBL-1 strain. (top panel) Interacting epitope map of thio 3'SL as derived by STD NMR data. (bottom panel) The 1H-NMR spectrum and the STD NMR spectrum of MuV-SBL-1–thio-3'SL mixture with a molecular ratio of 1:70, at 283 K.

As expected, the strongest STD enhancements were observed for the sialic acid moiety and especially for its acetyl group and protons of Neu5Ac lateral chain (H7-H8-H9). Significant STD effects were further detected for protons H4-H5 and H6 of sialic acid ring. Lower STD signals (below 50%) were instead observed for protons belonging to galactose and glucose residues indicating that they pointed farer from the protein surface. The structure of thio-3'SL was then docked into the binding site of MuV-HN from SBL-1 strain. The docking calculations allowed to describe the relevant receptor-ligand binding interactions, highlighting that the protein accommodated the synthetic analogue and the 3'SLN with a similar binding mode. Indeed, the superimposition of the two complexes (Figure 7.29) clearly showed a nearly identical orientation of the ligands into the receptor binding site



Docking pose	Cluster rank	n. of cluster conformations	Est. free energy of binding (kcal/mol)	Est. inhibition constant (mM)	RMSD from reference (Å)
3SLN	3	100	-1.46	85.58	2.371
Thio-3SL	1	92	-2.24	22.67	1.632

Figure 7.29 3D model of MuV-HN from SBL-1 in complex with thio-3'SL a) Docking of 3'SLN and thio-3'SLN into MuV-HN from SBL-1 strain structural model. The superimposition of the best selected binding mode of both ligands is shown. **b)** Close up view of thio-3'SL binding pose at the MuV-HN active site. The main amino acid residues involved in the binding are indicated as lines. Sialic acid, Galactose and N-acetyl Glucosamine residues are colored in magenta, yellow and blue, respectively according to the SNFG nomenclature. **c)** Two-dimensional plot showing the interactions of thio- 3'SL with the MuV-HN binding site residues. The residues shown are involved into hydrophobic and polar interactions with the ligand.

The interaction network was overall conserved; the Neu5Ac carboxylate engaged electrostatic interactions with the receptor Arg180, Arg422 and Arg512; Glu 407, Glu264, and Tyr323 binding pocket residues were also interacting with sialic acid

unit. Moreover, the Glc moiety of the trisaccharide formed hydrophobic interactions with Tyr369 and Val 476 residues; this latter residue was also in close contact with Gal unit.

7.4 Discussion

MuV is one of the primary causes of respiratory infections in humans, especially children. With the development of live attenuated vaccines, the prevalence of MuV infections has been radically reduced;[361] despite this, several outbreaks of MuV infection still occurred worldwide in the recent years, even in highly vaccinated populations.[362,363,364] The reasons of the vaccine resistance still need to be explained, although it is believed to be due either to the weaning of the immunity (secondary vaccine failure) or the elusion from the vaccine induced antibodies, allowed by antigenic variations between the vaccine and outbreak strains.[365,366,367]

Mumps virus possesses different surface glycoproteins and, among them, hemagglutinin-neuraminidase activity is essential for the infection and propagation of viruses belonging to the Paramyxoviridae family, including parainfluenza, the Newcastle disease and mumps viruses.

HN represents a multitasking protein mediating both the early and late stages of viral infection. Indeed, upon recognition of sialylated-glycan structures of host cell surface glycolipids and glycoproteins, particularly abundant in the mucus covering the respiratory epithelia, MuV-HN induces the structural change of the adjacent MuV-F, promotes the fusion activity and also serves to remove sialic acids from the virus-budding cell surface and progeny viral particles. This neuraminidase activity is crucial to avoid the viral tethering on cell surface, the nonspecific binding onto the cell matrix components and the viral self-

agglutination. Moreover, the HN protein represents the main antigen inducing neutralizing antibodies and plays key roles for the tropism of mump viruses. Thus, MuV-HN is an appealing target for the structure-based design of of novel antiviral drugs since interfering with these processes may affect the viral pathogenicity and infectivity, hence mitigating the diseases caused by mumps virus.

Over the past years, different types of inhibitors have been developed toward other viral neuraminidases, especially targeting hPIVs. However, the design and development of novel inhibitors and high-affinity ligands could prove worthwhile for a better understanding of virus tropism and pathogenesis and may help in the fight against viral infections allowing the advancement of new licensed antiviral drugs.

In this regard, the purpose of this study was to dissect, at molecular level and in solution state, the basis of sialoglycans recognition by MuV-HN, by establishing the proper glycan conformation selected upon binding and by describing its hydrolysis kinetic, as well as characterize a potential inhibitor of the neuraminidase activity.

Here, MuV-HN from Hoshino strain was characterized by means of NMR and computational studies. A detailed epitope mapping of different ligands was achieved by STD NMR, allowing the identification of the saccharide region making contact with MuV-HN binding pocket. In accordance with the defined crystal structure complex of MuV-HN with 3' sialyllactose previously reported[328,329] all the saccharide units of the sugar chain participated in the binding with MuV-HN, which was able to accommodate in its binding pocket the β-anomer of the sialic acid residue as well. Consequently, it was possible to affirm that MuV-HN was characterized by a considerably extended binding site compared with other neuraminidases.[368]

Then, tr-NOESY experiments complemented with docking and CORCEMA, allowed for the investigation the conformational behavior of sialoglycans when binding the MuV-HN, highlighting a conformer selection. Indeed, the results revealed the preference of the trisaccharide **1**/3'SLN for the *t* conformer, characterized by ϕ/ψ torsion angles of 180°/-11° around the Neu5Ac-Gal glycosidic linkage.

Docking calculations in the extended protein binding site established a network of interactions that defined the ability of MuV-HN to bind different sialoglycans, including the β-anomer of the sialic acid, but to accommodate also longer saccharide epitopes. Interestingly, the collected data revealed indeed that MuV-HN was able to recognize the complex glycan undecasaccaride **2**, displaying interactions comparable to those of MuV-HN/**1** complex, however, the steric hindrance of the saccharide chains, due to the presence of a branching of undecasaccharide **2**, enhanced the difficulties of the protein in the accommodation of the glycan-receptor. These data were further supported by kinetic studies that characterize the catalytic activity of MuV-HN by tracking the hydrolysis of the two sialoglycans in the presence of MuV-HN. The derived parameters, indeed, clearly showed a lower kinetic efficiency of the enzyme toward the substrate **2**.

Next, MuV-HN from SBL-1 strain was characterized by means of NMR and computational studies. The kinetic of trisaccharide **1**/3'SLN hydrolysis catalyzed by MuV-HN from SBL-1 strain was characterized by progress curve analysis derived by NMR experiments. Notably, the kinetic parameters obtained by the elaboration through the Lambert W function were in the same range of the values obtained for the same ligand interacting with MuV-HN from Hoshino strain, as shown in our previous work.[355] Moreover, STD NMR allowed to describe the ligand interacting epitope stressing the importance of the sialic acid in the

recognition and interaction process and that the binding pocket of the protein was able to accommodate the three composing sugars of the 3'SLN.

Then, homology modelling of the protein was also carried out with the aim to achieve a 3D view of the protein-ligand complex. The crystal structure of MuV-HN from the Hoshino strain was used for building the model, that revealed a huge similarity between the 3D structures. It is worth to note that, from the sequence analysis of the two proteins that belonged to different genotypes, few variable positions were observed. Few of the mutations affected the important functional regions, as the known B-cell epitopes, the most relevant was the 354 (P→Q) mutation at the 327–363 region, that modified the physiochemical features of the surface exposed area. This variation was already suggested to be significant for altered antibody recognition and neutralization by MuV-HN.[343] Nonetheless, the neuraminidase binding site was preserved, differently from that observed when comparing the sequences of other MuV-HN genotypes.[343]The refined 3D model of the protein has been used for docking analysis and MD simulations of the MuV-HN/3'SLN complex. The results demonstrated that the main interactions involving the three sugars of the glycan receptor observed when bound to MuV-HN from Hoshino strain were conserved in the binding with MuV-HN from SBL-1 strain. Indeed, the Neu5Ac residue was deeply located in the receptor pocket where it established strong polar interactions with the canonical neuraminidases sialic acid recognizing residues, including the Arginine triad (Arg180, Arg422, Arg512). Nonetheless, the previously described interactions involving Tyr369, which were reported to strongly influence the binding affinity of MuV-HN from Hoshino strain were also present in the case of MuV-HN from SBL-1 strain. Interestingly, the MD simulation studies performed on the apo and bound forms of MuV-HN revealed the stabilization of particular intra- and inter- sheets loops upon the

ligand binding, thus suggesting that the receptor loop flexibility plays a relevant role in the molecular recognition of sialoglycans by mumps hemagglutinin neuraminidase.

Moreover, the combination of synthetic chemistry, NMR, fluorescence and computational studies provided a 3D view of the MuV-HN from SBL-1 strain when interacting with the natural substrate, 3'SLN, and the synthetic thio-3'SL, showing a similar binding mode. The synthesis of a thio-trisaccharide, thio-3'SL, was carried out by Dr. Oscar Francesconi (University of Florence) and it was investigated as a potential inhibitor of the neuraminidase activity of the protein. Hence, the ability of thio-3'SL to interact with MuV-HN from SBL-1 strain was demonstrated by monitoring the intrinsic fluorescence of the Trp residues of the receptor and by NMR. Moreover, docking analysis provided a picture of the protein-ligand complex, showing that the thio-3'SL established contacts with MuV-HN receptor which were comparable to those made from 3'SLN, consistently to the similar bound ligands's topology. The main interactions were established by the Neu5Ac unit, that interacted with the characteristic receptor binding site amino acids Arg180, Arg422, Arg512, Glu407 and Tyr323. Accordingly, the third sugar exhibited stacking interactions with the Tyr369.

In conclusion, the present study advances the knowledge about the ability of MuV-HN to specifically recognize α(2,3)-linked sialic acid at the terminus of *N*-linked and *O*-linked glycans commonly found on cell-surface glycoproteins, as well as a synthetic sialoside. Overall, the analyses here performed demonstrated that, in both the investigated viral strains, the main interactions involved the three sugars of the glycan receptor were comparable. This study also revealed molecular details of the viral entry and release mechanism into/from host cells by biophysical characterization of glycan-receptors binding to the MuV-HN protein and their

hydrolysis reactions. These results may serve for the structure-based design and synthesis of sialic acid-derived compounds analogues acting as inhibitors of MuV-HN and may help in the identification of new inhibitor scaffolds which could prove worthwhile in the fight against Mumps.

Chapter 8-Concluding remarks

The glycocalyx is an essential element of living organisms, representing the host primary interface which discriminates between healthy or diseased cells as well as invading pathogens. In particular, the ubiquitous sialic acids, being exposed on mammalian cells surfaces are considered as effective markers of "self" and, upon interaction with specific sialic acid binding proteins act as key regulators of the immune system. Among sialic acid binding proteins, Siglecs are immune system transmembrane receptors able to recognize the sialic acid epitope presented by glycoconjugates on cell surfaces.

Since their discovery in the late 1980, an intense research activity aimed at dissecting Siglecs' roles in immunity and led to deeply rationalize the role played by each Siglec, dependent on their expression pattern, relative population, signalling properties, location and specific recognition of sialylated structures. These studies constitute a key step toward the comprehension of Siglecs relevance in both health and disease. Indeed, aberrant Siglec-sialylated ligands interactions have been associated to a broad spectrum of illnesses ranging from allergies, autoimmunity, neurodegeneration and cancer. Currently, Siglecs are considered as immune checkpoints in human diseases and the specificity and restricted cell-type expression make them suitable for the design of therapeutic agents for the treatment of such diseases.

In the last decades, many attempts have been made to develop antibodies- or glycan-based therapies targeting Siglecs. Nonetheless, several antibody-based therapies targeting Siglecs entered in clinical trials. Moreover, the therapeutic potential of high affinity and selective glycan-based ligands targeting Siglecs emerged from new pioneering studies that highlighted the potential of glycan

ligands for their impact in the modulation of cell signalling by controlled Siglecs recruitment to the appropriate receptor.[369]

Currently, as the understanding of the biological functions of Siglecs keep expanding, so the opportunities for modulating disease by targeting the Siglecs are increasing.

Thus, despite some encouraging results, still many efforts are required to design drug candidates for successful clinical trials. In the context of Siglecs glycomimetics development, a viable strategy is represented by the rational design of high affinity and high avidity Siglec ligands. To do this, the molecular basis of Siglec-sialoglycans interaction needs to be characterized for several Siglecs.

To this end, the approach based on the combination of ligand-based NMR techniques and molecular modelling has been extensively applied to the elucidation of various Siglec-sialoglycans interactions at the atomic level and have provided new insights for the understanding of the molecular recognition events underlying the functions of these systems of biological interest. A pluridisciplinary approach, in a collaborative work with different groups, has been carried out for the detailed study of Siglec-2 interactions with natural and synthetic ligands. The detailed characterization of complex N-glycans behavior when interacting with h-CD22 was carried out, demonstrating the ability of h-CD22 to equally bind both arms of biantennary glycans. Interestingly, the interacting surface of both antennae was almost invariant regardless of glycan investigated, indicating that this superficial recognition is not affected by inner residues. To further define the molecular recognition features of the biologically relevant B-cell inhibitory receptor, the effect of sialic acid glycolylation on the binding with CD22 was evaluated, comparing the behavior of different ligands in complex with murine and human CD22, improving the knowledge of the structural basis of the recognition of sialylated *N*-glycans by CD22 receptor. Indeed, a global vision of how the most diffuse

neuraminic acid forms of sialylated *N*-glycans in mammals are arranged in the binding pocket of CD22 was provided. The achieved results indicated that, despite the different nature of sialic acid residue, the recognition region of h-CD22 is almost invariant comparing Neu5Ac and Neu5Gc containing glycans.

The interaction of Siglec-10 with differently sialylated glycans was then probed. Although the crystal structure of Siglec-10 has not been solved yet, the orthogonal methods applied here allowed a molecular description of the recognition and binding processes driving Siglec-10-sialyloglycans complexes, providing unique information about dynamics and molecular interactions. The ligands' epitopes in their bound conformations were indeed profiled and consistent 3D models of the interactions were presented. Thus, the obtained outcomes have improved the knowledge of the molecular mechanisms occurring between Siglec-10 and sialylated glycans, providing a structural point of view for the design and development of high-affinity ligands, in their bioactive conformations, able to control the receptor functionality.

Siglec-sialylated pathogens interplay represents a pivotal yet underrated aspect of Siglecs function. Nonetheless, a deeper comprehension of the mechanisms occurring in the interaction with dangerous human pathogen could provide the solution to many diseases lacking a definitive cure. Here, a suitable structural model for the binding of the *F. nucleatum* sialylated O-antigen by Siglec-7 was provided. The studies allowed for the determination of the key structural features for the recognition of the *F. nucleatum* sialylated O-antigen in the interplay with Siglec-7, both important actors in tumorigenesis. The obtained information will be valuable in the development of therapies effective against the opportunistic pathogen diseases, as well as for the modulation of Siglec-7 activity.

Within this frame, the Thesis work significantly contributed to improve the comprehension of the structural and molecular basis for the glycans receptor

binding to Siglec-2, Siglec-7 and Siglec-10. Such achievements will be crucial for the rational design and synthesis of novel therapeutics based on high affinity sialosides with the above described applications in health and disease. Despite this, further structural and *in vivo* studies will be necessary to provide a complete depiction of the complex mechanisms at the basis of Siglecs modulation.

In the light of the achieved results, the development of selective sialosides probes, based on the atomistic description here provided, may allow to advance the current Siglec targeting strategies. Indeed, the application of glycan-based drug delivery vehicles may lead to increasing the sensitivity and sensibility toward changes in the availability of Siglec binding sites on different cells and under different conditions.[229]

Noteworthy, the orthogonal approach based on NMR and molecular modelling was applied to other side projects, such as to demonstrate that human C-type lectins are potential binders of bacterial LPS in solution. In particular, it was provided evidence that the human MGL is able to detect not only tumor-related carbohydrate structures present in mucin-like glycopeptides, but also glycan structures exposed on the bacterial surface. Moreover, it was employed to improve the knowledge about of hemagglutinin neuraminidase protein recognition of α-2,3-linked sialic acid at the terminus of *N*-linked and *O*-linked glycans commonly found on cell-surface glycoproteins. This study revealed molecular details of the viral entry and release mechanism into/from host cells by biophysical characterization of glycan-receptors binding to the viral receptor and their hydrolysis reactions.

SECTION IV- EXPERIMENTAL SECTION

Chapter 9 Experimental section

9.1 Materials

9.1.1 Recombinant expression of Siglecs

Siglec-2/CD22 (Section II-Chapter 3)

Both human and murine Siglec-2 were expressed by Prof. Paul Crocker's team at University of Dundee in UK in form of IgG Fc chimeras, fusion proteins comprising the Fc portion of human IgG and the extracellular Ig-like domains of CD22 and expressed and purified as follows:[373] A Chinese Hamster Ovary cell line stably expressing the 3 N-terminal Ig-like domains of human CD22 fused to the Fc region of mouse IgG2b and the 3 N-terminal Ig-like domains of mouse CD22 fused to the Fc region of human IgG2b, were kindly provided by Dr Ajit Varki (University of California San Diego). Cells were expanded in flasks in Hams F10 + 10% FCS and when confluent the medium was changed to Hams F10 + 0.5% FCS and cells incubated for up to 7 days. 4 litres of supernatants were pooled, 0.2 µM filtered and passed over a 1.0 ml protein A-Sepharose column. Following washing with 100 ml PBS, the CD22-Fc protein was eluted with 0.1 M glycine pH 3.0 and neutralised with one tenth volume of 1.0 M Tris buffer pH 8.0. Fractions containing > 0.1 mg/ml protein were pooled, dialysed against PBS and stored at -80C. Purity of the protein was assessed by SDS-PAGE under reducing and non-reducing conditions followed by Coomassie Blue staining.

Siglec-10 (Section II-Chapter 4)

<u>Expression of Siglec-10 Fc</u>

The cDNA encoding a recombinant form of human Siglec-10 containing Chinese Hamster Ovary cell line stably expressing the 5 3 N-terminal Ig-like domains of human Siglec-10 fused to the Fc region of human mouse IgG1G2b was

kindly provided by Dr Ajit Varki (University of California San Diego) and cloned into the pEE14 glutamine synthetase expression system by Prof. Paul D. Crocker team. Siglec-10 construct was then expressed by Prof. Angela Arciello team at University of Napoli Federico II according to the following procedure.[374]

A CHO cell clone stably expressing Siglec-10-Fc was expanded in glutamine-free Glasgow's modification of Eagles Medium (GMEM) containing 200 micromolar L-methionine sulfoximine and 10% FCS. Once cells were ~90% confluent, the medium was switched to GMEM + 0.5% Hyclone Fetalclone II (low IgG) and cells left for up to 7 days. Three Cells were expanded in flasks in Hams F10 + 10% FCS and when confluent the medium was changed to Hams F10 + 0.5% FCS and cells incubated for up to 7 days. 4 litres of supernatants were pooled, 0.2 µM filtered and passed over a 1.0 ml protein A-Sepharose column. Following washing with 100 ml PBS, the Siglec-10-Fc protein was eluted with 0.1 M glycine pH 3.0 and neutralised with one tenth volume of 1.0 M Tris buffer pH 8.0. Fractions containing > 0.1 mg/ml protein were pooled, dialysed against PBS and stored at -80C. Purity of the protein was assessed by SDS-PAGE under reducing and nonreducing conditions followed by Coomassie Blue staining.

Expression of Siglec-10 CRD in *E. Coli*

The nucleotide sequence encoding for human Siglec-10 CRD was obtained by NCBI database (acces.No. Q96LC7.3) (residues 18-140), and the sequence including C36S mutation was cloned into pET29b plasmid (Twist Bioscience). pET29b_Siglec-10_C36S plasmid was then transformed into Rosetta (DE3) and Origami 2 strains, streaked onto LB agar plates containing 15 g/ml of Kanamicin, 34 g/ml of chloramphenicol, and incubated at 37 ˚C over night. All liquid culture media were supplemented with 15 g/ml of Kanamicin and 34 g/ml of chloramphenicol. For expression of unlabeled Siglec -10 a single colony was inoculated into LB medium and grown overnight at 37 ˚C and 250 rpm. The LB

main cultures were inoculated and grown at 37 °C and 130 rpm until an OD_{600} of 0.6, and then induced by adding 1 mM IPTG and further grown at 15 °C for 38h. Cells were harvested by centrifugation at 7500 rpm for 15 min at 4 °C. Then, bacterial cell pellets were resuspended in ice cold buffer (20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4) additionally supplemented with Roche complete EDTA-free protease-inhibitor mix and subsequently lysed by sonication. Insoluble material was removed by centrifugation at 40000 rpm for 30 min at 4°C.

All cell pellets samples were analyzed by SDS-PAGE using precast 20% polyacrylamide gels (Bio-Rad) in denaturing conditions according to the following procedure: Samples were mixed with 1/5 volume 5 x Laemmli sample buffer (125 mM tris, pH 6.8; 6% glycerol, 2% SDS; 5% beta-mercaptoethanol; 0.025% bromophenol blue) followed by boiling at 95°C for 5 min. After cooling, samples were separated on 20% SDS-PAGE at 200 V for 30 min in Tris/Glicine/SDS running buffer. Protein bands were visualized using Coomassie Blue staining.

For the gateway cloning technology to obtain Siglec-10 CRD fused to HGB1, MBP and GST tags, the principles and protocol of this technology described elsewhere.[375]

The original plasmid encoding for Siglec-10 C36S CRD were amplified by PCR using the suitable primers and inserted into the pENTR/TEV/D-TOPO vector by a TOPO cloning. Subsequently, the LR clonase reaction was performed using different pDEST vectors (e.g. pETG-30A: GST), pDEST-HisMBP: MBP and pTH34:HGB1), all the resulting plasmid showed high level of expression of the protein constructs in BL21(DE3) Gold strains.

9.1.2 Recombinant expression of other proteins

Expression of MGL-ECD (Section III-Chapter 6)

MuV-HN proteins were expressed by Prof. Franck Fieschi team at University Grenoble des Alpes according to the following methodologies.[376] MGL amino acids 61 to 292 corresponding to extracellular domain (MGL-ECD) were cloned downstream to StrepTagII, Factor Xa protease cleavage site and 3 glycines in between Nde I and Xho I restriction site of pET-30 multiple cloning site region. MGL-ECD was expressed in E. coli BL21(DE3) in 1 L of LB medium supplemented with 50 µg/mL kanamycin at 37°C. Expression was induced for 3 h by addition of 1 mM isopropyl 1-thio-D-galactopyranoside when the culture had reached an OD_{600} of 0.8. The protein was expressed in the cytoplasm as inclusion bodies. Cells were pelleted by a 20 min centrifugation at 4000 g at 4°C. The pellet was re-suspended in 30 mL of buffer 150 mM NaCl, 25 mM Tris-HCl pH 8, 4 mM CaCl2 (calcium buffer) and one anti-protease tablet (Complete EDTA free, Roche), using a Potter-Elvehjem. Bacteria were disrupted by sonication and cell debris were eliminated by centrifugation at 100000 g for 30 min at 4°C. The pellet was successively washed in 30 mL of buffer 2 M Urea, 150 mM NaCl, 25mM Tris-HCl pH 8 and 1 % triton X-100, centrifuged at 100000 g, washed in 30 mL of calcium buffer, centrifuged at 100000 g. Inclusion bodies were finally solubilized in 30 mL of buffer 6 M guanidine, 150 mM NaCl, 25 mM tris-HCl and 0,01 % β- mercaptoethanol (guanidine buffer) and insoluble material was removed by a centrifugation of 30 min at 100000 g at 4°C. Based on calculated MGL-ECD $\varepsilon(280)=69440\ M^{-1cm-1}$, the protein concentration of supernatant was adjusted to 2 mg/mL with guanidine buffer. The refolding of the protein was performed by a drop-by-drop dilution in a buffer 150 mM NaCl, 100 mM Tris-HCl pH 8 and 25 mM CaCl2. Product of refolding was dialyzed 3 times against calcium buffer 25

mM Tris-HCl pH 8, 150 mM NaCl and 4 mM CaCl2. Insoluble material was eliminated by a final centrifugation step at 100000 g for 1 h at 4°C. The protein was purified using a two steps protocol on GalNAc-Agarose affinity column (Sigma), eluted with buffer 150 mM NaCl, 25 mM tris-HCl and 5 mM EDTA. The eluted sample is loaded onto a Toyopearl HW-50S gel filtration column (Tosoh Bioscience), equilibrated with the calcium buffer. After analysis by SDS-PAGE gel fractions containing MGL-ECD were pooled and concentrated to 5.6 mg/mL by ultrafiltration using a Vivaspin 20 PES 10 kDa MWCO.

Expression of MuV-HN from Hoshino and SBL-1 strains (Section III-Chapter 7)

MuV-HN proteins were expressed by Prof. Takao Hachiguchi team at University of Kyushu according to the following procedure.[377]

The DNA fragments encoding HN proteins (amino acid positions 96–582) were amplified by PCR from the template plasmids of the MuV Hoshino strain or the SBL-1 strain. They were individually cloned into the expression vector pHLsec containing the N-terminal secretion signal sequence and the C-terminal His_6-tag sequence. These expression plasmids were transiently transfected into 80% confluent HEK293S cells lacking *N*-acetylglucosaminyltransferase I [293S GnTI(-) cells][378] using polyethyleneimine-MAX (Polysciences) and incubated at 37 °C and 5% CO_2. At 6 d post-transfection, the supernatant containing the secreted MuV-HN was harvested and centrifuged to eliminate cell components. The supernatant was incubated at 4 °C overnight and filtered to further eliminate insoluble components. MuV-HNs were purified using Ni^{2+}-NTA affinity column (COSMOGEL His-Accept; Nacalai Tesque) in purification buffer (50 mM NaH_2PO_4, 150 mM NaCl, 10 mM imidazole, pH 8.0) and then eluted with elution buffer (50 mM NaH_2PO_4, 150 mM NaCl, 500 mM imidazole, pH 8.0). The eluted MuV-HNs were further purified using a size exclusion column (Superdex 200 Increase GL 10/300; GE Healthcare) in PBS without potassium. MuV-HNs were

concentrated and adjusted to 1.5 mg/mL using Amicon Ultra centrifugal filters (Merck Millipore).

9.1.3 Extraction of the LPS from *F. nucleatum* strain 10953 (Section II-Chapter 5)

The LPS extraction was performed by laboratory colleague Ferran Nieto-Fabregat. *F. nucleatum* ATTC 10953 cell pellets (provided by Prof. Nathalie Juge) were harvested by centrifugation, lyophilized, and extracted by the hot phenol/water method as previously described.[379] Each phase was dialyzed against distilled water to remove the phenol, freeze-dried, and analyzed by 12% sodium dodecyl sulphate SDS-PAGE. After the water/phenol extraction, LPSs from strains 10953 was detected in the water phase by silver nitrate staining, The phases containing the LPSs were purified by enzymatic digestion (DNAse, RNAse and proteinase K) as previously described,[48]ollowed by centrifugation at 6,000 rpm, for 30 min at 4 °C and ultracentrifugation at 30,000 rpm for 4 h at 4°C. In order to perform the STD NMR experiment a mild acid hydrolysis was performed with acetic acid 1% (100°C, 2-3 h) in order to separate the O-antigen (OPS) and the lipid A portion. A partial depolymerization of the O- antigen was also performed.

The polysaccharide part present in the supernatant was further purified by size exclusion chromatography (78.6 mL volume, 78.6 mL/h flow, NH_4HCO_3 50mM) using a Sephacryl S200 column. The eluate was monitored by a refractive index detector (Knauer GmbH – WellChrom Differential Refractometer K-2301) and fractions were pooled accordingly.to isolate the O-antigen.

9.1.5 Extraction of the LOS from *E. coli* OSR1(Section III-Chapter 6)

The LOS extraction was performed by laboratory collegue Meriem Maalej. Bacterial E. coli R1 and R3 strains were grown, starting from glycerol stock, in LB broth medium at 37°C and the cells were harvested by centrifugation.

The freeze-dried cells were extracted according to the petroleum ether–chloroform–phenol (PCP) as described.[380]For each strain, the LOS fraction was precipitated and analyzcharacteìed through DOC-polyacrylamide gel electrophoresis 13,5% followed by silver nitrate staining[381] highlighting the presence of lipooligosaccharide material. An aliquot of LOS was treated with anhydrous hydrazine (1 ml for 20 mg of sample), stirred at 37°C for 90 min, cooled, poured into ice-cold acetone, and allowed to precipitate. The precipitate was then centrifuged, washed twice with ice-cold acetone, dried, and dissolved in water and lyophilized.[382] The O-deacylated sample was subsequently N-deacylated with a strong alkaline treatment (KOH 4 M, 120 °C). After desalting using a column of Sephadex G-10, the resulting oligosaccharide fraction was further separated utilizing size exclusion chromatography on Biogel P2, from which oligosaccharide OS was obtained.[383]

9.2 NMR methods

9.2.1 Free ligand spectroscopic characterization

All the systems were dissolved in 50 mM PBS/D_2O buffer, pH 7.4. All of the spectra were recorded at 298 K and/or 283K. The complete assignment of the ^{1}H and ^{13}C resonances, for each free ligand analyzed in this Thesis, was achieved combining the information from 1D-NMR and 2D-NMR DQF-COSY, TOCSY, NOESY, t-ROESY and ^{1}H-^{13}C HSQC experiments. All NMR spectra were recorded on a Bruker 600 MHz DRX spectrometer equipped with a triple resonance probe head (^{1}H, ^{13}C, and ^{15}N). The complete assignment of the ^{1}H and ^{13}C resonances, for each free ligand analyzed in this Thesis, was achieved combining the information from 1D-NMR and 2D-NMR DQF-COSY, TOCSY, NOESY, t-ROESY and ^{1}H-^{13}C HSQC experiments. All NMR spectra were recorded on a Bruker 600 MHz DRX spectrometer equipped with a triple resonance probe head

(^{1}H, ^{13}C, and ^{15}N) with a cryo probe. All spectrum were recorded in deuterate phosphate saline buffer (PBS/H2O) The spectra were calibrated with [D4](trimethylsilyl)propionic acid, sodium salt (TSP; 0.05mM) as an internal reference. ROESY and NOESY spectra were measured using data sets (t1xt2) of 4096x256 points with mixing times between 100 ms and 500 ms. Double quantum filtered phase sensitive COSY experiments were performed using data sets of 4096x512 points; total correlation spectroscopy experiments (TOCSY) were performed with a spinlock time of 100 ms, using data sets (t1 xt2) of 4096x256 points. In all homonuclear experiments the data matrix was zero-filled in the F1 dimension to give a matrix of 4096x2048 points and was resolution enhanced in both dimensions by a cosine-bell function before Fourier transformation. Coupling constants were determined on a first- order basis from 2D phase sensitive double quantum filtered correlation spectroscopy (DQF-COSY) Heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond correlation (HMBC) experiments were measured in the ^{1}H-detected mode via single quantum coherence with proton decoupling in the ^{13}C domain, using data sets of 2048x256 points. Experiments were carried out in the phase-sensitive mode. A 60 ms delay was used for the evolution of long-range connectivity in the HMBC experiment. In all heteronuclear experiments the data matrix was extended to 2048x1024 points using forward linear prediction extrapolation.

9.2.2 Bound ligand spectroscopic characterization

All the systems were dissolved in 50 mM PBS/D_2O buffer, pH 7.4. All of the spectra were recorded at 298 K and/or 283K.

For the bound ligands, STD and TR-NOE experiments were performed as following. TR-NOESY spectra were recorded by using protein/ligand molar ratios of 1:5→1:25 with mixing times of 150–600 ms. The concentration of the protein was

10-20μM. For the acquisition of STD NMR spectra, the protein/ligand molar ratio varied from 1:20 to 1:150, and saturation times between 0.5 and 5 s were used. A T1ρ filter (spin-lock pulse: 50 ms) to eliminate the unwanted broad resonance signals of the protein was used. The on-resonance frequency was set at different ppm values according to the analyzed system, since it has to fall in a region in which ligand signals must be absent, necessary condition to avoid false-positive results in the STD spectrum. The off-resonance frequency was always maintained at δ=40 ppm, where no protein or ligand signals resonated. For protein saturation, a train of 40 Gaussian shaped pulses each of 50 ms (delay between pulses: 1 ms, field strength: 21 Hz) was employed. The saturated and reference spectra were acquired simultaneously by creating a pseudo-2D experiment. Reference experiments on free ligands, in the same conditions, were carried out to assure the absence of direct irradiation of the ligand. STD spectra were performed with 16 k and 32 k data points. The original free induction decay (FID) was zero-filled to 64 k, and Fourier transformation with use of an exponential window function was applied (lb = 1–2 Hz). The 2D STD NMR spectrum 2D STD TOCSY (Section II-Chapter 3) was recorded by using data sets of 2048_256 points, and the same parameters used for the STD and TOCSY/HSQC spectra were employed. The STD NMR required the acquisition of an additional STD spectrum of the receptor in absence of ligands, in the same conditions, that is then subtracted from the STD spectrum acquires in the presence of ligands. To determine the magnitude of the STD effects, the intensity of the signals in the STD NMR spectra were compared with the signal intensities of a reference spectrum (off-resonance). The STD signal with the highest intensity was set to 100% and the other signals were normalized to this peak. In order to refine these qualitative results, STD build up curves were derived (Section II-Chapter 3).[150,151]The spectra acquired at different saturation times were fitted to a monoexponential equation: $STD = STDmax\ (1 - exp^{-(ksat\ t)})$,

where STD is the STD intensity of a given proton at a saturation time t, STDmax is the asymptotic maximum of the curve, and ksat stands for the observed saturation rate constant. The epitope map of each ligand was obtained by normalizing all the values of different protons ligand to the largest STDfit, giving STDepitopes fit. Data acquisition and processing were performed with TOPSPIN software.

9.2.3 NMR kinetics (Section III- Chapter 7)

For the analysis of the enzyme kinetic of HN, a 1:50 p/l molar ratio was used. The enzyme and the substrate were dissolved in 50 mM PBS/D_2O buffer, pH 7.4. All of the spectra were recorded at 298 K. Before the addition MuV-HN, a 1D proton spectrum with the application of composite pulses to carry out water presaturation (zgppr) was acquired with 32 transients (2 min experiment duration), to obtain the intensities corresponding to t=0. Then, HN was added to the NMR tube and the solution was mixed quickly. The sample was then reintroduced into the magnet and after applying a short shimming routine, ^1H NMR spectra using the aforementioned acquisition parameters were recorded at different time points for several hours. A timer was set to measure the delay between the addition of HN and the collection of the first NMR spectrum of the mixture. The delay time was incorporated into the previous calculations. FIDs were multiplied with an exponential function prior to Fourier transformation. All well dispersed resonances of substrate and product were integrated at each time point. The concentrations each time point was determined with respect to the integrals of the H1 and $H3_{eq}$ resonances of trisaccharide **1** and undecasaccharide **2** at t=0.The progress curves for the transformation of the substrates were fitted in Sigma Plot with the equation adapted from reference.[384]

9.3 Molecular mechanics and molecular dynamics simulations

The determination of the conformation and dynamics of glycans, in free and bound state, can be performed by using molecular modelling techniques supported by NMR spectroscopy data. The individual monosaccharides show a limited flexibility and within an oligosaccharide, they are assumed as rigid ring structures, so the conformation of a glycan can be established just defining the behavior of each glycosidic linkage through molecular mechanics and dynamic simulation.

9.3.1 Molecular mechanics simulations

MM studies were carried out using Maestro software. Since the conformation of a saccharidic chain is mainly defined by the relative orientation of the sugar moieties, i.e. by Φ and Ψ torsion angles, the first step in the conformational analysis was to build the potential energy surfaces for each disaccharide connected by a glycosidic linkage; Φ represents the torsion angle about H1- C1-O-CX' whereas Ψ about C1-O-CX'-HX'. Molecular mechanics calculations were performed using the MM3* force field as included in MacroModel. The bulk of the solvent was simulated using a constant dielectric of 80. For each disaccharide both Φ and Ψ dihedral angles were varied incrementally using a grid step of 18 degrees. The corresponding flexible maps were drawn as 2D contours plots using the graphical tools of MacroModel tool.

Molecular mechanics provided a first estimation of the conformational regions energetically accessible. The resulting adiabatic energy maps indicated global and local minima.

9.3.2 Molecular dynamics simulations

Molecular dynamics simulations are typically carried out in four steps under isotermic isobaric conditions.

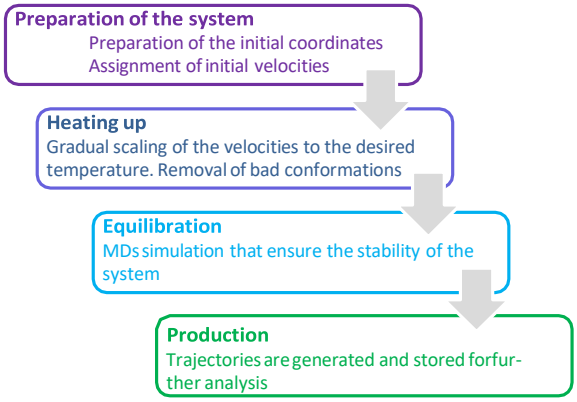


Figure 9.1 General protocol for running MDs simulations. In the first step, the system, derived from NMR, X-ray or homology modelling, is prepared, adding missing atoms, and submitted to minimization cycles. The second step consists of heating the system in order to remove bad contacts, increasing and assigning new atom velocities. The initial velocities are generally determined from the standard temperature-dependent Maxwell-Boltzmann distribution. The third step is called equilibration, in which energy, temperature and RMSD (Root Mean Square Deviation) of the system converge to stable values that allow the collection of trajectories (production step) that will be successively analyzed.

Molecular dynamic calculations were performed in explicit water using the Glycam06j-1 for the glycans, FF14SB for the proteins and Gaff for organic moieties (Section II-Chapter III) force fieds within the AMBER 14 (Section II-Chapter 3) and AMBER 18 (Section II, Chapter 4-7) software packages.

All the oligosaccharides were built up and minimized by using Maestro package and the carbohydrate builder utility of the glycam website (www.glycam.com), in order to have a file .pdb with the appropriate atom types as requested by the AMBER software and then the torsional angles were set to the values obtained through the molecular mechanics calculations. For the protein preparation, missing

hydrogen atoms were added, and protonation state of ionizable groups and cap termini were computed by using Maestro Protein Preparation Wizard.

For the preparation of STnThra ligand (Section-II Chapter 3) the xleap, antechamber and parmchk2 modules were used to parametrize the molecule. A similar approach was used to build the D-FucNAcN residue of the hexasaccharide and the dodecasaccharide from the O- antigen of *F. Nucleatum* (Section II- Chapter 5).

The input files, were generated using the tleap (.prmtop and .inpcrd) modules of the software package, all the minimization were performed using Sander module and molecular dynamic calculations were performed using the PMEMD module. The corresponding molecules were immersed within an octahedral box of water (TIP3P) of the proper size and the remote interactions were calculated using a cut off of 10 angstroms and counterions were added to neutralize the system. After the preparation of the input files, a energy minimization process was performed to refine the initial structure. 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. The system was minimized, at first, holding the solute fixed, while a second minimization was performed on the entire system. Furthermore, the whole system was slowly heated from 0 to 300 K using a weak restrain on the solute and then, the system was equilibrated at 300 K using constant pressure and removing the restrains on the solute. The system coordinates were saved and used for the 100ns simulations using the PMEMD module implemented in AMBER. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 10000 structures for each complex, which were finally analyzed. The free binding energy along the simulations was calculated by means of the MM-PBSA and MM-GBSA methods (Section II-Chapter 3). The stability of energy, pressure, temperature and other thermodynamic parameters

were monitored along the trajectory and then, RMSD, RMSF, torsions, distances and hydrogen bonds were extracted. Ptraj was the utility used for analyzing and processing trajectories and coordinate files created from the MD simulations. VMD was used to analyze the trajectory.

9.3.3 Homology modelling

m-CD22 (Section II-Chapter 3)

The sequence encoding for m-CD22 (acces.No. NP_033975.3) was obtained from NCBI (http://www.ncbi.nlm.nih.gov). For computational 3D structure calculation by homology modelling, the extracellular V-set, and C2 set domains of murine CD22 were considered. The sequence interval corresponding to the extracellular portion was aligned to hCD22 template (PDB ID: 5VKM), using BLAST homology model was generated by means of SWISS-MODEL.

Siglec-10 (Section II-Chapter 4)

The carbohydrate recognition domain of human Siglec-10 was considered for computational 3D structure calculation by homology modelling. The sequence encoding for the human Siglec-10 (acces.No. Q96LC7.3) was extracted from NCBI database. The carbohydrate recognition domain was identified by submitting the query sequence to CD-search.[385].s result, the region between the residues 23-140 of the protein query was found to match ('specific hit') to the Immunoglobulin (Ig) domain at the N terminus of Siglec (acces.No. cd05712). The previous interval was then submitted to the BLAST search, using protein data bank as database to identify potential templates for the homology modelling of Siglec-10. Siglec-8 (PDB ID:2HRL), Siglec-7 (PDB ID:1NKO), Siglec-5 (PDB ID:2ZG2) and Siglec-3 (PDB ID:5IHB) were selected as the templates because of the highest identity (47-56%) and query cover (85-99%). Homology models were generated using four different servers: SWISS-MODEL, I-TASSER, PHYRE2, and

RAPTOR-X. Then, the best models were submitted to 100ns molecular dynamics simulations to optimize the geometry and study the stability.

MGL (Section III-Chapter 6)

The homology modelling of the MGL CRD was built by using asialoglycoprotein receptor (PDB ID: 1DV8) as template (see Supplementary Material and Methods). The model retained the overall protein fold of the template, with a CRD constituted by core region of β– strands and by two α-helices located on the core sides. The structure displayed three calcium ions coordinated by the protein residues, in three different calcium binding sites. The putative sugar-binding site, which was centred around the second calcium ion was formed by Gln243, Asp245, Glu256, Asn268, Asp269. These residues were also involved in the coordination of the calcium ion. The top edge of the binding was constituted by a curved loop that contained the conserved Trp247 residue, which represents a feature shared by most the Gal-specific proteins characterized so far.

MuV -HN (Section III-Chapter 7)

The sequence encoding for MuV-HN (UniProtKB: P19762-1) was obtained from Uniprot database (http://www.uniprot.org). For computational 3D structure calculation by homology modelling, the 3D coordinates of MuV-HN head domain (PDB ID: 5B2C), were used as template. The sequence of the target was aligned to the template using BLAST and the target-template alignment was submitted to generate the homology model by means of SWISS-MODEL. Then, the obtained structure was optimized geometry optimization and to evaluate the stability of the model.

9.3.4 Docking

Docking calculations were performed with AutoDock4. For the docking calculations, Gasteiger charges of protein and ligands were assigned by

AutoDockTools. With AutoDock4, the Lamarckian genetic algorithm was used to sample different conformations of the ligands, by randomly changing all the torsion angles and overall orientation of the molecule. A three-dimensional grid was defined centred on an equidistant point to the side chains of key binding residues of the target protein, and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock were also used. Docking calculations and analysis of all compounds were performed by AutoDock 4.2.2. The grid point spacing was centered among residues of the binding site residues of the proteins. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100, and 250000 energy evaluations. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

9.3.5 CORCEMA-ST

The protons saturation of the ligand was predicted using CORCEMA-ST program and subsequently compared to the corresponding experimental results through the NOE R factor. The pdb coordinates of the analyzed complexes were generated from docking calculations. The representative complexes structures selected for each system were submitted to the CORCEMA-ST protocol. The input parameters employed for the calculations according to the experimental conditions were saturation time of 2 s, protein concentration and ligand concentrations inputted according to the experimentally used to record STD experiments. A generalized order parameter S2 of 0.85 and a uniform leakage relaxation of 0.30 s^{-1} were postulated. As the dissociation constant (K_D) was set in the range of 10^{-4}–10^{-6} M and modified further to achieve the best fit. The direct saturation of the aromatic residues His, Trp, Phe and Thr protons was assumed, as the STD

irradiation frequency was 6.5 ppm. The conformation of the ligand was considered to be the same in both free and bound state. The k_{on} was set to 10^{-8} L mol^{-1} s^{-1}, in the assumption of a diffusion controlled kinetic model; a correlation time, τ_r, of the ligand in the free and bound state was set at 2 ns and $3*10^{-8}$ s was estimated through an empirical approximation.

9.4 Ligand binding assays

9.4.1 Surface Plasmon Resonance (Section II-Chapter 3)

SPR experiments were performed by Prof. Bruno Pagano group, using a Biacore X100 system (GE Healthcare) at 25°C. The sample cell of a CM5 sensor chip, equilibrated in HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% surfactant P20, pH 7.4), was activated with the h-CD22 protein (40 µg/mL in 10 mM sodium acetate, pH 5.0) at a flow rate of 10 µL/min using standard amine coupling (N-ethyl-N-(dimethylaminopropyl)-carbodiimide (EDC) and N-hydroxysuccinimide (NHS)). h-CD22 protein was immobilized at 2600 response units (RU). Unreacted activated groups were blocked by injection of 1 M ethanolamine at 10 µL/min. Trisaccharide **1** was injected as analyte at various concentrations (from 0.031 to 0.5 mM). Injections were performed with a flow rate of 30 µL/min using PBS at pH 7.4 as running buffer. Contact time of 120 s and dissociation time of 600 s were used. No regeneration after each sample was necessary. The response from the reference surface (ethanolamine derivatized) was subtracted from the sample flow cell derivatized with h-CD22 to correct for refractive index changes, nonspecific binding, and instrument drift. To determine dissociation constants (K_D), the equilibrium binding data were analyzed by non-linear curve fitting using the Biacore Evaluation software provided with the device.

9.4.2 AlphaScreen (Section II-Chapter III)

The Alpha Screen was performed by laboratory colleague Cristina Di Carluccio. The experimental conditions were optimized in preliminary experiments, keeping the concentrations of beads constant (20 μg/mL final concentration of each bead) and varying only the concentration of the tagged protein and the tracer in a 40 μL (final assay volume) reaction. Excitation of donor beads was at 680nM. Samples were measured at 520-620 nm in EnspireTM Alpha (Perkin-Elmer). Once chosen the concentration of the tagged protein and the tracer, the displacement assay was performed by using as buffer assay PBS containing 0.5% BSA. Final concentrations of 0.3 nM and 3 nM of h-CD22 and tracer were used respectively. Fc-tagged human CD22 protein (10 μL/well) was mixed with 6'SLN-PAA-biotin (10 μL/well, GLycoNz), and a final concentration range (1, 3, 10, 30, 100, 300 μM) of glycans (10μL/well). After incubation 1h at 25 °C, anti-mouse IgG Acceptor beads (10 μL/well, final concentration 20 μg/mL) in buffer assay were added to the plate and incubated for 60 min. Finally, Streptavidin Donor beads (10 μL/well, final concentration 20 μg/mL) in buffer assay were added to the plate. The alpha counts were normalized according to the following equation: Y = (X – background) * 100 / (Z – background), X = [alpha counts in the presence of each ligand], Z = [alpha counts in the absence of ligands], background = [alpha counts in the absence of ligands and 6'-SLN-PAA-biotin] and the dose-response curves were obtained by fitting the data with the Log(inhibitor)vs normalized response – variable slope formula using GraphPad Prism. Each data point represents the mean: SD of triplicate measurements.

9.4.3 Binding analysis by ELISA (Section II-Chapter 4)

ELISA assay was performed by Dr. Fabrizio Chiodo (CNR Pozzuoli, Italy), according to the following procedure: 50 μL of synthetic polyacrylamide

(PAA) polymers functionalized with Neu5Acα2-3Galβ1-4Glcβ-$NHCOCH_2NH_2$ or Neu5Acα2-6Galβ1-4Glcβ-$OCH_2CH_2NH_2$ (purchased from GlycoNZ, Auckland, New Zealand, carrying the same moles of carbohydrate) at 200 µg/mL, or a solution of 50 µL of synthetic glyco-peptides containing sialylated *N*-glycans at 1mg/mL, in PBS (10mM, pH=7.4), were used to coat the Nunc MaxiSorp plate 2h at room temperature. After discarding and washing with PBS (1x150µL), the wells were blocked with 100 µL of 1% BSA (Sigma-Aldrich, lyophilized powder, ≥96%, agarose gel electrophoresis) in PBS at room temperature for 30 min. The blocking solution was discarded and 50µL of Fc-Siglec-10 at 1 µg/mL (Recombinant Human Siglec-10 Fc Chimera Protein from R&D) in assay buffer (0.5% BSA) were added to the wells. After 1h at 37°C, the wells were washed with PBS (2x150µL) and then 100 µL of anti-human horseradish peroxidase (0.8 µg/mL, life technologies, Novex® Goat anti-Human IgG-HRP) were added. After 30 min at room temperature, the wells were washed with PBS (2x150µL). Finally, 100 µL of substrate solution (3,3′,5,5′-Tetramethylbenzidine, TMB, in citric/acetate buffer, pH=4, and H_2O_2) were added and after 5 min incubation at room temperature the reaction was stopped with 50 µL of H_2SO_4 (0.8 M) and the optical density was measured at 450 nm in an ELISA reader. The experiment was performed in duplicate and data were normalized over the signal at 450 nm from the BSA-containing wells.

9.4.4 Binding analysis by fluorescence spectroscopy (Section II- Chapters 3,4,5 and Section III- Chapter 7)

Steady-state fluorescence spectroscopy have been acquired on a Fluoromax-4 spectrofluorometer (Horiba, Edison, NJ, USA) at the fixed temperature of 5°C. Emission spectra were recorded in the emission range of 290–500 nm upon excitation at 280 nm. The slit widths were fixed at 4 nm for the excitation and 10

nm for the emission wavelength. A quartz cuvette with a path length of 1 cm and a chamber volume of 1 mL was used under constant stirring. A 0.9 mL volume of proteins solution at fixed concentration of 0.25 μM was titrated by adding small aliquots (1–20 μL of a ligand stock solution of 500 μM) of ligands. The PBS buffer at pH 7.4 was used for all solutions. The optical density of the solution at the excitation wavelength was kept less than 0.05. Small errors due to dilution upon ligand additions were neglected. The protein fluorescence was found to quench in the presence of the ligands. The binding curve was obtained by plotting ΔF/F0 values versus ligand concentration. Data analysis were performed using the following equation:

$$\frac{\Delta I_f}{I_0} = \frac{\Delta I_{\max}}{I_0} X_{FY}$$

where ΔI_f is the fluorescence intensity change upon addition of the ligand and ΔI_{max} the maximal fluorescence intensity change.

For the fluorophore, *F*, and the interference specie, *Y*, associate to form a complex, the system is described by the association constant, *Kb*, and considering *Xf* the molar fraction due to total concentration of *F* in solution, it can be written as:

$$X_{FY} = \frac{\vdash b \pm \sqrt{b^2 - 4ac}}{2a}$$

Were a = [F]tKb, b = 1 + Kb[Y]t + Kb[F]t and c = Kb[Y]t

10 References

1 Varki, A., & Gagneux, P. (2017). Biological Functions of Glycans. In A. Varki (Eds.) et. al., Essentials of Glycobiology. (3rd ed., pp. 77–88). Cold Spring Harbor Laboratory Press.

2 Zhou, J. Y., & Cobb, B. A. (2021). Glycans in Immunologic Health and Disease. Annual review of immunology, 39, 511–536.

3 Zeng, W., Ford, K. L., Bacic, A., & Heazlewood, J. L. (2018). N-linked Glycan Micro-heterogeneity in Glycoproteins of Arabidopsis. Molecular & cellular proteomics : MCP, 17(3), 413–421.

4 Gimeno, A., Valverde, P., Ardá, A., & Jiménez-Barbero, J. (2020). Glycan structures and their interactions with proteins. A NMR view. Current opinion in structural biology, 62, 22–30.

5 Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. Nature reviews. Cancer, 15(9), 540–555.

6 Sasisekharan, R., Raman, R., & Prabhakar, V. (2006). Glycomics approach to structure-function relationships of glycosaminoglycans. Annual review of biomedical engineering, 8, 181–231.

7 Michaud, M., Prinz, W. A., & Jouhet, J. (2017). Glycerolipid synthesis and lipid trafficking in plant mitochondria.

8 Yu, R. K., Suzuki, Y., & Yanagisawa, M. (2010). Membrane glycolipids in stem cells. FEBS letters, 584(9), 1694–1699.

9 Di Carluccio, C., Forgione, R.E., Molinaro, A., Crocker, P., Marchetti, R., and Silipo, A. (2021). Exploring the fascinating world of sialoglycans in the interplay with Siglecs. Carbohydr. Chem. 44, 31–55.

10 Möckl L. (2020). The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. Frontiers in cell and developmental biology, 8, 253.

11 Fisher, P., Thomas-Oates, J., Wood, A. J., & Ungar, D. (2019). The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. Frontiers in cell and developmental biology, 7, 157.

12 Sprovieri, P., & Martino, G. (2018). The role of the carbohydrates in plasmatic membrane. Physiological research, 67(1), 1–11.

13 Oswald, D. M., & Cobb, B. A. (2018). Emerging glycobiology tools: A renaissance in accessibility. Cellular immunology, 333, 2–8.

14 Du, H., Clarke, A. E., & Bacic, A. (1996). Arabinogalactan-proteins: a class of extracellular matrix proteoglycans involved in plant growth and development. Trends in cell biology, 6(11), 411–414.
15 Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. Cell, 126(5), 855–867. 9.
16 Gagneux, P., Aebi, M., & Varki, A. (2017). Evolution of Glycan Diversity. In A. Varki (Eds.) et. al., Essentials of Glycobiology. (3rd ed., pp. 253–264). Cold Spring Harbor Laboratory Press.
17 Gaudette, S., Hughes, D., & Boller, M. (2020). The endothelial glycocalyx: Structure and function in health and critical illness. Journal of veterinary emergency and critical care (San Antonio, Tex.: 2001), 30(2), 117–134.
18 Filloux, A., & Whitfield, C. (2016). Editorial: The many wonders of the bacterial cell surface. FEMS microbiology reviews, 40(2), 161–163.
19 Geoghegan, I., Steinberg, G., & Gurr, S. (2017). The Role of the Fungal Cell Wall in the Infection of Plants. Trends in microbiology, 25(12), 957–967.
20 Sticklen, M. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat Rev Genet **9,** 433–443 (2008).
21 Marth, J. D., & Grewal, P. K. (2008). Mammalian glycosylation in immunity. Nature reviews. Immunology, 8(11), 874–887.
22 Rillahan, C. D., & Paulson, J. C. (2011). Glycan microarrays for decoding the glycome. Annual review of biochemistry, 80, 797–823.
23 Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A., Woods, R. J., & SNFG Discussion Group (2019). Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology, 29(9), 620–624.
24 Purcell, S. C., & Godula, K. (2019). Synthetic glycoscapes: addressing the structural and functional complexity of the glycocalyx. Interface focus, 9(2), 20180080.
25 Varki A. (2017). Biological roles of glycans. Glycobiology, 27(1), 3–49.
26 Cohen, M., & Varki, A. (2014). Modulation of glycan recognition by clustered saccharide patches. International review of cell and molecular biology, 308, 75–125.
27 Reitsma, S., Slaaf, D. W., Vink, H., van Zandvoort, M. A., & oude Egbrink, M. G. (2007). The endothelial glycocalyx: composition, functions, and visualization. Pflugers Archiv : European journal of physiology, 454(3), 345–359.

28 Hart, G. W., & Copeland, R. J. (2010). Glycomics hits the big time. Cell, 143(5), 672–676.
29 Dwek R. A. (1996). Glycobiology: Toward Understanding the Function of Sugars. Chemical reviews, 96(2), 683–720.
30 Latousakis, D., & Juge, N. (2018). How Sweet Are Our Gut Beneficial Bacteria? A Focus on Protein Glycosylation in Lactobacillus. International journal of molecular sciences, 19(1), 136.
31 Defaus, S., Gupta, P., Andreu, D., & Gutiérrez-Gallego, R. (2014). Mammalian protein glycosylation--structure versus function. The Analyst, 139(12), 2944–2967.
32 Pomin, V. H., & Mulloy, B. (2018). Glycosaminoglycans and Proteoglycans. Pharmaceuticals (Basel, Switzerland), 11(1), 27.
33 Iozzo, R. V., & Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. Matrix biology: journal of the International Society for Matrix Biology, 42, 11–55.
34 Banks D. D. (2015). Nonspecific shielding of unfavorable electrostatic intramolecular interactions in the erythropoietin native-state increase conformational stability and limit non-native aggregation. Protein science : a publication of the Protein Society, 24(5), 803–811.
35 Chen, M. M., Bartlett, A. I., Nerenberg, P. S., Friel, C. T., Hackenberger, C. P., Stultz, C. M., Radford, S. E., & Imperiali, B. (2010). Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. Proceedings of the National Academy of Sciences of the United States of America, 107(52), 22528–22533.
36 Benyair, R., Ogen-Shtern, N., & Lederkremer, G. Z. (2015). Glycan regulation of ER-associated degradation through compartmentalization. Seminars in cell & developmental biology, 41, 99–109.
37 Xu, C., & Ng, D. T. (2015). Glycosylation-directed quality control of protein folding. Nature reviews. Molecular cell biology, 16(12), 742–752.
38 Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., & Tabak, L. A. (2012). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology, 22(6), 736–756.
39 Qu, D., Wang, G., Yu, L., Tian, F., Chen, W., & Zhai, Q. (2021). The effects of diet and gut microbiota on the regulation of intestinal mucin glycosylation. Carbohydrate polymers, 258, 117651.

40 Suzuki, M., Angata, K., Nakayama, J., & Fukuda, M. (2003). Polysialic acid and mucin type o-glycans on the neural cell adhesion molecule differentially regulate myoblast fusion. The Journal of biological chemistry, 278(49), 49459–49468.
41 Angata, K., Long, J. M., Bukalo, O., Lee, W., Dityatev, A., Wynshaw-Boris, A., Schachner, M., Fukuda, M., & Marth, J. D. (2004). Sialyltransferase ST8Sia-II assembles a subset of polysialic acid that directs hippocampal axonal targeting and promotes fear behavior. The Journal of biological chemistry, 279(31), 32603–32613.
42 Varshney, S., & Stanley, P. (2018). Multiple roles for O-glycans in Notch signalling. FEBS letters, 592(23), 3819–3834.
43 Marchetti, R., Forgione, R. E., Fabregat, F. N., Di Carluccio, C., Molinaro, A., & Silipo, A. (2021). Solving the structural puzzle of bacterial glycome. Current opinion in structural biology, 68, 74–83.
44 Park, D., Xu, G., Barboza, M., Shah, I. M., Wong, M., Raybould, H., Mills, D. A., & Lebrilla, C. B. (2017). Enterocyte glycosylation is responsive to changes in extracellular conditions: implications for membrane functions. Glycobiology, 27(9), 847–860.
45 Bojar, D., Powers, R. K., Camacho, D. M., & Collins, J. J. (2021). Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions. Cell host & microbe, 29(1), 132–144.e3.
46 Janeway, C. A., Jr, & Medzhitov, R. (2002). Innate immune recognition. Annual review of immunology, 20, 197–216.
47 Mogensen T. H. (2009). Pathogen recognition and inflammatory signaling in innate immune defenses. Clinical microbiology reviews, 22(2), 240–273.
48 Di Lorenzo F., De Castro C., Lanzetta R, Parrilli M., Silipo A., Antonio M. (2015) Carbohydrates in Drug Design and Discovery, Chapter 3 Lipopolysaccharides as Microbe-associated MolecularPatterns: A Structural Perspective. 3-7, 54-56.
49 Trinchieri, G., & Sher, A. (2007). Cooperation of Toll-like receptor signals in innate immune defence. Nature reviews. Immunology, 7(3), 179–190.
50 Kogelberg, H., & Feizi, T. (2001). New structural insights into lectin-type proteins of the immune system. Current opinion in structural biology, 11(5), 635–643.

51 Mathews, R. J., Sprakes, M. B., & McDermott, M. F. (2008). NOD-like receptors and inflammation. Arthritis research & therapy, 10(6), 228.
52 Caruso, R., Warner, N., Inohara, N., & Núñez, G. (2014). NOD1 and NOD2: signaling, host defense, and inflammatory disease. Immunity, 41(6), 898–908. 0
53 Roh, J. S., & Sohn, D. H. (2018). Damage-Associated Molecular Patterns in Inflammatory Diseases. Immune network, 18(4), e27.
54 Schaefer L. (2014). Complexity of danger: the diverse nature of damage-associated molecular patterns. The Journal of biological chemistry, 289(51), 35237–35245.
55 Sangiuliano, B., Pérez, N. M., Moreira, D. F., & Belizário, J. E. (2014). Cell death-associated molecular-pattern molecules: inflammatory signaling and control. Mediators of inflammation, 2014, 821043.
56 Varki A. (2011). Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. Glycobiology, 21(9), 1121–1124.
57 Andreassen, P. R., Trappetti, C., Minhas, V., Nielsen, F. D., Pakula, K., Paton, J. C., & Jørgensen, M. G. (2020). Host-glycan metabolism is regulated by a species-conserved two-component system in Streptococcus pneumoniae. PLoS pathogens, 16(3), e1008332.
58 White, B. A., Lamed, R., Bayer, E. A., & Flint, H. J. (2014). Biomass utilization by gut microbiomes. Annual review of microbiology, 68, 279–296.
59Alsmark, C., Foster, P. G., Sicheritz-Ponten, T., Nakjang, S., Martin Embley, T., & Hirt, R. P. (2013). Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome biology, 14(2), R19.
60 Rabinovich, G. A., van Kooyk, Y., & Cobb, B. A. (2012). Glycobiology of immune responses. Annals of the New York Academy of Sciences, 1253, 1–15.
61 Varki, A., Etzler, M. E., Cummings, R. D., & Esko, J. D. (2009). Discovery and Classification of Glycan-Binding Proteins. In A. Varki (Eds.) et. al., Essentials of Glycobiology. (2nd ed.). Cold Spring Harbor Laboratory Press.
62 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.). (2015). Essentials of Glycobiology. (3rd ed.). Cold Spring Harbor Laboratory Press.

63 Varki, A., Schnaar, R. L., & Crocker, P. R. (2017). I-Type Lectins. In A. Varki (Eds.) et. al., Essentials of Glycobiology. (3rd ed., pp. 453–467). Cold Spring Harbor Laboratory Press.
64 Johannes, L., Wunder, C. & Shafaq-Zadah, M. (2016). Glycolipids and Lectins in Endocytic Uptake Processes. J. Mol. Biol. 428, 4792–4818.
65 Vasta, G. R., Ahmed, H., Nita-Lazar, M., Banerjee, A., Pasek, M., Shridhar, S., Guha, P., & Fernández-Robledo, J. A. (2012). Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. Frontiers in immunology, 3, 199.
66 McEver R. P. (2015). Selectins: initiators of leucocyte adhesion and signalling at the vascular wall. Cardiovascular research, 107(3), 331–339.
67 Chang, Y. C., Olson, J., Louie, A., Crocker, P. R., Varki, A., & Nizet, V. (2014). Role of macrophage sialoadhesin in host defense against the sialylated pathogen group B Streptococcus. Journal of molecular medicine (Berlin, Germany), 92(9), 951–959.
68 Schnaar R. L. (2016). Glycobiology simplified: diverse roles of glycan recognition in inflammation. Journal of leukocyte biology, 99(6), 825–838.
69 Thompson A. J., de Vries R.P., Paulson J.C. (2019) .Virus recognition of glycan receptors. Curr. Opin. Virol.34:117–29.
70 Ilver, D., Arnqvist, A., Ogren, J., Frick, I. M., Kersulyte, D., Incecik, E. T., Berg, D. E., Covacci, A., Engstrand, L., & Borén, T. (1998). Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science (New York, N.Y.), 279(5349), 373–377.
71 Mahdavi, J., Sondén, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K. A., Altraja, S., Wadström, T., Kersulyte, D., Berg, D. E., Dubois, A., Petersson, C., Magnusson, K. E., Norberg, T., Lindh, F., Lundskog, B. B., … Borén, T. (2002). Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. Science (New York, N.Y.), 297(5581), 573–578.
72 Dwek R. A. (1996). Glycobiology: Toward Understanding the Function of Sugars. Chemical reviews, 96(2), 683–720.
73 Varki, A., Schnaar, R. L., & Schauer, R. (2017). Sialic Acids and Other Nonulosonic Acids. In A. Varki (Eds.) et. al., Essentials of Glycobiology. (3rd ed., pp. 179–195). Cold Spring Harbor Laboratory Press.

74 Chang, Y. C., & Nizet, V. (2014). The interplay between Siglecs and sialylated pathogens. Glycobiology, 24(9), 818–825.
75 Mahajan, V. S., & Pillai, S. (2016). Sialic acids and autoimmune disease. Immunological reviews, 269(1), 145–161.
76 Bergfeld, A. K., Pearce, O. M., Diaz, S. L., Pham, T., & Varki, A. (2012). Metabolism of vertebrate amino sugars with N-glycolyl groups: elucidating the intracellular fate of the non-human sialic acid N-glycolylneuraminic acid. The Journal of biological chemistry, 287(34), 28865–28881.
77 Kawanishi, K., Dhar, C., Do, R., Varki, N., Gordts, P., & Varki, A. (2019). Human species-specific loss of CMP-N-acetylneuraminic acid hydroxylase enhances atherosclerosis via intrinsic and extrinsic mechanisms. Proceedings of the National Academy of Sciences of the United States of America, 116(32), 16036–16045.
78 Okerblom, J., & Varki, A. (2017). Biochemical, Cellular, Physiological, and Pathological Consequences of Human Loss of N-Glycolylneuraminic Acid. Chembiochem : a European journal of chemical biology, 18(13), 1155–1171.
79 Khan, N., de Manuel, M., Peyregne, S., Do, R., Prufer, K., Marques-Bonet, T., Varki, N., Gagneux, P., & Varki, A. (2020). Multiple Genomic Events Altering Hominin SIGLEC Biology and Innate Immunity Predated the Common Ancestor of Humans and Archaic Hominins. Genome biology and evolution, 12(7), 1040–1050.
80 Hedlund, M., Padler-Karavani, V., Varki, N. M., & Varki, A. (2008). Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. Proceedings of the National Academy of Sciences of the United States of America, 105(48), 18936–18941.
81 Malykh, Y. N., Schauer, R., & Shaw, L. (2001). N-Glycolylneuraminic acid in human tumours. Biochimie, 83(7), 623–634.
82 Varki A. (2017). Are humans prone to autoimmunity? Implications from evolutionary changes in hominin sialic acid biology. Journal of autoimmunity, 83, 134–142.
83 Soulillou, J. P., & Padler-Karavani, V. (2020). Editorial: Human Antibodies Against the Dietary Non-human Neu5Gc-Carrying Glycans in Normal and Pathologic States. Frontiers in immunology, 11, 1589.

84 Duan, S., & Paulson, J. C. (2020). Siglecs as Immune Cell Checkpoints in Disease. Annual review of immunology, 38, 365–395.
85 Varki A. (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature, 446(7139), 1023–1029.
86 Angata, T., & Varki, A. (2002). Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chemical reviews, 102(2), 439–469.
87 Schauer, R., & Kamerling, J. P. (2018). Exploration of the Sialic Acid World. Advances in carbohydrate chemistry and biochemistry, 75, 1–213.
88 Crocker, P. R., Paulson, J. C., & Varki, A. (2007). Siglecs and their roles in the immune system. Nature reviews. Immunology, 7(4), 255–2666.
89 Varki, A., & Angata, T. (2006). Siglecs--the major subfamily of I-type lectins. Glycobiology, 16(1), 1R–27R.
90 Macauley, M. S., Crocker, P. R., & Paulson, J. C. (2014). Siglec-mediated regulation of immune cell function in disease. Nature reviews. Immunology, 14(10), 653–666.
91 Crocker P. R. (2005). Siglecs in innate immunity. Current opinion in pharmacology, 5(4), 431–437.
92 Crocker, P. R., Clark, E. A., Filbin, M., Gordon, S., Jones, Y., Kehrl, J. H., Kelm, S., Le Douarin, N., Powell, L., Roder, J., Schnaar, R. L., Sgroi, D. C., Stamenkovic, K., Schauer, R., Schachner, M., van den Berg, T. K., van der Merwe, P. A., Watt, S. M., & Varki, A. (1998). Siglecs: a family of sialic-acid binding lectins. Glycobiology, 8(2), v.
93 Pillai, S., Netravali, I. A., Cariappa, A., & Mattoo, H. (2012). Siglecs and immune regulation. Annual review of immunology, 30, 357–392.
94 Crocker, P. R., & Varki, A. (2001). Siglecs in the immune system. Immunology, 103(2), 137–145.
95 Bornhöfft, K. F., Goldammer, T., Rebl, A., & Galuska, S. P. (2018). Siglecs: A journey through the evolution of sialic acid-binding immunoglobulin-type lectins. Developmental and comparative immunology, 86, 219–231.
96 Angata, T., Margulies, E. H., Green, E. D., & Varki, A. (2004). Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proceedings of the National Academy of Sciences of the United States of America, 101(36), 13251–13256.

97 Ravetch, J. V., & Lanier, L. L. (2000). Immune inhibitory receptors. Science (New York, N.Y.), 290(5489), 84–89.
98 Ulyanova, T., Shah, D. D., & Thomas, M. L. (2001). Molecular cloning of MIS, a myeloid inhibitory siglec, that binds protein-tyrosine phosphatases SHP-1 and SHP-2. The Journal of biological chemistry, 276(17), 14451–14458.
99 Avril, T., Freeman, S. D., Attrill, H., Clarke, R. G., & Crocker, P. R. (2005). Siglec-5 (CD170) can mediate inhibitory signaling in the absence of immunoreceptor tyrosine-based inhibitory motif phosphorylation. The Journal of biological chemistry, 280(20), 19843–19851.
100 Yu, Z., Maoui, M., Wu, L., Banville, D., & Shen, S. (2001). mSiglec-E, a novel mouse CD33-related siglec (sialic acid-binding immunoglobulin-like lectin) that recruits Src homology 2 (SH2)-domain-containing protein tyrosine phosphatases SHP-1 and SHP-2. The Biochemical journal, 353(Pt 3), 483–492.
101 Linnartz, B., Wang, Y., & Neumann, H. (2010). Microglial immunoreceptor tyrosine-based activation and inhibition motif signaling in neuroinflammation. International journal of Alzheimer's disease, 2010, 587463.
102 Arase, H., & Lanier, L. L. (2004). Specific recognition of virus-infected cells by paired NK receptors. Reviews in medical virology, 14(2), 83–93.
103 Quarles R. H. (2007). Myelin-associated glycoprotein (MAG): past, present and beyond. Journal of neurochemistry, 100(6), 1431–1448.
104 O'Neill, A. S., van den Berg, T. K., & Mullen, G. E. (2013). Sialoadhesin - a macrophage-restricted marker of immunoregulation and inflammation. Immunology, 138(3), 198–207.
105 Freeman, S. D., Kelm, S., Barber, E. K., & Crocker, P. R. (1995). Characterization of CD33 as a new member of the sialoadhesin family of cellular interaction molecules. Blood, 85(8), 2005–2012.
106 May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., & Jones, E. Y. (1998). Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 A resolution. Molecular cell, 1(5), 719–728.
107 Attrill, H., Imamura, A., Sharma, R. S., Kiso, M., Crocker, P. R., & van Aalten, D. M. (2006). Siglec-7 undergoes a major conformational change when complexed with the alpha(2,8)-disialylganglioside GT1b. The Journal of biological chemistry, 281(43), 32774–32783.
108 Lübbers, J., Rodríguez, E., & van Kooyk, Y. (2018). Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions. Frontiers in immunology, 9, 2807.

109 Hartnell, A., Steel, J., Turley, H., Jones, M., Jackson, D. G., & Crocker, P. R. (2001). Characterization of human sialoadhesin, a sialic acid binding receptor expressed by resident and inflammatory macrophage populations. Blood, 97(1), 288–296.
110 Khatua, B., Roy, S., & Mandal, C. (2013). Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens. The Indian journal of medical research, 138(5), 648–662.
111 Paulson, J. C., Macauley, M. S., & Kawasaki, N. (2012). Siglecs as sensors of self in innate and adaptive immune responses. Annals of the New York Academy of Sciences, 1253(1), 37–48.
112 Nitschke, L., & Tsubata, T. (2004). Molecular interactions regulate BCR signal inhibition by CD22 and CD72. Trends in immunology, 25(10), 543–550.
113 Nitschke L. (2014). CD22 and Siglec-G regulate inhibition of B-cell signaling by sialic acid ligand binding and control B-cell tolerance. Glycobiology, 24(9), 807–817.
114 Bochner, B. S., & Zimmermann, N. (2015). Role of siglecs and related glycan-binding proteins in immune responses and immunoregulation. The Journal of allergy and clinical immunology, 135(3), 598–608.
115 Khatua, B., Ghoshal, A., Bhattacharya, K., Mandal, C., Saha, B., Crocker, P. R., & Mandal, C. (2010). Sialic acids acquired by Pseudomonas aeruginosa are involved in reduced complement deposition and siglec mediated host-cell recognition. FEBS letters, 584(3), 555–561.
116 Bax, M., Kuijf, M. L., Heikema, A. P., van Rijs, W., Bruijns, S. C., García-Vallejo, J. J., Crocker, P. R., Jacobs, B. C., van Vliet, S. J., & van Kooyk, Y. (2011). Campylobacter jejuni lipooligosaccharides modulate dendritic cell-mediated T cell polarization in a sialic acid linkage-dependent manner. Infection and immunity, 79(7), 2681–2689.
117 Chang, Y. C., Uchiyama, S., Varki, A., & Nizet, V. (2012). Leukocyte inflammatory responses provoked by pneumococcal sialidase. mBio, 3(1), e00220-11.
118 Stephenson, H. N., Mills, D. C., Jones, H., Milioris, E., Copland, A., Dorrell, N., Wren, B. W., Crocker, P. R., Escors, D., & Bajaj-Elliott, M. (2014). Pseudaminic acid on Campylobacter jejuni flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. The Journal of infectious diseases, 210(9), 1487–1498.

119 Mikulak, J., Di Vito, C., Zaghi, E., & Mavilio, D. (2017). Host Immune Responses in HIV-1 Infection: The Emerging Pathogenic Role of Siglecs and Their Clinical Correlates. Frontiers in immunology, 8, 314.
120 Chang, Y. C., Olson, J., Louie, A., Crocker, P. R., Varki, A., & Nizet, V. (2014). Role of macrophage sialoadhesin in host defense against the sialylated pathogen group B Streptococcus. Journal of molecular medicine (Berlin, Germany), 92(9), 951–959.
121 Rodriguez, E., Schetters, S., & van Kooyk, Y. (2018). The tumour glyco-code as a novel immune checkpoint for immunotherapy. Nature reviews. Immunology, 18(3), 204–211.
122 Li, F., & Ding, J. (2019). Sialylation is involved in cell fate decision during development, reprogramming and cancer progression. Protein & cell, 10(8), 550–565.
123 Cagnoni, A. J., Pérez Sáez, J. M., Rabinovich, G. A., & Mariño, K. V. (2016). Turning-Off Signaling by Siglecs, Selectins, and Galectins: Chemical Inhibition of Glycan-Dependent Interactions in Cancer. Frontiers in oncology, 6, 109.
124 Lübbers, J., Rodríguez, E., & van Kooyk, Y. (2018). Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions. Frontiers in immunology, 9, 2807.
125 Sullivan-Chang, L., O'Donnell, R. T., & Tuscano, J. M. (2013). Targeting CD22 in B-cell malignancies: current status and clinical outlook. BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy, 27(4), 293–304.
126 Jandus, C., Boligan, K. F., Chijioke, O., Liu, H., Dahlhaus, M., Démoulins, T., Schneider, C., Wehrli, M., Hunger, R. E., Baerlocher, G. M., Simon, H. U., Romero, P., Münz, C., & von Gunten, S. (2014). Interactions between Siglec-7/9 receptors and ligands influence NK cell-dependent tumor immunosurveillance. The Journal of clinical investigation, 124(4), 1810–1820.
127 Hudak, J. E., Canham, S. M., & Bertozzi, C. R. (2014). Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. Nature chemical biology, 10(1), 69–75.
128 Barkal, A. A., Brewer, R. E., Markovic, M., Kowarsky, M., Barkal, S. A., Zaro, B. W., Krishnan, V., Hatakeyama, J., Dorigo, O., Barkal, L. J., & Weissman, I. L. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature, 572(7769), 392–396.

129 Wang, J., Sun, J., Liu, L. N., Flies, D. B., Nie, X., Toki, M., Zhang, J., Song, C., Zarr, M., Zhou, X., Han, X., Archer, K. A., O'Neill, T., Herbst, R. S., Boto, A. N., Sanmamed, M. F., Langermann, S., Rimm, D. L., & Chen, L. (2019). Siglec-15 as an immune suppressor and potential target for normalization cancer immunotherapy. Nature medicine, 25(4), 656–666.
130 Gao, P. S., Shimizu, K., Grant, A. V., Rafaels, N., Zhou, L. F., Hudson, S. A., Konno, S., Zimmermann, N., Araujo, M. I., Ponte, E. V., Cruz, A. A., Nishimura, M., Su, S. N., Hizawa, N., Beaty, T. H., Mathias, R. A., Rothenberg, M. E., Barnes, K. C., & Bochner, B. S. (2010). Polymorphisms in the sialic acid-binding immunoglobulin-like lectin-8 (Siglec-8) gene are associated with susceptibility to asthma. European journal of human genetics: EJHG, 18(6), 713–719.
131 Läubli, H., & Varki, A. (2020). Sialic acid-binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. Cellular and molecular life sciences: CMLS, 77(4), 593–605.
132 Sharon, N., & Lis, H. (2004). History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology, 14(11), 53R–62R.
133 Bishop, J. R., & Gagneux, P. (2007). Evolution of carbohydrate antigens--microbial forces shaping host glycomes?. Glycobiology, 17(5), 23R–34R.
134 Imberty, A., & Varrot, A. (2008). Microbial recognition of human cell surface glycoconjugates. Current opinion in structural biology, 18(5), 567–576.
135 Rillahan, C. D., & Paulson, J. C. (2011). Glycan microarrays for decoding the glycome. Annual review of biochemistry, 80, 797–823.
136 Williams, S. J., & Davies, G. J. (2001). Protein--carbohydrate interactions: learning lessons from nature. Trends in biotechnology, 19(9), 356–362.
137 Purcell, S. C., & Godula, K. (2019). Synthetic glycoscapes: addressing the structural and functional complexity of the glycocalyx. Interface focus, 9(2), 20180080.
138 Silverstein, R. M.; Webster, F. X.; KiemleJohn, D. J. (2005). Spectrometric Identification of Organic Compounds, 8th Edition,Wiley & Sons.
139 Claridge, T. D. W. (2008). High-Resolution NMR Techniques in Organic Chemistry, Volume 2, 2nd Edition, Tetrahedron Organic Chemistry.
140 Keeler, J. J. (2010). Understanding NMR Spectroscopy, Wiley & Sons, 2010.
141 Bubb, W.A. (2003). NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. Concepts Magn. Reson., 19A: 1-19.

142 Viegas, A., Macedo, A. L., & Cabrita, E. J. (2009). Ligand-based nuclear magnetic resonance screening techniques. Methods in molecular biology (Clifton, N.J.), 572, 81–100.
143 Gossert, A. D., & Jahnke, W. (2016). NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. Progress in nuclear magnetic resonance spectroscopy, 97, 82–125.
144 Meyer, B., & Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angewandte Chemie (International ed. in English), 42(8), 864–890.
145 Marchetti, R., Perez, S., Arda, A., Imberty, A., Jimenez-Barbero, J., Silipo, A., & Molinaro, A. (2016). "Rules of Engagement" of Protein-Glycoconjugate Interactions: A Molecular View Achievable by using NMR Spectroscopy and Molecular Modeling. ChemistryOpen, 5(4), 274–296.
146 Pan, A. C., Borhani, D. W., Dror, R. O., & Shaw, D. E. (2013). Molecular determinants of drug-receptor binding kinetics. Drug discovery today, 18(13-14), 667–673.
147 Furukawa, A., Konuma, T., Yanaka, S., & Sugase, K. (2016). Quantitative analysis of protein-ligand interactions by NMR. Progress in nuclear magnetic resonance spectroscopy, 96, 47–57.
148 Di Carluccio, C., Forgione, M. C., Martini, S., Berti, F., Molinaro, A., Marchetti, R., & Silipo, A. (2021). Investigation of protein-ligand complexes by ligand-based NMR methods. Carbohydrate research, 503, 108313.
149 Dias, D. M., & Ciulli, A. (2014). NMR approaches in structure-based lead discovery: recent developments and new frontiers for targeting multi-protein complexes. Progress in biophysics and molecular biology, 116(2-3), 101–112.
150 Angulo, J., & Nieto, P. M. (2011). STD-NMR: application to transient interactions between biomolecules-a quantitative approach. European biophysics journal: EBJ, 40(12), 1357–1369.
151 Angulo, J., Enríquez-Navas, P. M., & Nieto, P. M. (2010). Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. Chemistry (Weinheim an der Bergstrasse, Germany), 16(26), 7803–7812.

152 Sapay, N., Nurisso, A., & Imberty, A. (2013). Simulation of carbohydrates, from molecular docking to dynamics in water. Methods in molecular biology (Clifton, N.J.), 924, 469–483.
153 Woods, R. J., & Tessier, M. B. (2010). Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. Current opinion in structural biology, 20(5), 575–583.
154 Vanommeslaeghe, K., Guvench, O., & MacKerell, A. D., Jr (2014). Molecular mechanics. Current pharmaceutical design, 20(20), 3281–3292.
155 Grant, O. C., & Woods, R. J. (2014). Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. Current opinion in structural biology, 28, 47–55.
156 Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. Journal of computational chemistry, 29(4), 622–655.
157 Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W., & Mackerell, A. D. (2009). CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. Journal of chemical theory and computation, 5(9), 2353–2370.
158 Lins, R. D., & Hünenberger, P. H. (2005). A new GROMOS force field for hexopyranose-based carbohydrates. Journal of computational chemistry, 26(13), 1400–1412.
159 Morris, G. M., Goodsell, D. S., Huey, R., & Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. Journal of computer-aided molecular design, 10(4), 293–304.
160 Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, 31(2), 455–461.
161 Glide version 6.1, Schrödinger, LLC, New York, NY, 2012.
162 Vakser I. A. (2014). Protein-protein docking: from interaction to interactome. Biophysical journal, 107(8), 1785–1793.
163 Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. Biophysical reviews, 9(2), 91–102.

164 Li, J., Fu, A., & Zhang, L. (2019). An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. Interdisciplinary sciences, computational life sciences, 11(2), 320–328.
165 Michalewicz, Z. (1996) Evolutionary programming and genetic programming. In Genetic algorithms+ data structures= evolution programs, Springer: 283-287.
166 Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem.**,** 19 (14), 1639-1662.
167 Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert opinion on drug discovery, 10(5), 449–461.
168 Böhm H. J. (1994). 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, 8(3), 243–256.
169 Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. Journal of molecular biology, 295(2), 337–356.
170 Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry, 30(16), 2785–2791.
171 Vanommeslaeghe, K., Guvench, O., & MacKerell, A. D., Jr (2014). Molecular mechanics. Current pharmaceutical design, 20(20), 3281–3292.
172 Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. Neuron, 99(6), 1129–1143.
173 Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T.; Joannopoulos, J. (1992). Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Rev. Mod. Phys 64 (4), 1045.
174 Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79 (2), 926-935.

175 Darden, T.; York, D.; Pedersen, L. (1993) Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. J. Chem. Phys 98 (12), 10089-10092.
176 D.A. Case, et al. (2021). Amber 2021, University of California, San Francisco.
177 Dickson, C. J., Madej, B. D., Skjevik, A. A., Betz, R. M., Teigen, K., Gould, I. R., & Walker, R. C. (2014). Lipid14: The Amber Lipid Force Field. Journal of chemical theory and computation, 10(2), 865–879.
178 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of chemical theory and computation, 11(8), 3696–3713.
179 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. Journal of computational chemistry, 25(9), 1157–1174.
180 Sapay, N., Nurisso, A., & Imberty, A. (2013). Simulation of carbohydrates, from molecular docking to dynamics in water. Methods in molecular biology (Clifton, N.J.), 924, 469–483.
181 Protein Data Bank. http://www.rcsb.org
182 Universal Protein Resource (UNIPROT). http://www.uniprot.org
183 National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/
184 Basic Local Alignment Search Tool (BLAST). https://blast.ncbi.nlm.nih.gov
185 Pearson W. R. (2013). An introduction to sequence similarity ("homology") searching. Current protocols in bioinformatics, Chapter 3, Unit3.1.
186 Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. Nature protocols, 4(1), 1–13.
187 Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nature methods, 12(1), 7–8.
188 Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protocols 10 (6), 845-858.

189 Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., **26**, 283-291.
190 SAVES. https://services.mbi.ucla.edu/SAVES
191 Jayalakshmi, V., & Krishna, N. R. (2002). 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), 155(1), 106–118.
192 Krishna, N. R., & Jayalakshmi, V. (2008). Quantitative Analysis of STD-NMR Spectra of Reversibly Forming Ligand-Receptor Complexes. Topics in current chemistry, 273, 15–54.
193 Enríquez-Navas, P. M., Guzzi, C., Muñoz-García, J. C., Nieto, P. M., & Angulo, J. (2015). Structures of glycans bound to receptors from saturation transfer difference (STD) NMR spectroscopy: quantitative analysis by using CORCEMA-ST. Methods in molecular biology (Clifton, N.J.), 1273, 475–487.
194 Müller, J., & Nitschke, L. (2014). The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease. Nature reviews. Rheumatology, 10(7), 422–428.
195 Ereño-Orbea, J., Sicard, T., Cui, H., Mazhab-Jafari, M. T., Benlekbir, S., Guarné, A., Rubinstein, J. L., & Julien, J. P. (2017). Molecular basis of human CD22 function and therapeutic targeting. Nature communications, 8(1), 764.
196 Han, S., Collins, B. E., Bengtson, P., & Paulson, J. C. (2005). Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. Nature chemical biology, 1(2), 93–97.
197 Collins, B. E., Blixt, O., DeSieno, A. R., Bovin, N., Marth, J. D., & Paulson, J. C. (2004). Masking of CD22 by cis ligands does not prevent redistribution of CD22 to sites of cell contact. Proceedings of the National Academy of Sciences of the United States of America, 101(16), 6104–6109.
198 Cornall, R. J., Cyster, J. G., Hibbs, M. L., Dunn, A. R., Otipoby, K. L., Clark, E. A., & Goodnow, C. C. (1998). Polygenic autoimmune traits: Lyn, CD22, and SHP-1 are limiting elements of a biochemical pathway regulating BCR signaling and selection. Immunity, 8(4), 497–508.
199 Saphire, E. O. et al. (2001). Crystal structure of a neutralizing human IgG against HIV-1: a template for vaccine design. Science 293, 1155–1159.

200 Radaev, S. et al. Structural and functional studies of Igαβ and its assembly with the B cell antigen receptor. Structure 18, 934–943 (2010).
201 Croll, T. I. et al. (2016). Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. Structure 24, 469–476.
202 Macauley, M. S., Kawasaki, N., Peng, W., Wang, S. H., He, Y., Arlian, B. M., McBride, R., Kannagi, R., Khoo, K. H., & Paulson, J. C. (2015). Unmasking of CD22 Co-receptor on Germinal Center B-cells Occurs by Alternative Mechanisms in Mouse and Man. The Journal of biological chemistry, 290(50), 30066–30077.
203 Kimura, N., Ohmori, K., Miyazaki, K., Izawa, M., Matsuzaki, Y., Yasuda, Y., Takematsu, H., Kozutsumi, Y., Moriyama, A., & Kannagi, R. (2007). Human B-lymphocytes express alpha2-6-sialylated 6-sulfo-N-acetyllactosamine serving as a preferred ligand for CD22/Siglec-2. The Journal of biological chemistry, 282(44), 32200–32207.
204 Ai J, Advani A. 2015. Current status of antibody therapy in ALL. Br. J. Haematol. 168:471–80.
205 Angata, T., Nycholat, C. M., & Macauley, M. S. (2015). Therapeutic Targeting of Siglecs using Antibody- and Glycan-Based Approaches. Trends in pharmacological sciences, 36(10), 645–660.
206 Gottenberg, J. E., Dörner, T., Bootsma, H., Devauchelle-Pensec, V., Bowman, S. J., Mariette, X., Bartz, H., Oortgiesen, M., Shock, A., Koetse, W., Galateanu, C., Bongardt, S., Wegener, W. A., Goldenberg, D. M., Meno-Tetang, G., Kosutic, G., & Gordon, C. (2018). Efficacy of Epratuzumab, an Anti-CD22 Monoclonal IgG Antibody, in Systemic Lupus Erythematosus Patients With Associated Sjögren's Syndrome: Post Hoc Analyses From the EMBODY Trials. Arthritis & rheumatology (Hoboken, N.J.), 70(5), 763–773.
207 Büll, C., Heise, T., Adema, G. J., & Boltje, T. J. (2016). Sialic Acid Mimetics to Target the Sialic Acid-Siglec Axis. Trends in biochemical sciences, 41(6), 519–531.
208 Prescher, H., Schweizer, A., Kuhfeldt, E., Nitschke, L., & Brossmer, R. (2017). New Human CD22/Siglec-2 Ligands with a Triazole Glycoside. Chembiochem : a European journal of chemical biology, 18(13), 1216–1225.
209 Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., & Sasisekharan, R. (2008). Glycan topology

determines human adaptation of avian H5N1 virus hemagglutinin. Nature biotechnology, 26(1), 107–113.
210 Sassaki, G. L., Elli, S., Rudd, T. R., Macchi, E., Yates, E. A., Naggi, A., Shriver, Z., Raman, R., Sasisekharan, R., Torri, G., & Guerrini, M. (2013). Human (α2→6) and avian (α2→3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution. Biochemistry, 52(41), 7217–7230.
211 Poppe L, R Struike-Prill, B Meyer and H van Halbeek. 1992. The solution conformation of sialyl-α(2→6)-lactose studied by modern NMR techniques and Monte Carlo simulation. J Biomol NMR 2: 109–136.
212 Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., McCammon, J. A., Li, W. W., & Arzberger, P. W. (2009). Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective. Journal of molecular biology, 387(2), 465–491.
213 Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. Nature reviews. Nephrology, 15(6), 346–366.
214 Di Carluccio, C., Crisman, E., Manabe, Y., Forgione, R. E., Lacetera, A., Amato, J., Pagano, B., Randazzo, A., Zampella, A., Lanzetta, R., Fukase, K., Molinaro, A., Crocker, P. R., Martín-Santamaría, S., Marchetti, R., & Silipo, A. (2020). Characterisation of the Dynamic Interactions between Complex N-Glycans and Human CD22. Chembiochem : a European journal of chemical biology, 21(1-2), 129–140.
215 Guillot, A., Dauchez, M., Belloy, N., Jonquet, J., Duca, L., Romier, B., Maurice, P., Debelle, L., Martiny, L., Durlach, V., Baud, S., & Blaise, S. (2016). Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans. Scientific reports, 6, 35666.
216 Jo, S., Qi, Y., & Im, W. (2016). Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins. Glycobiology, 26(1), 19–29.
217 Nishima, W., Miyashita, N., Yamaguchi, Y., Sugita, Y., & Re, S. (2012). Effect of bisecting GlcNAc and core fucosylation on conformational properties of biantennary complex-type N-glycans in solution. The journal of physical chemistry. B, 116(29), 8504–8512.
218 Takahashi, M., Kuroki, Y., Ohtsubo, K., & Taniguchi, N. (2009). Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. Carbohydrate research, 344(12), 1387–1390.

219 Yan, J., Kline, A. D., Mo, H., Shapiro, M. J., & Zartler, E. R. (2003). The effect of relaxation on the epitope mapping by saturation transfer difference NMR. Journal of magnetic resonance (San Diego, Calif. : 1997), 163(2), 270–276.
220 Patel, D. S., Pendrill, R., Mallajosyula, S. S., Widmalm, G., & MacKerell, A. D., Jr (2014). Conformational properties of α- or β-(1→6)-linked oligosaccharides: Hamiltonian replica exchange MD simulations and NMR experiments. The journal of physical chemistry. B, 118(11), 2851–2871.
221 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191. doi:10.1093/bioinformatics/btp033
222 Papi, F., Pâris, A., Lafite, P., Daniellou, R., & Nativi, C. (2020). Synthesis of an STnThr analogue, structurally based on a TnThr antigen mimetic. Organic & biomolecular chemistry, 18(37), 7366–7372.
223 Ardá, A., Bosco, A., Sastre, J., Cañada, F. J., André, S., Gabius, H.-J. J, Richichi, B., Jiménez-Barbero J. and Nativi, C. (2015) European J. Org. Chem., 6823–6831
224 Nativi, C., Papi, F. & Roelens, S. (2019). Tn antigen analogues: the synthetic way to "upgrade" an attracting tumour associated carbohydrate antigen (TACA). Chemical communications (Cambridge, England), 55(54), 7729–7736.
225 Läubli, H., & Varki, A. (2020). Sialic acid-binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. Cellular and molecular life sciences: CMLS, 77(4), 593–605.
226 O'Reilly, M. K., & Paulson, J. C. (2010). Multivalent ligands for siglecs. Methods in enzymology, 478, 343–363.
227 Chen, G. Y., Tang, J., Zheng, P., & Liu, Y. (2009). CD24 and Siglec-10 selectively repress tissue damage-induced immune responses. Science (New York, N.Y.), 323(5922), 1722–1725.
228 von Gunten, S., & Bochner, B. S. (2008). Basic and clinical immunology of Siglecs. Annals of the New York Academy of Sciences, 1143, 61–82.
229 Whitney, G., Wang, S., Chang, H., Cheng, K. Y., Lu, P.,Zhou, X. D., Yang, W. P., McKinnon, M., & Longphre, M. (2001). A new siglec family member, siglec-10, is expressed in cells of the immune system and has signaling properties similar to CD33. European journal of biochemistry, 268(23), 6083–6096.

230 Chen, G. Y., Brown, N. K., Zheng, P., & Liu, Y. (2014). Siglec-G/10 in self-nonself discrimination of innate and adaptive immunity. Glycobiology, 24(9), 800–806.
231 Bandala-Sanchez, E., G Bediaga, N., Goddard-Borger, E. D., Ngui, K., Naselli, G., Stone, N. L., Neale, A. M., Pearce, L. A., Wardak, A., Czabotar, P., Haselhorst, T., Maggioni, A., Hartley-Tassell, L. A., Adams, T. E., & Harrison, L. C. (2018). CD52 glycan binds the proinflammatory B box of HMGB1 to engage the Siglec-10 receptor and suppress human T cell function. Proceedings of the National Academy of Sciences of the United States of America, 115(30), 7783–7788.
232 Bandala-Sanchez, E., Bediaga, N. G., Naselli, G., Neale, A. M., & Harrison, L. C. (2020). Siglec-10 expression is up-regulated in activated human CD4+ T cells. Human immunology, 81(2-3), 101–104.
233 Jellusova, J., & Nitschke, L. (2012). Regulation of B cell functions by the sialic acid-binding receptors siglec-G and CD22. Frontiers in immunology, 2, 96.
234 Meyer, S. J., Linder, A. T., Brandl, C., & Nitschke, L. (2018). B Cell Siglecs-News on Signaling and Its Interplay With Ligand Binding. Frontiers in immunology, 9, 2820.
235 Pfrengle, F., Macauley, M. S., Kawasaki, N., & Paulson, J. C. (2013). Co-presentation of antigen and ligands of Siglec-G induces B cell tolerance independent of CD22. Journal of immunology (Baltimore, Md. : 1950), 191(4), 1724–1731.
236 Zhang, P., Lu, X., Tao, K., Shi, L., Li, W., Wang, G., & Wu, K. (2015). Siglec-10 is associated with survival and natural killer cell dysfunction in hepatocellular carcinoma. The Journal of surgical research, 194(1), 107–113.
237 Sammar, M., Siwetz, M., Meiri, H., Fleming, V., Altevogt, P., & Huppertz, B. (2017). Expression of CD24 and Siglec-10 in first trimester placenta: implications for immune tolerance at the fetal-maternal interface. Histochemistry and cell biology, 147(5), 565–574.
238 Phongsisay V. (2016). The immunobiology of Campylobacter jejuni: Innate immunity and autoimmune diseases. Immunobiology, 221(4), 535–543.
239 Schnaar R. L. (2016). Glycobiology simplified: diverse roles of glycan recognition in inflammation. Journal of leukocyte biology, 99(6), 825–838.
240 Nitschke L. (2009). CD22 and Siglec-G: B-cell inhibitory receptors with distinct functions. Immunological reviews, 230(1), 128–143.

241 http://raptorx.uchicago.edu/
242 Zhuravleva, M. A., Trandem, K., & Sun, P. D. (2008). Structural implications of Siglec-5-mediated sialoglycan recognition. Journal of molecular biology, 375(2), 437–447.
243 Attrill, H., Takazawa, H., Witt, S., Kelm, S., Isecke, R., Brossmer, R., Ando, T., Ishida, H., Kiso, M., Crocker, P. R., & van Aalten, D. M. (2006). The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. The Biochemical journal, 397(2), 271–278.
244 Pröpster, J. M., Yang, F., Rabbani, S., Ernst, B., Allain, F. H., & Schubert, M. (2016). Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. Proceedings of the National Academy of Sciences of the United States of America, 113(29), E4170–E4179.
245 Thibaudeau, C., Stenutz, R., Hertz, B., Klepach, T., Zhao, S., Wu, Q., Carmichael, I., & Serianni, A. S. (2004). Correlated C-C and C-O bond conformations in saccharide hydroxymethyl groups: parametrization and application of redundant 1H-1H, 13C-1H, and 13C-13C NMR J-couplings. Journal of the American Chemical Society, 126(48), 15668–15685.
246 D. S. Patel, R. Pendrill, S. S. Mallajosyula, G. Widmalm, A. D. MacKerell Jr. Conformational Properties of α- or β-(1→6)-Linked Oligosaccharides: Hamiltonian Replica Exchange MD Simulations and NMR Experiments (2014) J. Phys. Chem. B, 118 (11), 2851–2871.
247 Pröpster, J. M., Yang, F., Ernst, B., Allain, F. H., & Schubert, M. (2015). Functional Siglec lectin domains from soluble expression in the cytoplasm of Escherichia coli. Protein expression and purification, 109, 14–22.
248 Reece-Hoyes, J. S., & Walhout, A. (2018). Gateway Recombinational Cloning. Cold Spring Harbor protocols, 2018(1), 094912.
249 Rillahan, C. D., Schwartz, E., McBride, R., Fokin, V. V., & Paulson, J. C. (2012). Click and pick: identification of sialoside analogues for siglec-based cell targeting. Angewandte Chemie (International ed. in English), 51(44), 11014–11018.
250 Movsisyan, L. D., & Macauley, M. S. (2020). Structural advances of Siglecs: insight into synthetic glycan ligands for immunomodulation. Organic & biomolecular chemistry, 18(30), 5784–5797.

251 Nicoll, G. et al. (1999). Identification and characterization of a novel siglec, siglec-7, expressed by human natural killer cells and monocytes. J. Biol. Chem. 274, 34089–95.
252 Angata, T. & Varki, A. Siglec-7: a sialic acid-binding lectin of the immunoglobulin superfamily. Glycobiology **10**, 431–8 (2000).
253 Hudak, J. E., Canham, S. M., & Bertozzi, C. R. (2014). Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. Nature chemical biology, 10(1), 69–75.
254 Alphey, M. S., Attrill, H., Crocker, P. R., & van Aalten, D. M. (2003). High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. The Journal of biological chemistry, 278(5), 3372–3377.
255 Attrill, H., Takazawa, H., Witt, S., Kelm, S., Isecke, R., Brossmer, R., Ando, T., Ishida, H., Kiso, M., Crocker, P. R., & van Aalten, D. M. (2006). The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. The Biochemical journal, 397(2), 271–278.
256 Attrill, H., Imamura, A., Sharma, R. S., Kiso, M., Crocker, P. R., & van Aalten, D. M. (2006). Siglec-7 undergoes a major conformational change when complexed with the alpha(2,8)-disialylganglioside GT1b. The Journal of biological chemistry, 281(43), 32774–32783.
257 Yamakawa, N., Yasuda, Y., Yoshimura, A., Goshima, A., Crocker, P. R., Vergoten, G., Nishiura, Y., Takahashi, T., Hanashima, S., Matsumoto, K., Yamaguchi, Y., Tanaka, H., Kitajima, K., & Sato, C. (2020). Discovery of a new sialic acid binding region that regulates Siglec-7. Scientific reports, 10(1), 8647.
258 Legrand, F., Landolina, N., Zaffran, I., Emeh, R. O., Chen, E., Klion, A. D., & Levi-Schaffer, F. (2019). Siglec-7 on peripheral blood eosinophils: Surface expression and function. Allergy, 74(7), 1257–1265.
259 Daly, J., Carlsten, M., & O'Dwyer, M. (2019). Sugar Free: Novel Immunotherapeutic Approaches Targeting Siglecs and Sialic Acids to Enhance Natural Killer Cell Cytotoxicity Against Cancer. Frontiers in immunology, 10, 1047.
260 Rosenstock, P., Horstkorte, R., Gnanapragassam, V. S., Harth, J., & Kielstein, H. (2017). Siglec-7 expression is reduced on a natural killer (NK) cell subset of obese humans. Immunologic research, 65(5), 1017–1024.
261 Varchetta, S., Mele, D., Lombardi, A., Oliviero, B., Mantovani, S., Tinelli, C., Spreafico, M., Prati, D., Ludovisi, S., Ferraioli, G., Filice, C., Aghemo, A., Lampertico, P., Facchetti, F., Bernuzzi, F., Invernizzi, P., & Mondelli, M. U.

(2016). Lack of Siglec-7 expression identifies a dysfunctional natural killer cell subset associated with liver inflammation and fibrosis in chronic HCV infection. Gut, 65(12), 1998–2006.
262 Brunetta, E., Fogli, M., Varchetta, S., Bozzo, L., Hudspeth, K. L., Marcenaro, E., Moretta, A., & Mavilio, D. (2009). The decreased expression of Siglec-7 represents an early marker of dysfunctional natural killer-cell subsets associated with high levels of HIV-1 viremia. Blood, 114(18), 3822–3830.
263 Brunetta, E., Hudspeth, K. L., & Mavilio, D. (2010). Pathologic natural killer cell subset redistribution in HIV-1 infection: new insights in pathophysiology and clinical outcomes. Journal of leukocyte biology, 88(6), 1119–1130.
264 Varchetta, S., Lusso, P., Hudspeth, K., Mikulak, J., Mele, D., Paolucci, S., Cimbro, R., Malnati, M., Riva, A., Maserati, R., Mondelli, M. U., & Mavilio, D. (2013). Sialic acid-binding Ig-like lectin-7 interacts with HIV-1 gp120 and facilitates infection of CD4pos T cells and macrophages. Retrovirology, 10, 154.
265 Ikehara, Y., Ikehara, S. K., & Paulson, J. C. (2004). Negative regulation of T cell receptor signaling by Siglec-7 (p70/AIRM) and Siglec-9. The Journal of biological chemistry, 279(41), 43117–43125.
266 Varchetta, S., Brunetta, E., Roberto, A., Mikulak, J., Hudspeth, K. L., Mondelli, M. U., & Mavilio, D. (2012). Engagement of Siglec-7 receptor induces a pro-inflammatory response selectively in monocytes. PloS one, 7(9), e45821.
267 Legrand, F., Landolina, N., Zaffran, I., Emeh, R. O., Chen, E., Klion, A. D., & Levi-Schaffer, F. (2019). Siglec-7 on peripheral blood eosinophils: Surface expression and function. Allergy, 74(7), 1257–1265.
268 Büll, C., Collado-Camps, E., Kers-Rebel, E. D., Heise, T., Søndergaard, J. N., den Brok, M. H., Schulte, B. M., Boltje, T. J., & Adema, G. J. (2017). Metabolic sialic acid blockade lowers the activation threshold of moDCs for TLR stimulation. Immunology and cell biology, 95(4), 408–415.
269 Avril, T., Wagner, E. R., Willison, H. J., & Crocker, P. R. (2006). Sialic acid-binding immunoglobulin-like lectin 7 mediates selective recognition of sialylated glycans expressed on Campylobacter jejuni lipooligosaccharides. Infection and immunity, 74(7), 4133–4141.
270 Khatua, B., Ghoshal, A., Bhattacharya, K., Mandal, C., Saha, B., Crocker, P. R., & Mandal, C. (2010). Sialic acids acquired by Pseudomonas aeruginosa are involved in reduced complement deposition and siglec mediated host-cell recognition. FEBS letters, 584(3), 555–561.

271 Avila, M., Ojcius, D. M., & Yilmaz, O. (2009). The oral microbiota: living with a permanent guest. DNA and cell biology, 28(8), 405–411.
272 Zijnge, V., van Leeuwen, M. B., Degener, J. E., Abbas, F., Thurnheer, T., Gmür, R., & Harmsen, H. J. (2010). Oral biofilm architecture on natural teeth. PloS one, 5(2), e9321.
273 Kolenbrander, P. E., Palmer, R. J., Jr, Periasamy, S., & Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell-cell distance. Nature reviews. Microbiology, 8(7), 471–480.
274 Brennan, C. A., & Garrett, W. S. (2019). Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. Nature reviews. Microbiology, 17(3), 156–166.
275 Han Y. W. (2015). Fusobacterium nucleatum: a commensal-turned pathogen. Current opinion in microbiology, 23, 141–
276 Sveen, K., & Hofstad, T. (1991). Synergistic effect on blastogenesis in murine spleen cells of lipopolysaccharide, lipid A, and acid-degraded polysaccharide from Fusobacterium nucleatum. FEMS microbiology immunology, 3(1), 47–52.
277 Okahashi, N., Koga, T., Nishihara, T., Fujiwara, T., & Hamada, S. (1988). Immunobiological properties of lipopolysaccharides isolated from Fusobacterium nucleatum and F. necrophorum. Journal of general microbiology, 134(6), 1707–1715.
278 Di Lorenzo, F., De Castro, C., Silipo, A., & Molinaro, A. (2019). Lipopolysaccharide structures of Gram-negative populations in the gut microbiota and effects on host interactions. FEMS microbiology reviews, 43(3), 257–272.
279 Hofstad, T., Skaug, N., & Sveen, K. (1993). Stimulation of B lymphocytes by lipopolysaccharides from anaerobic bacteria. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 16 Suppl 4, S200–S202.
280 Vinogradov, E., St Michael, F., & Cox, A. D. (2018). Structure of the LPS O-chain from Fusobacterium nucleatum strain ATCC 23726 containing a novel 5,7-diamino-3,5,7,9-tetradeoxy-l-gluco-non-2-ulosonic acid presumably having the d-glycero-l-gluco configuration. Carbohydrate research, 468, 69–72.
281 Vinogradov, E., St Michael, F., & Cox, A. D. (2018). Structure of the LPS O-chain from Fusobacterium nucleatum strain MJR 7757 B. Carbohydrate research, 463, 37–39.

282Vinogradov, E., St Michael, F., Homma, K., Sharma, A., & Cox, A. D. (2017). Structure of the LPS O-chain from Fusobacterium nucleatum strain 10953, containing sialic acid. Carbohydrate research, 440-441, 38–42.
283 Vinogradov, E., St Michael, F., & Cox, A. D. (2017). Structure of the LPS O-chain from Fusobacterium nucleatum strain 12230. Carbohydrate research, 448, 115–117.
284 Vinogradov, E., St Michael, F., & Cox, A. D. (2017). The structure of the LPS O-chain of Fusobacterium nucleatum strain 25586 containing two novel monosaccharides, 2-acetamido-2,6-dideoxy-l-altrose and a 5-acetimidoylamino-3,5,9-trideoxy-gluco-non-2-ulosonic acid. Carbohydrate research, 440-441, 10–15.
285 Onoue, S., Niwa, M., Isshiki, Y., & Kawahara, K. (1996). Extraction and characterization of the smooth-type lipopolysaccharide from Fusobacterium nucleatum JCM 8532 and its biological activities. Microbiology and immunology, 40(5), 323–331.
286 Garcia-Vello, P., Di Lorenzo, F., Lamprinaki, D., Notaro, A., Speciale, I., Molinaro, A., Juge, N., & De Castro, C. (2021). Structure of the O-Antigen and the Lipid A from the Lipopolysaccharide of Fusobacterium nucleatum ATCC 51191. Chembiochem : a European journal of chemical biology, 22(7), 1252–1260.
287 Vinogradov, E., St Michael, F., Homma, K., Sharma, A., & Cox, A. D. (2017). Structure of the LPS O-chain from Fusobacterium nucleatum strain 10953, containing sialic acid. Carbohydrate research, 440-441, 38–42.
288 Davis, M. R., Jr, & Goldberg, J. B. (2012). Purification and visualization of lipopolysaccharide from Gram-negative bacteria by hot aqueous-phenol extraction. Journal of visualized experiments : JoVE, (63), 3916.
289 Ribeiro, M. M., Franquelim, H. G., Castanho, M. A., & Veiga, A. S. (2008). Molecular interaction studies of peptides using steady-state fluorescence intensity. Static (de)quenching revisited. Journal of peptide science : an official publication of the European Peptide Society, 14(4), 401–406.
1Wang, X., & Bai, Y. (2016). The global Minmax k-means algorithm. SpringerPlus, 5(1), 1665.
290 Fittolani, G., Seeberger, P.H., Delbianco, M. (2020). Helical polysaccharides. Peptide Science. 112:e24124.

291 Kusumoto, S., Fukase, K., & Shiba, T. (2010). 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, 86(4), 322–337.
292 Tytgat, H. L., & Lebeer, S. (2014). The sweet tooth of bacteria: common themes in bacterial glycoconjugates. Microbiology and molecular biology reviews : MMBR, 78(3), 372–417.
293 Di Lorenzo, F., De Castro, C., Lanzetta, R. Parrilli, M., Silipo, A., Molinaro, A. (2015). Lipopolysaccharides as Microbe-associated Molecular Patterns: A Structural Perspective**,** RSC Drug Discovery Series, 43, 38-63.
294 Gamage, S. D., McGannon, C. M., & Weiss, A. A. (2004). Escherichia coli serogroup O107/O117 lipopolysaccharide binds and neutralizes Shiga toxin
295 Journal of bacteriology, 186(16), 5506–5512.
296 Silipo, A., & Molinaro, A. (2010). The diversity of the core oligosaccharide in lipopolysaccharides. Sub-cellular biochemistry, 53, 69–99.
297 Park, B. S., Song, D. H., Kim, H. M., Choi, B. S., Lee, H., & Lee, J. O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature, 458(7242), 1191–1195.
298 Wesener, D. A., Dugan, A., & Kiessling, L. L. (2017). Recognition of microbial glycans by soluble human lectins. Current opinion in structural biology, 44, 168–178.
299 Muller-Loennies, S., Brade, L., MacKenzie, C. R., Di Padova, F. E., & Brade, H. (2003). Identification of a cross-reactive epitope widely present in lipopolysaccharide from enterobacteria and recognized by the cross-protective monoclonal antibody WN1 222-5. The Journal of biological chemistry, 278(28), 25618–25627.
300 Orozco, C. A., Martinez-Bosch, N., Guerrero, P. E., Vinaixa, J., Dalotto-Moreno, T., Iglesias, M., Moreno, M., Djurec, M., Poirier, F., Gabius, H. J., Fernandez-Zapico, M. E., Hwang, R. F., Guerra, C., Rabinovich, G. A., & Navarro, P. (2018). Targeting galectin-1 inhibits pancreatic cancer progression by modulating tumor-stroma crosstalk. Proceedings of the National Academy of Sciences of the United States of America, 115(16), E3769–E3778.
301 Napoletano, C., Zizzari, I. G., Rughetti, A., Rahimi, H., Irimura, T., Clausen, H., Wandall, H. H., Belleudi, F., Bellati, F., Pierelli, L., Frati, L., & Nuti, M.

(2012). Targeting of macrophage galactose-type C-type lectin (MGL) induces DC signaling and activation. European journal of immunology, 42(4), 936–945.
302 Stowell, S. R., Arthur, C. M., McBride, R., Berger, O., Razi, N., Heimburg-Molinaro, J., Rodrigues, L. C., Gourdine, J. P., Noll, A. J., von Gunten, S., Smith, D. F., Knirel, Y. A., Paulson, J. C., & Cummings, R. D. (2014). Microbial glycan microarrays define key features of host-microbial interactions. Nature chemical biology, 10(6), 470–476.
303 van Vliet, S. J., Gringhuis, S. I., Geijtenbeek, T. B., & van Kooyk, Y. (2006). Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. Nature immunology, 7(11), 1200–1208.
304 Valverde, P., Martínez, J. D., Cañada, F. J., Ardá, A., & Jiménez-Barbero, J. (2020). Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. Chembiochem : a European journal of chemical biology, 21(21), 2999–3025.
305 van Kooyk, Y., Ilarregui, J. M., & van Vliet, S. J. (2015). Novel insights into the immunomodulatory role of the dendritic cell and macrophage-expressed C-type lectin MGL. Immunobiology, 220(2), 185–192.
306 Mnich, M. E., van Dalen, R., Gerlach, D., Hendriks, A., Xia, G., Peschel, A., van Strijp, J., & van Sorge, N. M. (2019). The C-type lectin receptor MGL senses N-acetylgalactosamine on the unique Staphylococcus aureus ST395 wall teichoic acid. Cellular microbiology, 21(10), e13072.
307 Pozzi, C., Olaniyi, R., Liljeroos, L., Galgani, I., Rappuoli, R., & Bagnoli, F. (2017). Vaccines for Staphylococcus aureus and Target Populations. Current topics in microbiology and immunology, 409, 491–528.
308 Mayer, M., and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed Engl 38 1784–1788.
309 Marcelo, F., Garcia-Martin, F., Matsushita, T., Sardinha, J., Coelho, H., Oude-Vrielink, A., Koller, C., André, S., Cabrita, E. J., Gabius, H. J., Nishimura, S., Jiménez-Barbero, J., & Cañada, F. J. (2014). Delineating binding modes of Gal/GalNAc and structural elements of the molecular recognition of tumor-associated mucin glycopeptides by the human macrophage galactose-type lectin. Chemistry (Weinheim an der Bergstrasse, Germany), 20(49),
310 Kolatkar, A. R., and Weis, W. I. (1996) Structural basis of galactose recognition by C-type animal lectins. J. Biol. Chem. 271 (12), 6679−6685.

311 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England), 23(21), 2947–2948.
312 Gabba, A., Bogucka, A., Luz, J. G., Diniz, A., Coelho, H., Corzana, F., Cañada, F. J., Marcelo, F., Murphy, P. V., & Birrane, G. (2021). Crystal Structure of the Carbohydrate Recognition Domain of the Human Macrophage Galactose C-Type Lectin Bound to GalNAc and the Tumor-Associated Tn Antigen. Biochemistry, 60(17), 1327–1336.
313 Sugawara, H., Kusunoki, M., Kurisu, G., Fujimoto, T., Aoyagi, H., & Hatakeyama, T. (2004). Characteristic recognition of N-acetylgalactosamine by an invertebrate C-type Lectin, CEL-I, revealed by X-ray crystallographic analysis. The Journal of biological chemistry, 279(43), 45219–45225.
314 Hatakeyama, T., Kamiya, T., Kusunoki, M., Nakamura-Tsuruta, S., Hirabayashi, J., Goda, S., & Unno, H. (2011). Galactose recognition by a tetrameric C-type lectin, CEL-IV, containing the EPN carbohydrate recognition motif. The Journal of biological chemistry, 286(12), 10305–10315.
315 Feinberg, H., Taylor, M. E., Razi, N., McBride, R., Knirel, Y. A., Graham, S. A., Drickamer, K., & Weis, W. I. (2011). Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. Journal of molecular biology, 405(4), 1027–1039.
316 Cambi, A., Koopman, M., & Figdor, C. G. (2005). How C-type lectins detect pathogens. Cellular microbiology, 7(4), 481–488.
317 Steeghs, L., van Vliet, S. J., Uronen-Hansson, H., van Mourik, A., Engering, A., Sanchez-Hernandez, M., Klein, N., Callard, R., van Putten, J. P., van der Ley, P., van Kooyk, Y., & van de Winkel, J. G. (2006). Neisseria meningitidis expressing lgtB lipopolysaccharide targets DC-SIGN and modulates dendritic cell function. Cellular microbiology, 8(2), 316–325.
318 van Vliet, S. J., Saeland, E., & van Kooyk, Y. (2008). Sweet preferences of MGL: carbohydrate specificity and function. Trends in immunology, 29(2), 83–90.
319 Vukman, K. V., Ravidà, A., Aldridge, A. M., & O'Neill, S. M. (2013). Mannose receptor and macrophage galactose-type lectin are involved in Bordetella pertussis mast cell interaction. Journal of leukocyte biology, 94(3), 439–448.

320 van Sorge, N. M., Bleumink, N. M., van Vliet, S. J., Saeland, E., van der Pol, W. L., van Kooyk, Y., & van Putten, J. P. (2009). N-glycosylated proteins and distinct lipooligosaccharide glycoforms of Campylobacter jejuni target the human C-type lectin receptor MGL. Cellular microbiology, 11(12), 1768–1781.
321 Zizzari, I. G., Martufi, P., Battisti, F., Rahimi, H., Caponnetto, S., Bellati, F., Nuti, M., Rughetti, A., & Napoletano, C. (2015). The Macrophage Galactose-Type C-Type Lectin (MGL) Modulates Regulatory T Cell Functions. PloS one, 10(7), e0132617.
322 Latner, D. R., & Hickman, C. J. (2015). Remembering mumps. PLoS pathogens, 11(5), e1004791.
323 Ninth Report of the International Committee on Taxonomy of Viruses (Eds.: A. M. Q. King, M. J. Adams, E. B. Carstens, E. J. Lefkowitz), Elsevier, New York, 2012, pp. 672-685.
324 Rubin, S., Eckhaus, M., Rennick, L. J., Bamford, C. G., & Duprex, W. P. (2015). Molecular biology, pathogenesis and pathology of mumps virus. The Journal of pathology, 235(2), 242–252.
325 Willocks, L. J., Guerendiain, D., Austin, H. I., Morrison, K. E., Cameron, R. L., Templeton, K. E., DE Lima, V., Ewing, R., Donovan, W., & Pollock, K. (2017). An outbreak of mumps with genetic strain variation in a highly vaccinated student population in Scotland. Epidemiology and infection, 145(15), 3219–3225.
326 Donahue, M., Hendrickson, B., Julian, D., Hill, N., Rother, J., Koirala, S., Clayton, J. L., Safranek, T., & Buss, B. (2020). Multistate Mumps Outbreak Originating from Asymptomatic Transmission at a Nebraska Wedding - Six States, August-October 2019. MMWR. Morbidity and mortality weekly report, 69(22), 666–669.
327 Paterson, R. G., & Lamb, R. A. (1990). RNA editing by G-nucleotide insertion in mumps virus P-gene mRNA transcripts. Journal of virology, 64(9), 4137–4145.
328 Elliott, G. D., Yeo, R. P., Afzal, M. A., Simpson, E. J., Curran, J. A., & Rima, B. K. (1990). Strain-variable editing during transcription of the P gene of mumps virus may lead to the generation of non-structural proteins NS1 (V) and NS2. The Journal of general virology, 71 (Pt 7), 1555–1560.
329Wilson, R. L., Fuentes, S. M., Wang, P., Taddeo, E. C., Klatt, A., Henderson, A. J., & He, B. (2006). Function of small hydrophobic proteins of paramyxovirus. Journal of virology, 80(4), 1700–1709.

330 Elango, N., Varsanyi, T. M., Kövamees, J., & Norrby, E. (1988). Molecular cloning and characterization of six genes, determination of gene order and intergenic sequences and leader sequence of mumps virus. The Journal of general virology, 69 (Pt 11), 2893–2900.
331 Elliott, G. D., Afzal, M. A., Martin, S. J., & Rima, B. K. (1989). Nucleotide sequence of the matrix, fusion and putative SH protein genes of mumps virus and their deduced amino acid sequences. Virus research, 12(1), 61–75.
332 R. A. Lamb, G. D. Parks, in Field Virology (Eds.: D. M. Knipe, P. M. Howley), Lippincott Williams & Wilkins, Philadelphia, 6th Ed, Vol I, 2013, pp. 957–995.
333 Orvell, C., Alsheikhly, A. R., Kalantari, M., & Johansson, B. (1997). Characterization of genotype-specific epitopes of the HN protein of mumps virus. The Journal of general virology, 78 (Pt 12), 3187–3193.
334 Mumps virus nomenclature update: 2012. (2012). Releve epidemiologique hebdomadaire, 87(22), 217–224.
335 S. A. Rubin, C. J. Sauder, K. M. Carbone in Field Virology (Eds.: D. M. Knipe, P. M. Howley), Lippincott Williams & Wilkins, Philadelphia, 6th Ed, Vol I, 2013, pp. 1024–1041.
336 Cui, A., Zhu, Z., Hu, Y., Deng, X., Sun, Z., Zhang, Y., Mao, N., Xu, S., Fang, X., Gao, H., Si, Y., Lei, Y., Zheng, H., He, J., Wu, H., & Xu, W. (2017). Mumps Epidemiology and Mumps Virus Genotypes Circulating in Mainland China during 2013-2015. PloS one, 12(1), e0169561.
337 Lamb R. A. (1993). Paramyxovirus fusion: a hypothesis for changes. Virology, 197(1), 1–11.
338 Schauer R. (1982). Chemistry, metabolism, and biological functions of sialic acids. Advances in carbohydrate chemistry and biochemistry, 40, 131–234.
339 Takimoto, T., Taylor, G. L., Connaris, H. C., Crennell, S. J., & Portner, A. (2002). Role of the hemagglutinin-neuraminidase protein in the mechanism of paramyxovirus-cell membrane fusion. Journal of virology, 76(24), 13028–13033.
340 Lamb R.A., Kolakofsky D. Paramyxoviridae: the viruses and their replication. In: Knipe D.M., Howley P.M. eds. Fields Virology (4th ed.). Philadelphia: Lippincott, Williams & Wilkins, 2001.
341 Kubota, M., Takeuchi, K., Watanabe, S., Ohno, S., Matsuoka, R., Kohda, D., Nakakita, S. I., Hiramatsu, H., Suzuki, Y., Nakayama, T., Terada, T., Shimizu, K., Shimizu, N., Shiroishi, M., Yanagi, Y., & Hashiguchi, T. (2016). Trisaccharide

containing α2,3-linked sialic acid is a receptor for mumps virus. Proceedings of the National Academy of Sciences of the United States of America, 113(41), 11579–11584.
342 Kubota, M., Matsuoka, R., Suzuki, T., Yonekura, K., Yanagi, Y., & Hashiguchi, T. (2019). Molecular Mechanism of the Flexible Glycan Receptor Recognition by Mumps Virus. Journal of virology, 93(15), e00344-19.
343 Kötzler, M. P., Blank, S., Bantleon, F. I., Spillner, E., & Meyer, B. (2012). Donor substrate binding and enzymatic mechanism of human core α1,6-fucosyltransferase (FUT8). Biochimica et biophysica acta, 1820(12), 1915–1925.
344 McCarter, J. D., & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. Current opinion in structural biology, 4(6), 885–892.
345 Brzozowski, A. M., & Davies, G. J. (1997). Structure of the Aspergillus oryzae alpha-amylase complexed with the inhibitor acarbose at 2.0 A resolution. Biochemistry, 36(36), 10837–10845.
346 Exnowitz, F., Meyer, B., & Hackl, T. (2012). NMR for direct determination of K(m) and V(max) of enzyme reactions based on the Lambert W function-analysis of progress curves. Biochimica et biophysica acta, 1824(3), 443–449.
347 Goličnik M. (2011). Explicit analytic approximations for time-dependent solutions of the generalized integrated Michaelis-Menten equation. Analytical biochemistry, 411(2), 303–305.
348 Golicnik M. (2010). Explicit reformulations of time-dependent solution for a Michaelis-Menten enzyme reaction model. Analytical biochemistry, 406(1), 94–96.
349Voet D., Voet J., Pratt, C. (2008). Principles of Biochemistry, Wiley.
350 Ohrui H., Nishida Y., Itoh H., Meguro H. (1991). -ray and NMR studies on the rotational isomerism about the C(5)-C(6) bond of 6-substituted methyl 2,3,4,-tri-O-acetyl-6-X-α-D-glucopyranoside derivatives, J. Org. Chem. 56, 5.
351 Poppe L., Brown G. S., Philo J. S., Nikrad P. V., Shah B. H., J. Am. Chem. Soc. 1997, 119, 7, 1727–1736.
352 McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E., & McKimm-Breschkin, J. L. (2019). Influenza Virus Neuraminidase Structure and Functions. *Frontiers in microbiology*, *10*, 39.
353 Chan, J., Watson, J. N., Lu, A., Cerda, V. C., Borgford, T. J., & Bennet, A. J. (2012). Bacterial and viral sialidases: contribution of the conserved active site glutamate to catalysis. Biochemistry, 51(1), 433–441.

354 Goudar, C. T., Harris, S. K., McInerney, M. J., & Suflita, J. M. (2004). Progress curve analysis for enzyme and microbial kinetic reactions using explicit solutions based on the Lambert W function. Journal of microbiological methods, 59(3), 317–326.
355 Forgione, R. E., Di Carluccio, C., Kubota, M., Manabe, Y., Fukase, K., Molinaro, A., Hashiguchi, T., Marchetti, R., & Silipo, A. (2020). Structural basis for Glycan-receptor binding by mumps virus hemagglutinin-neuraminidase. Scientific reports, 10(1), 1589.
356 Gouma, S., Vermeire, T., Van Gucht, S., Martens, L., Hutse, V., Cremer, J., Rota, P. A., Leroux-Roels, G., Koopmans, M., van Binnendijk, R., & Vandermarliere, E. (2020). Publisher Correction: Differences in antigenic sites and other functional regions between genotype A and G mumps virus surface proteins. Scientific reports, 10(1), 5456.
357 Colman PM, Hoyne PA, Lawrence MC (1993) Sequence and structure alignment of paramyxovirus hemagglutinin-neuraminidase with influenza virus neuraminidase. J Virol 67(6):2972–2980
358 Winger, M., & von Itzstein, M. (2012). Exposing the flexibility of human parainfluenza virus hemagglutinin-neuraminidase. Journal of the American Chemical Society, 134(44), 18447–18452.
359Turnbull W.B., Field R.A., J. Chem. Soc., Perkin Trans., 2000, 1, 1859-1866
360 Yoneda, Y., Kawada, T., Rosenau, T., & Kosma, P. (2005). Synthesis of methyl 4'-O-methyl-13C12-beta-D-cellobioside from 13C6-D-glucose. Part 1: reaction optimization and synthesis. Carbohydrate research, 340(15), 2428–2435.
361 A. M. Galazka, S. E. Robertson, A. Kraigher, Bull. World Health Organ. 1999; 77, pp. 3-14.
362 Rubin, S. A., Link, M. A., Sauder, C. J., Zhang, C., Ngo, L., Rima, B. K., & Duprex, W. P. (2012). Recent mumps outbreaks in vaccinated populations: no evidence of immune escape. Journal of virology, 86(1), 615–620.
363 Yoshida, N., Fujino, M., Miyata, A., Nagai, T., Kamada, M., Sakiyama, H., Ihara, T., Kumagai, T., Okafuji, T., Okafuji, T., & Nakayama, T. (2008). Mumps virus reinfection is not a rare event confirmed by reverse transcription loop-mediated isothermal amplification. Journal of medical virology, 80(3), 517–523.
364 Savage, E., Ramsay, M., White, J., Beard, S., Lawson, H., Hunjan, R., & Brown, D. (2005). Mumps outbreaks across England and Wales in 2004: observational study. BMJ (Clinical research ed.), 330(7500), 1119–1120.

365 R. Rydbeck, A. Löve, C. Örvell, E. J. Norrby, Gen. Virol. 1986, 67, 2, 281-287.
366 Sauder, C. J., Vandenburgh, K. M., Iskow, R. C., Malik, T., Carbone, K. M., & Rubin, S. A. (2006). Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity. Virology, 350(1), 48–57.
367 Peltola, H., Kulkarni, P. S., Kapre, S. V., Paunio, M., Jadhav, S. S., & Dhere, R. M. (2007). Mumps outbreaks in Canada and the United States: time for new thinking on mumps vaccines. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 45(4), 459–466.
368 Crennell, S., Takimoto, T., Portner, A., & Taylor, G. (2000). Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. Nature structural biology, 7(11), 1068–1074.
369 Angata, T., Nycholat, C. M., & Macauley, M. S. (2015). Therapeutic Targeting of Siglecs using Antibody- and Glycan-Based Approaches. Trends in pharmacological sciences, 36(10), 645–660.
370 Kajihara, Y., Suzuki, Y., Yamamoto, N., Sasaki, K., Sakakibara, T., & Juneja, L. R. (2004). Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its application to the solid-phase synthesis of a glycopeptide with Asn-linked sialyl-undeca- and asialo-nonasaccharides. *Chemistry (Weinheim an der Bergstrasse, Germany)*, *10*(4), 971–985.
371 Nagasaki, M., Manabe, Y., Minamoto, N., Tanaka, K., Silipo, A., Molinaro, A., & Fukase, K. (2016). Chemical Synthesis of a Complex-Type N-Glycan Containing a Core Fucose. *The Journal of organic chemistry*, *81*(22), 10600–10616.
372 Y. Kajihara, Y. Suzuki, N. Yamamoto, K. Sasaki, T. Sakakibara, L.R. Juneja. (2004) Prompt chemoenzymatic synthesis of diverse complex-type oligosaccharides and its application to the solid-phase synthesis of a glycopeptide with Asn-linked sialyl-undeca- and asialo-nonasaccharides. Chem. Eur. J. 10, 971-985.
373 Di Carluccio, C., Forgione, R. E., Montefiori, M., Civera, M., Sattin, S., Smaldone, G., Fukase, K., Manabe, Y., Crocker, P. R., Molinaro, A., Marchetti, R., & Silipo, A. (2020). Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22. iScience, 24(1), 101998.
374 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., Martín-Santamaría, S., Crocker, P. R., Marchetti, R., & Silipo, A. (2020).

Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. iScience, 23(6), 101231.
375 Buj, R., Iglesias, N., Planas, A. M., & Santalucía, T. (2013). A plasmid toolkit for cloning chimeric cDNAs encoding customized fusion proteins into any Gateway destination expression vector. BMC molecular biology, 14(1), 18.
376 Maalej, M., Forgione, R. E., Marchetti, R., Bulteau, F., Thépaut, M., Lanzetta, R., Laguri, C., Simorre, J. P., Fieschi, F., Molinaro, A., & Silipo, A. (2019). Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of Escherichia coli Lipooligosaccharide. Chembiochem : a European journal of chemical biology, 20(14), 1778–1782.
377 Kubota M., Hashiguchi T. (2020). Large-Scale Expression and Purification of Mumps Virus Hemagglutinin-Neuraminidase for Structural Analyses and Glycan-Binding Assays. Methods Mol. Biol.
378 Reeves P. J., Callewaert N., Contreras R., Khorana H. G. (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. Proc. Natl. Acad. Sci. U S A. 99(21),13419-24.
379 Rezania S., Amirmozaffari N., Tabarraei B., Jeddi-Tehrani M., Zarei O., Alizadeh R., Masjedian F. & Zarnani A. H. (2011). Extraction, Purification and Characterization of Lipopolysaccharide from Escherichia coli and Salmonella typhi. Avicenna journal of medical biotechnology, 3(1), 3–9.
380 Leone, S., Silipo, A., L Nazarenko, E., Lanzetta, R., Parrilli, M., & Molinaro, A. (2007). Molecular structure of endotoxins from Gram-negative marine bacteria: an update. Marine drugs, 5(3), 85–112.
381 Kittelberger, R., & Hilbink, F. (1993). Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. *Journal of biochemical and biophysical methods*, *26*(1), 81–86.
382 Holst O. (2000). Deacylation of lipopolysaccharides and isolation of oligosaccharide phosphates. *Methods in molecular biology (Clifton, N.J.)*, *145*, 345–353.
383 Holst O., Muller-Loennies S., Lindner B., Brade H. (1993). Chemical structure of the lipid A of Escherichia coli J-5, Eur J Biochem, 214, 695–701.
384 Her C., Alonzo A. P., Bang J. Y., Torres E., Krishnan V. V., (2015). Real-Time Enzyme Kinetics by Quantitative NMR Spectroscopy and Determination of

the Michaelis–Menten Constant Using the Lambert-W Function, J. Chem. Educ. 92, 11, 1943-194.
385 Marchler-Bauer A. et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res, 45, D200-D203.

AKNOWLEDGMENTS

First of all, I would like to thank my supervisor Prof. Alba Silipo for the outstanding and constant supervision of my PhD Thesis work and for her support, even when thing were very hard for me from a personal perspective. Also, I want to thank her for and for giving me many growth opportunities, for putting a lot of trust in myself, and last but not least, for transmitting me the passion for the glycans world.

Also, I would like to thank Prof. Antonio Molinaro for conveying is knowledge and experience to us students, and for the many important scientific advices provided to me throughout the PhD journey.

I also thank Prof. Roberta Marchetti for her key help in the PhD project achievements, for being there when I needed someone to talk or some assistance.

I would like also to thank Prof. Sonsoles Martín Santamaría and Dr. Joan Guzmàn-Caldentey for tutoring me about Docking and MD techniques during the period I spent (in person and virtually) at CSIC and Madrid. These periods have been very important for me because allowed me to understand my predisposition for this field.

I acknowledge Prof. Marco Fragai and Dr. Stefano Giuntini for teaching me protein recombinant expression methodologies during the period I spent at Giotto Biotech and for the opportunity to join a great scientific community.

I would like to thank my examiner Prof. Maria Monti, for the useful scientific suggestions provided during our meetings.

Thanks to Prof. Angela Lombardi, for having a major impact in my decision to pursue a PhD.

I would like to acknowledge the FSE, PON Ricerca e Innovazione 2014-2020 grant for funding my PhD.

I would like to thank my colleagues Cristina, Maria, Ferran, Molly, Pilar, Giulia, Rossella and all the people come through the lab, for contributing in different ways to reach this important goal.

Special thanks to Fabio and Felicia, for sharing with me all the highs and lows of the PhD journey, that we started together.

Thanks to all my friends across the world who have made me feel their presence despite the distance.

I want to thank my partner Alfonso, for the constant encouragement, for the patience, for the crucial pep talks. For being always here for me and for supporting me no matter what. I am very lucky to have you by my side.

I would like to dedicate this Thesis to my family. To my mother for showing me how to be strong and resilient and to my brother for the essential support and suggestions.

Amongst all, I want to thank my father. I may have lost him last year, but I wouldn't be at this point without him and all the things he did for me until his very last day. The only thing that would make me happier it would be the possibility to share this achievement with you, dad.

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