UNIVERSITÀ DEGLI STUDI FIRENZE

HR EXCELLENCE IN RESEARCH

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CYCLE XXXVII

COORDINATOR Prof. Roberta Pierattelli

NMR-based structural studies of pathologically relevant protein-sialoglycan interactions and drug delivery systems

Settore Scientifico Disciplinare CHIM/03

Doctoral Candidate
Dott. Luis David Padilla Cortés

Supervisor
Prof. Marco Fragai

Coordinator
Prof. Roberta Pierattelli

November 2021 – October 2024

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Universiteit Utrecht

KLINIKUM DER JOHANN WOLFGANG GOETHE-UNIVERSITÄT FRANKFURT AM MAIN

Contents

1. Introduction

Molecular recognition of a specific receptor by a given ligand is an invaluable source of information in an organism, and a cascade of such interactions constitutes the basis for all the biological processes in life. A complete understanding of the biomolecular interactions and the structural mechanisms governing molecular recognition is key to describing these biological processes. By elucidating the three-dimensional structure of biomolecular complexes, we can precisely map the binding interfaces, as well as conformational changes and dynamic features that drive molecular recognition. This knowledge is critical not only for understanding healthy biological systems, but also for identifying how structural defects lead to pathological outcomes, guiding the rational design of therapeutic interventions.

Different methodologies are able to capture specific aspects of a biomolecule's behavior, hence, an integrative approach is crucial for fully describing the biophysical and structural features of a system. The size of the system, its dynamics, the affinity (in case of an interaction), or simply, the sample preparation requirements and types of response detected are crucial when choosing the right techniques for the study.

Among the different methods in structural biology, Nuclear Magnetic Resonance (NMR) comes as a powerful tool for the characterization of proteins and their interactions, providing atomic resolution insights into the structure of a complex. It is also a major tool for the study of dynamics or changes in the dynamic features of a system, such as molecular motions and conformational changes upon an interaction.

Several experiments are available to explore each of these properties and are broadly divided into ligand-based and protein-based NMR experiments. Ligand-based experiments such as Saturation Transfer Difference (STD) and WaterLOGSY, are useful experiments focused on the detection of changes in the NMR spectrum of a ligand upon binding to a specific receptor. These experiments are particularly useful for ligand screening, the assessment of the binding affinity of a complex, the detection of transient and weak interactions, as well as to map ligand epitopes involved in the interaction. Protein-based

NMR experiments on the contrary, monitor the effect of a ligand on the NMR spectrum of a protein, most of the time, isotopically enriched in ^{15}N and/or ^{13}C isotopes. A larger number of protein-based NMR experiments are available to monitor protein-ligand interactions, among which, 2D NMR titrations monitored by ^{1}H-^{15}N HSQC experiments, NOESY, T1 and T2 relaxation measurements are the most widely used. Important information such as the binding affinity, as well as other thermodynamic and kinetic parameters, mapping of the binding site, information about the protein residues involved in the interaction, and dynamics and conformational changes in the protein upon binding, can be obtained.

More technical aspects of specific NMR experiments are further discussed in the methodology section of this thesis, however, the ability to monitor both ligand and protein signals in solution provides a complementary and comprehensive understanding of an interaction. This wide variety of approaches that can be applied through NMR makes it an essential tool for exploring the molecular details of protein-glycan interactions as well as for the structural characterization biologics.

1.1 Biophysical characterization of carbohydrate binding proteins

The sialic acid family

Sialic acids (Sias), also known as neuraminic acids (Neus) are a member of the nonulosonic acid (NulO) family, a group of α-keto acid monosaccharides with a 9-carbon backbone[1]. They are characterized for their diversity and high number of functional groups for a single monosaccharide, which combined with the various epimers and modifications, make them a complex molecule. In sialylated glycans, this complexity is further increased by the different types of glycosidic bonds that can link the Sia residue and the underlying glycan, making sialoglycans ideal as information carriers during cell-cell and host-pathogen recognition[2].

The structural diversity in sialic acids arises from the different chemical modifications the molecule can undergo. N-acetyl neuraminic acid (Neu5Ac), with an N-acetyl group in the C-5 position; N-glycolyl neuraminic acid (Neu5Gc), with a hydroxylated 5-N-acetyl group;

and 2-keto-3-deoxynonic acid (Kdn), with a hydroxyl group in C-5 position, are the most common forms of sialic acid (Figure 1). However, the less common neuraminic acid (Neu), with a non-acetylated 5-amino group can also be found in nature. Additional phosphate groups, O-acetyl, O-methyl, O-sulfate and O-lactyl modifications of these four core Sia forms (Neu5Ac, Neu5Gc, Kdn and Neu) on the hydroxyl groups at C-4, C-7, C-8 and C-9 are also found. Furthermore, the C-1 carboxylate group, is ionized at physiological pH, but it can also be converted into a lactone with hydroxyl groups of adjacent glycans, or into a lactam with the free amino group at C-5. All of these modifications, in combination with the different glycosidic linkages, generate an immense structural diversity in which Sia can be presented on the cell surface[3,4].

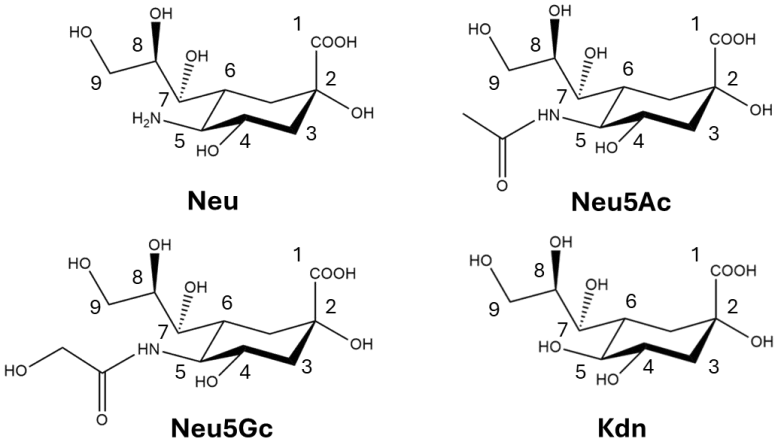


Figure 1. Chair conformations in the α configuration in C-2 position, with numbering of the positions indicated for: A) neuraminic acid (Neu), B) N-acetyl neuraminic acid (Neu5Ac), C) N-glycolyl neuraminic acid (Neu5Gc) and D) 2-keto-3-deoxynonic acid (Kdn).

Typically, sialic acids are found as part of several glycoconjugates, capping different glycosphingolipids, as well as N- and O-glycans on the cell surface and conferring unique properties to the molecules containing them. In general, Sias modulate the stability and function of the glycoprotein or glycosphingolipid and influence the net charges on the cell surface[5]. Hence, aberrant sialoglycan expression has been associated with several

pathological conditions. In particular, the roles of hypersialylation in several types of cancer, as well as in bacterial and viral infections have been widely investigated[6–10].

Biosynthesis of sialylated glycoconjugates

Three enzymes participate in the biosynthesis of Neu5Ac, the predominant form of sialic acid in humans. First, N-acetylmannosamine (ManNAc) is formed from the epimerization of UDP-acetylglucosamine (UDP-GlcNAc). ManNAc is then phosphorylated in the C-6 position by the same bifunctional enzyme, an UDP-GlcNAc-2-epimerase/Nacetylmannosamine kinase forming a ManNAc-6P intermediate. This 6-carbon intermediate is further condensed to the 3-carbon molecule phosphoenolpyruvate (PEP) by the N-acetyl neuraminic acid-9P synthetase (NANS) to yield the 9-carbon molecule N-acetyl neuraminic acid-9P. The molecule is then dephosphorylated by the N-acetyl neuraminate 9-phosphate phosphohydrolase (NANP), resulting in Neu5Ac. The whole process occurs in the cytosolic compartment as illustrated in Figure 2[1,2].

Neu5Ac is then converted to cytidine monophosphate-sialic acid (CMP-Neu5Ac), which acts as an activated form of Neu5Ac, in a process that happens in the nucleus for reasons that remain unclear. In vertebrates, sialyltransferases then transfer Neu5Ac from CMP-Neu5Ac to lipids and protein glycans in the Golgi, which are then further transported to the cell surface for glycan recognition. At last, these sialoglycans are taken up into lysosomes for degradation, and the metabolites reused. Additional modifications of Neu5Ac, such as the hydroxylation of N-acetyl group to form Neu5Gc or O-acetylation occur in the cytosol, while methylation or sulfation likely occur in the Golgi apparatus[1].

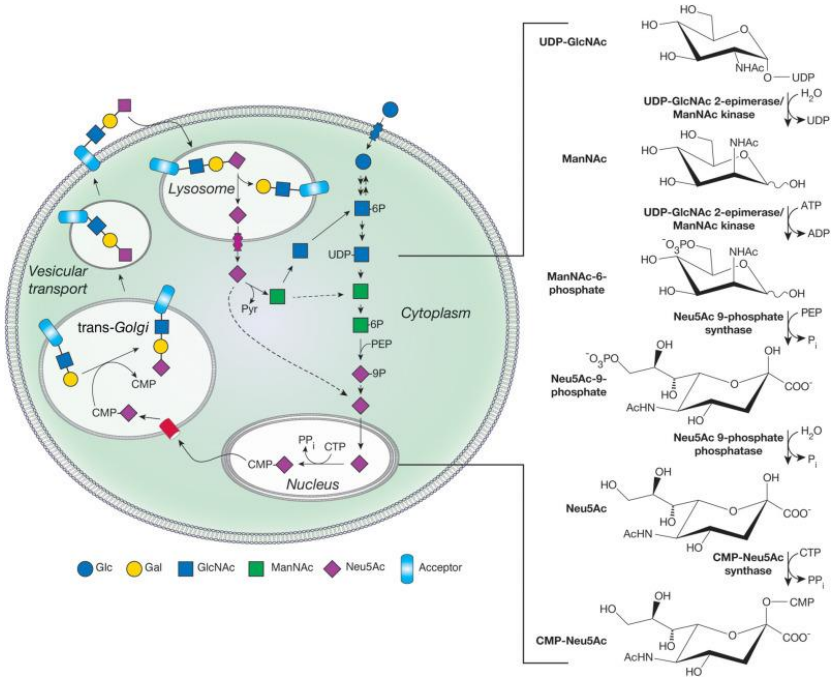


Figure 2. Metabolic pathway of N-acetyl neuraminic acid in vertebrate's cells. On the right column, a schematic representation of CMP-Neu5Ac biosynthesis from UDP-GlcNAc. The enzymes involved are highlighted in each step. On the left, a representation of the metabolism of Neu5Ac in the cell leading to sialoglycans presentation on the cell surface and further degradation. Reprinted from Essentials of Glycobiology, Chapter 15: Sialic Acids and Other Nonulosonic Acids (online version) by Varki A, Cummings RD, Esko JD, et al., 2022, Cold Spring Harbour, NY. Copyright © 2022 by The Consortium of Glycobiology Editors, La Jolla, California.

In humans, sialylation of different glycoconjugates is performed by 20 different sialyltransferases (STs)[11], a group of enzymes that transfer sialic acid from CMP-sialic acid to non-reducing ends of oligosaccharide chains in glycolipids and glycoproteins. These enzymes catalyze the formation of different types of glycosidic bonds, adding up to the diversity and complexity of the sialoglycans world. In addition to different glycosidic linkages, enzymes of the ST family are also "acceptor-specific", since they're able to discriminate in between different terminal carbohydrate sequences to be sialylated[1,11,12].

Typically, sialic acids are linked to galactose (Gal) or N-acetyl galactosamine (GalNAc) residues through α2,3 or α2,6 glycosidic bonds, as well as to other sialic acid residues through an α2,8 glycosidic linkage. Far less common α2,9 glycosidic linkages have

also been found in nature in linear sialic acid homopolymers or in linear copolymers with alternating α2,8- and α2,9-linked sialic acids[13–16]. Synthesis of each of these sialoglycans is catalyzed by different STs, that are generally categorized in 4 different families. The ST3Gal family, comprise enzymes that catalyze the transfer of sialic acid to a terminal galactose of N- and O- linked glycans and glycolipids forming an α2,3 glycosidic linkage; the ST6Gal family, with enzymes that add sialic acid to galactose residues of N-glycans with an α2,6 linkage; the ST6GalNAc family with enzymes that promote the addition of sialic acid to terminal N-acetyl galactosamine residues (GalNAc) in an α2,6 fashion; while enzymes from the ST8Sia family catalyze the formation of an α2,8 glycosidic bond between two sialic acid residues in N- and O- glycans[17,18].

Sialic acid-binding proteins

Given their diversity and terminal location in glycoproteins and glycolipids, sialic acids serve as ligands for a variety of lectins, often referred to as sialic acid binding proteins. These receptors play a crucial role in biological processes such as cell-cell interaction, regulation of immune system and pathogen recognition.

Lectins with remarkable degrees of specificity towards different types of sialic acids and their glycosidic linkages, have been discovered in different organisms unable to produce sialic acid. This phenomenon is not strange for organisms such as bacteria or viruses, whose sialic-acid binding proteins mediate infection processes. Nevertheless, the majority of sialic acid-binding proteins are found in vertebrates that express sialic acid endogenously and play important roles in regulating immune responses, cell-cell adhesion, and cell and pathogen recognition. The most widely known sialic acid-binding proteins are the previously mentioned sialyltransferases, as well as sialidases, selectins, some adhesins and siglecs. All these proteins belong to different families and structural classes, and as so, are involved in diverse physiological functions, yet they all share the capability of binding sialic acid[19].

Siglec family

Among the endogenous sialic acid-binding proteins in humans are those comprising the 15-member Siglec family. The sialic acid-binding immunoglobulin (Ig)-like lectins (Siglecs) are a group of transmembrane proteins that share a common N-terminal carbohydrate recognition domain (CRD), also known as V-set domain, able to recognize sialic acid containing glycoproteins and glycolipids. They also bear a variable number (1-16) of extracellular C2-set immunoglobulin domains, as well as cytoplasmic domains that contain different regulatory motifs responsible for their cell-signaling functions (Figure 3)[20–22].

They are usually classified in two different sub-categories depending on their evolutionary conservation and sequence similarity into conserved and CD33-related Siglecs. The evolutionary conserved Siglecs among all mammals include sialoadhesin (Siglec-1), CD22 (Siglec-2), myelin associated glycoprotein (MAG, Siglec-4) and Siglec-15. The remaining CD33 (Siglec-3), Siglec-5, -6, -7, -8, -9, -10, -11, -14 and -16 are part of the CD33-related category[6].

Not included in the category is Siglec-13 which, even though expressed in chimpanzees and baboons, is specifically deleted in humans[12,13]. Siglec-12, on the other hand, is expressed in humans but has lost the ability to bind sialic acid due to an R122C mutation in its binding site[23]. Interestingly, mutating C122 back to an arginine residue restores the capacity of the protein to bind sialic acid. Until today, there is debate on whether to include Siglec-12 in the human Siglec family and is always denoted as Siglec-XII to differentiate it from the non-mutated Siglec-12 in other mammals[24].

Except for sialoadhesin (Siglec-1), all the Siglecs bear structural features of regulatory motifs involved in cell-signaling, which can act as inhibitors or activators of immune responses. Activating signals arise from the immunoreceptor tyrosine-based activation motifs (ITAM). However, most Siglecs also contain one or more consensus immunoreceptor tyrosine-based inhibitory motifs (ITIM), including the classic ITIM motif (I/V/LxYxxL/V), and ITIM-like motif (D/E/xYxEC/IK/R), as well as the inhibitory tyrosine switch motif (ITSM, TxYxxV/I), that can participate in either inhibitory and activatory responses[20,25].

Upon sialoglycans binding, the tyrosine in the ITIM motif gets phosphorylated by Src kinases, which generate binding sites for the Src-homology 2 domain (SH2), present in the SHP-1 and SHP-2 phosphatases. This phosphatases dephosphorylate the signaling molecules in the ITAM motif, suppressing different activatory pathways, modulating different physiological responses, depending on the Siglec and the cell-type[20,26–28].

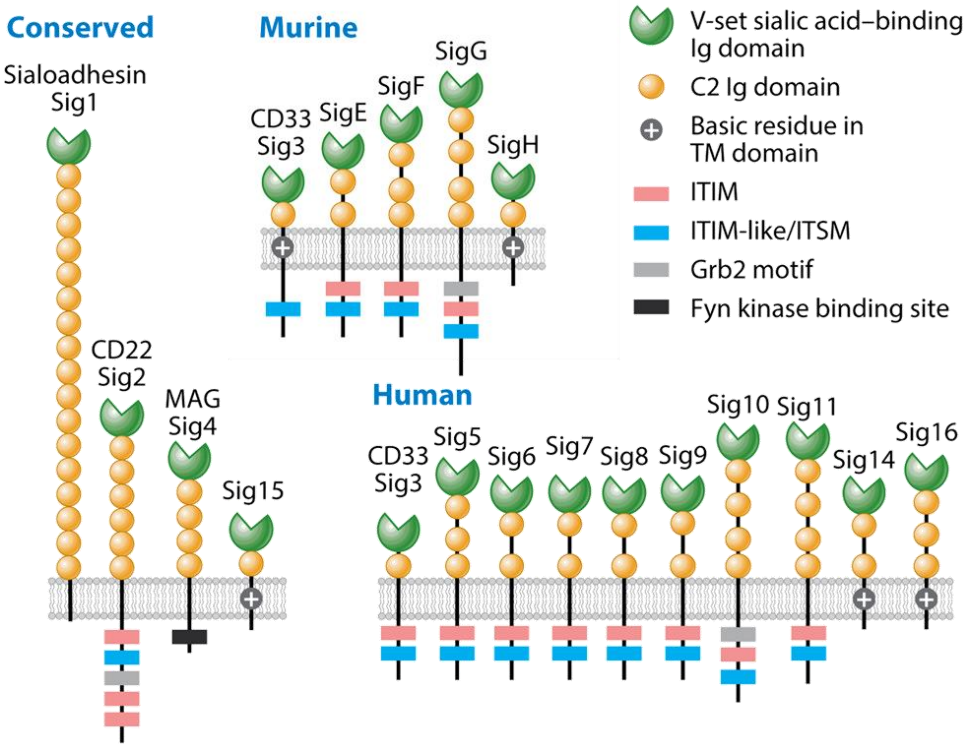


Figure 3. Human and murine receptors from the Siglec family, including the conserved members across mammals, as well as the 4 murine and 10 human CD33-related Siglecs and their structural features. Adapted from "Siglecs as Immune Checkpoints in Disease", by S. Duan and J. C. Paulson, 2020, *Annu. Rev. Immunol.*, 38, 365-395.Copyright © 2020 by Annual Reviews.

Siglecs have been found to be expressed in hematopoietic, as well as in most cells from the human immune systems, including several myeloid and lymphoid cells, with myelin associated glycoprotein (Siglec-4) being the only exception. Siglec-4 is instead found on oligodendrocytes and Schwann cells from the central and peripheral nervous system[29,30].

Some of the Siglecs are cell-type specific and carry out a particular function in a cell, others can be expressed on a larger variety of cells and perform a broader range of functions. Siglec-1, for example, has been found exclusively in macrophages of the spleen, liver, lymph node, bone marrow, colon and lungs and in certain subsets of dendritic cells[31,32], and play a role in pathogen recognition[33] and in B-cell activation[34,35]. The other cell-type specific Siglec is Siglec-2, which is expressed on B-cells, where it regulates their function[36]. Siglec-7 and Siglec-9, on the other hand, can be found in a variety of cells such as monocytes, macrophages, dendritic cells, neutrophils, T cells (Siglec-9) and NK-cells. Their functions can be as broad as pathogen recognition and phagocytosis, antigen presentation, modulation of T-cell responses, regulation of immune cell surveillance and inhibition of NK-cell cytotoxicity[37,38].

Siglecs ligand specificity

Another remarkable feature of the Siglec proteins is their capacity to recognize a wide variety of sialic acid-capped glycoconjugates with a unique yet overlapping specificity. This feature defines the function of each of the proteins of the family in the context of the specific cells they are expressed on. This ligand specificity is determined by the structure of the sialic acid residue, the type of glycosidic linkage to the underlying glycan and the structure of the underlying glycan chain, as mentioned above. Figure 4 illustrates the types of the main sialoglycans that have been shown to bind the different Siglecs, although the binding capability can even be broader for some of them and bind several sialoglycans with a different range of affinities[39].

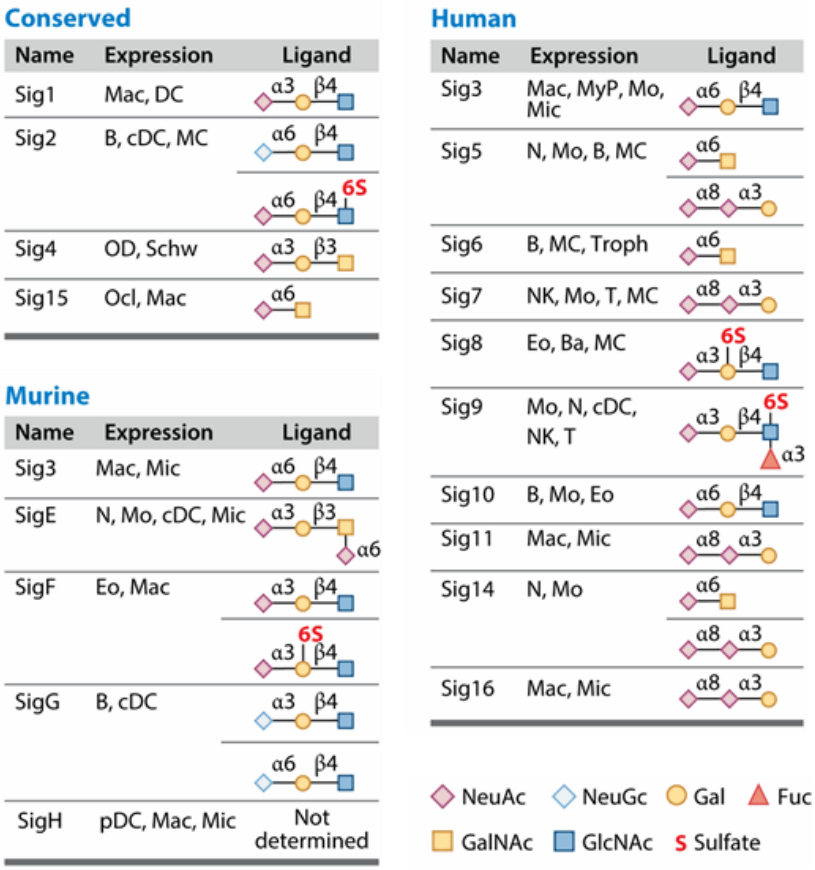


Figure 4. Cell types expressing each Siglec, with their preferred natural sialoglycans. Most siglecs naturally bind other types of sialosides with lower affinity, not shown in the figure. Cell types are abbreviated as B cells (B), basophils (Ba), conventional and plasmacytoid dendritic cells (cDC and pDC), eosinophils (Eo), macrophagues (Mac), mast cells (MC), microglia (Mic), monocytes (Mo), natural killer cells (NK), neutrophils (N), osteoclasts (Ocl), T cells (T), oligodendrocytes (OD) Schwann cells (Sch) and placental trophoblasts (troph). Adapted from "Siglecs as Immune Checkpoints in Disease", by S. Duan and J. C. Paulson, 2020, *Annu. Rev. Immunol.*, 38, 365-395.Copyright © 2020 by Annual Reviews.

Binding in Siglecs, as in most proteins, is greatly determined by its structural architecture. A shallow cavity on the N-terminal CRD acts as the sialic acid binding site. It contains a highly conserved arginine residue that mutagenesis experiments have confirmed to be essential for sialoglycan binding. This arginine is present in all the Siglecs, except for the previously mentioned Siglec-XII. At physiological pH, Siglecs interact with sialic acid

through a salt-bridge formed in between the guanidinium cation on its side chain and the negatively charged carboxylate group in C-1 position of the sialic acid[36].

Even though the CRD in Siglecs is highly conserved, besides the essential arginine, the residues in the binding site can differ from one Siglec to another, shaping the structure, charge and chemical composition of the active site for each of them. Furthermore, the binding preference is also determined by the different loops located in close proximity to the active site of the protein. These loops tend to be variable regions among different Siglecs and can play important roles that goes from masking the sialic acid binding sites for specific ligands, to shaping the active site of the protein. These differences in the binding sites account for the glycan selectivity and specificity and are crucial for their biological functions.

Furthermore, Siglec ligands are generally sialylated glycans presented in a multivalent fashion on the cell surfaces or pathogens, and hence, avidity has been shown to be an important factor that influences sialoglycan recognition. The binding affinities measured in non-multivalent systems, for the Siglec interactions, have been reported to be relatively weak, ranging from 100 μM to 3 mM[40]. Different studies have acknowledged the important role of multivalency in modulating ligand binding affinities to Siglec receptors[41–43].

Sialic acid-binding immunoglobulin-like lectin - 7

p75/AIRM1, also referred to as Siglec-7, is a member of the human CD33-related Siglecs that was almost contemporaneously cloned for the first time by the groups of Alessandro and Lorenzo Moretta, Paul Crocker and Ajit Varki, who published their findings in September 1999[44], November 1999[45] and April 2000[46], respectively. It is the main Siglec expressed on NK cells and is composed of an N-terminal (V-set) CRD domain, which contains the active site of the protein; two C2-set Ig domains that are thought to play a structural role, projecting the sialic acid binding site away from the membrane surface; and two ITIM and ITIM-like motifs in its cytosolic region, that make Siglec-7 an inhibitory receptor[37,44–46].

Before the discovery of Siglec-7, Siglec's binding capabilities were restricted to α2,3- and α2,6-linked sialic acids. Instead, Siglec-7 has unique binding preferences towards α2,8 disialylated or polysialylated molecules, as well as to α2,3 and α2,6-branched sialoglycans. Siglec-7 activity towards linear α2,3- and α2,6-linked sialylated glycoconjugates, as well as to free sialic acid has been reported to be either very weak or none.

The preferred ligands known so far for Siglec-7 are disialylated gangliosides that exhibit an α2,8 glycosidic bond in between the sialic acid residues. Gangliosides, which are sialic acid-containing glycosphingolipids, are endogenous Siglec-7 ligands. These are structurally composed of a hydrophobic lipid portion embedded in the cell membrane, known as ceramide, and an extracellular carbohydrate moiety exposed for recognition. Ganglioside recognition by Siglec-7 is essential for the biological roles of the protein in cell-cell communication, modulation of immune responses and host-pathogen interaction and recognition[46,47].

Structure of Siglec-7

The CRD of eight out of the fourteen Siglec structures have been determined so far, and even though Siglec-1[48], -7[49] and -5[50] saw their crystal structures elucidated on the earlier days of the Siglecs research (1998, 2003 and 2008, respectively), most of the structures remained unknown until the last eight years. In 2016, the X-ray structure of Siglec-4[51], as well as an NMR-based structure for Siglec-8[52] were published. The crystal structure of Siglec-2[53], -3[54] and 15[55], as well as the first X-ray structure of Siglec-8[56], followed in 2017, 2019, 2023 and 2022, respectively.

The elucidation of the structure of the Siglecs is fundamental to further understand the molecular basis for glycan recognition that govern their functions and provides new opportunities for the development of novel treatments for cancer, autoimmune disease and infectious diseases in which Siglecs are involved. These outbreaks in structure elucidation of the different members of the family, in fact, are a clear effect of the recent discoveries of the many biological roles of these proteins and their potential as therapeutic targets.

The structure of Siglec-7's CRD is a typical Ig-like fold, comprised of a β-sandwich formed by two antiparallel β-sheets made of strands A'GFCC' and ABED, as shown in Figure 5, that are stabilized by disulfide bonds. The strands are connected by the flexible A-B, B-C, C-C', C'-E, E-F, F-G and G-A loops, and some of them are particularly important in shaping the binding site of the protein[49].

The sialoglycan binding site is exposed to the solvent and is located in the A'GFCC' β-sheet, which contains the conserved arginine R124. This residue is key for sialoglycans binding and interacts with the carboxylate group of the sialic acid through a salt-bridge, as previously described. The broader ligand specificity of Siglec-7 is attributed to the B-C, C-C' and F-G loops located in proximity to the active site, that exhibit some degree of flexibility that allows the binding site to adapt to different types of sialoglycans.

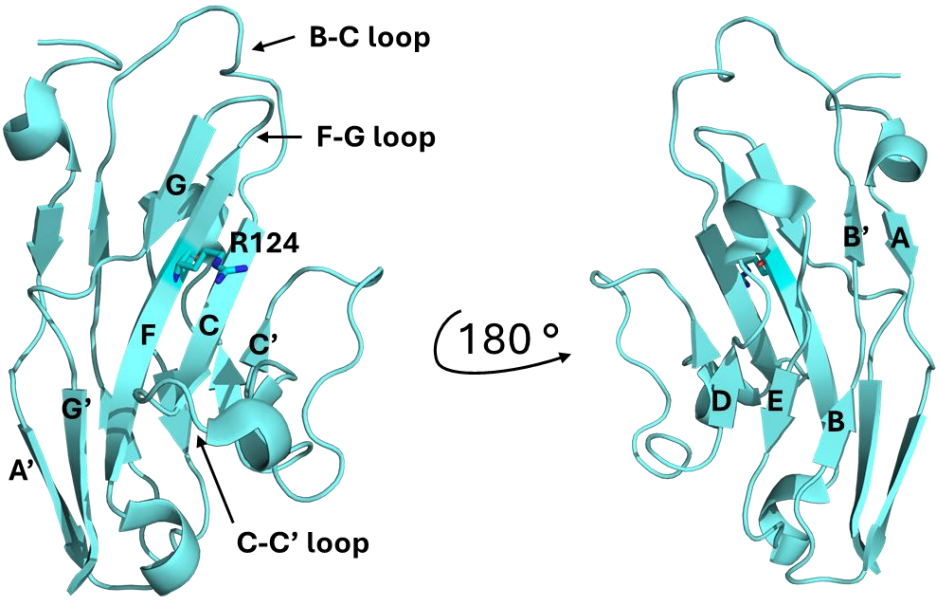


Figure 5. Crystal structure of Siglec-7 V-set Ig domain. The backbone of the protein is shown are ribbons and the conserved arginine in the active site of the protein (R124) is shown as sticks. Both faces of the a β-sandwich formed by the two antiparallel β-sheets made of strands A'GFCC' and ABED are indicated, as well as the loops shaping the binding site of the protein, B-C, C-C' and G-F. (PDB code 1O7V).

The research group of Prof. Chihiro Sato recently published the discovery of a second binding site, where R67 is the essential arginine for sialoglycans binding. The residue is

located at the end of the C β-strand and binding is thought to be highly influenced by the first binding site of the protein, controlled by the flexible C-C' loop[47].

Siglec-7 as a therapeutic target

Sigelec-7 is now considered an important immune checkpoint and has recently become an important therapeutic target in several contexts, such as cancer immunotherapy, as well as bacterial and viral infections.

Upregulation of sialic acid (known as hypersialylation) on the surface of cancer cells is a well-known phenomenon and Siglec-7 engagement helps cancer cells to escape NK-cell surveillance[57]. As a result, Siglec-7 has gained attention as a target molecule for immunotherapy in cancer. Different studies have shown that anti-Siglec-7 mAb inhibits the proliferation of chronic myeloid leukemia cells in vitro, as well as the growth of mast cells in vitro and in mice models[58], further confirming the potential of Siglec-7 as a treatment for cancer.

A therapeutic application that is currently being explored for Siglec-7, is its potential in enhancing NK cell activity, reduced when targeted by hypersialylated tumors or pathogens. The rationale is that the blocking of Siglec-7 on NK cells would prevent its inhibitory responses, enhancing NK cell cytotoxicity[57,59]. Strategies that employ monoclonal antibodies and small molecules are currently being investigated, however, the little understanding on Siglec-7 recognition of different glycans is still a limitation. This strategy has also been proposed in combination with blocking other immune checkpoints, such as PD1 and CTLA4 in order to synergistically enhance antitumor responses[60].

Besides its role in cancer, Siglec-7 has also been shown to be targeted by several bacteria and viruses such as *Fusobacterium nucelatum (F. nucleatum)* and HIV to downregulate immune responses. This results in an inhibitory response from Siglec-7, preventing NK cell responses during bacterial of viral infections[61–64]. Ongoing research, such as that reported in this thesis and manuscript, is being carried out in order to understand the structural basis for molecular recognition of Siglec-7 by pathogens.

The role of Siglec-7 in *Fusobacterium nucleatum* infections

Tumors that arise at the epithelial barrier surfaces are known to host high concentrations of microorganisms that alter and shape the way the tumor is recognized by the immune system. In certain types of colorectal cancer (CRC), *F. nucleatum* has been shown to be one of the predominant species on the cancer tissue[65,66], where it modulates the tumor microenvironment, and promotes cancer progression and metastasis. In fact, patients with *F. nucleatum*-associated carcinoma have been reported to have a shorter survival period[67], as well as higher levels of chemotherapy resistance[68,69].

Different strains of *F. nucleatum* have been shown to bind Siglec-7 using whole bacteria, *F.* nucleatum-derived outer membrane vesicles (OMVs) and bacterial lipopolysaccharide (LPS). These results strongly suggest Siglec-7 as the immunoreceptor targeted by the bacteria[63]. Interestingly enough, the LPS of one of the strains demonstrated to bind Siglec-7, *F. nucleatum ATCC 51191,* does not contain any terminal sialic acid residue. Treatment of the bacterial surface with sialidases does not abolish the binding, further confirming that the interaction is not mediated by sialic acid. STD NMR experiments further confirmed the binding of non-sialylated LPS from *ATCC 51191* to Siglec-7 in what is still one of the very isolated cases of non-sialylated ligands for Siglec-7[63].

F. nucletum ssp polymorphum 10953 is another strain highly abundant in CRC that has been shown to bind to Siglec-7. The structure of the repetitive unit of the LPS O-chain expressed on the surface of this strain is composed of N-acetyl neuraminic acid, galactose and 4N-acetylfucosamine residues. Even though *ssp polymorphum 10953* is able to synthesize sialic acid, the glycosidic linkages found in between the sialic acid residue and the neighboring monosaccharides of the LPS have not been reported previously to be recognized by Siglec-7, with a repetitive unit of [→4)-α-Neu*p*5Ac-(2→4)-β-Gal*p*-(1→3)-α-Fuc*p*NAc4NAc-(1→][70]. In fact, the molecular basis of Siglec-7 recognition of *F. nucleatum's surface*-exposed glycans remains a mystery and is addressed in this thesis work.

Bacterial siglec-like binding regions

Even though Siglec expression is restricted to eukaryotic organisms, other sialic acid binding proteins are found in prokaryotes. In particular, certain pathogenic bacteria that use sialic acid binding proteins to interact with hosts cells by recognizing sialylated glycoconjugates have been found to express these proteins. One of the most common examples are sialic acid binding adhesins (not to be confused with the mammalian sialoadhesin, Siglec-1).

Bacteria such as *Streptococcus gordonii, Streptococcus sanguinis* and *Streptococcus oralis* are a group of bacteria commensals to the oral microbiota, commonly found in patients with Infective Endocarditis (IE), a bacterial infection that affects the inner layer of the heart. The different strains of *Streptococcus* enter the bloodstream as a result of lesions in the oral epithelium, from where it can reach, and eventually infect different organs[71].

It is believed that adherence of the different strains of oral streptococci to the platelets in the bloodstream constitutes a crucial step in the pathogenesis of IE, in a process mediated by serine-rich repeat glycoproteins (SSRPs), anchored to the bacterial cell wall[72]. SRRPs are a group of proteins that belong to the family of bacterial sialic acid binding adhesins, known for their role in mediating adherence to the host's sialylated glycoproteins and in tissue colonization[72,73].

SRRP's structure includes an N-terminal signal peptide, a sialic acid binding region, known as siglec-like binding region (SLBR), located in between SRR1 and SRR2 regions and a C-terminus LPXTG motif anchored to the bacterial cell wall. In contrast to the Siglec proteins, the SLBRs are made up of two domains that are crucial for sialoglycans recognition: the Siglec domain, and the termed 'adjacent unique domain', involved in the modulation of the conformation of the first one. The most common ligands for SLRBs are α2,3-linked sialic acid-galactose residues located at the end of O-glycans, and normally found on the mucin and mucin-like glycoproteins. However, as in the Siglec proteins, different SLBRs exhibit different degrees of specificity towards different ligands, depending on the structure of the underlying glycan[71,74].

Among the few known members of SLBR family, the roles of SLBR-B and SLBR-H in pathogenicity have been widely studied both in vitro and in animal models of IE. SLBR-B has been proved to be selective for sialyl T-antigen (sTa) and SLBR-H exhibits a broader specificity towards several linear ligands, such as sTa, 3-sialylactosamine (3'SLn) and sialyl-Lewis C and their binding modes had been previously described[74]. In contrast with SLBR-B and SLBR-H and despite a sequence identity of 80%, Siglec-like binding region-N (SLBR-N) from *Streptococcus gordonii* strain UB10712, is characterized for recognizing fucosylated and sulfated derivates of 3'SLn, as well as a branched disialylated core 2 O-glycans. However, the structural basis for this recognition had not been investigated yet, until the studies reported in this thesis[74].

1.2 NMR for the characterization of Drug Delivery Systems

Erythrocytes as drug delivery systems

The development of drug delivery systems (DDS) is of great biomedical importance and has been a research area for several decades already. Optimization of DDS is critical for enhancing the efficacy and safety of traditional methods, that are normally limited by poor biocompatibility, poor biodegradability, rapid clearance, systemic side effects, and expensive costs[75,76]. The application of biological carriers, based on natural cells and their derivates is rapidly growing, and exploits the idea of combining their physiological and biophysical characteristics to overcome those limitations, with a particular focus on biocompatibility.

Erythrocytes, or red blood cells (RBC) are the most abundant cells in the blood and have a great potential in the field of DDS due to their unique biological and biophysical properties[77]. Indeed, they have long been proposed as vehicles for drug delivery because of their biocompatibility, as well as low immunogenicity, being native to the organism and hence, recognized as self-component by the immune system. Moreover, RBCs are particularly suited for drug delivery given their concave shape and lack of organelles,

allowing for the encapsulation in their entire inner volume of different therapeutic agents that range from small molecules to large multimeric proteins[78].

In addition, erythrocytes offer the advantage of an extended circulation time of around 120 days in humans. This prolonged circulation, which is considerably longer than the one of many synthetic drug carriers, allows for a sustained activity or release of a therapeutic agent over time, reducing the need for a frequent dose supply; this is particularly beneficial for chronic conditions requiring long-term treatments. Furthermore, RBCs have also been engineered to achieve a targeted delivery, both through modifications on the cell surface that allow the binding of specific receptors or by loading them with stimuli-responsive molecules that release their cargo in response to specific environmental triggers, such as pH changes.

The extensive research, and the above mentioned properties, are turning RBC into an increasingly relevant system as a drug carrier with current application in the biomedical field that range from immunotherapy to bioimaging[79–82].

Methods for encapsulation of biologics in Red Blood Cells

Two general approaches have been proposed for the loading of RBC. The most common one is the *ex-vivo* encapsulation of different therapeutic agents, however the attachment of the drug to the erythrocyte's membrane both *ex-vivo* and *in-vivo* have also been widely investigated.

Conjugation to the surface of erythrocytes represents an easier approach for the loading of RBC compared to encapsulation, and provides several advantages such as targeted delivery, rapid action of the conjugated drug and a less invasive method, since it doesn't require the disruption of the erythrocyte's membrane. However, this technique is often limited by the limited drug-load, shorter circulation time and mainly because its reduced biocompatibility gained by employing RBCs as carriers. When encapsulated, instead, drugs are isolated from the organism and can interact only with molecules that diffuse through the RBC's membrane, but not with molecules outside of the RBC, thus

greatly reducing immune reactions not only towards the drug carrier, but towards the drug itself.

Several methods have been developed for the erythrocyte encapsulation of therapeutic agents, of which, hypotonic dialysis is the most widely used. Other methods include hypotonic hemolysis, hypotonic pre-swelling, isotonic osmotic lysis, chemical perturbation of the membrane, entrapment by endocytosis and electrophoresis.

Hypotonic dialysis takes advantage of the RBC's ability to undergo reversible swelling in response to osmotic changes in the environment due to their shape and flexibility. In this method, the cells are first subjected to a hypotonic solution, causing water to enter the erythrocytes. This led to the swelling of the cells and the opening of pores in their membrane. These pores have been demonstrated to be big enough to allow the entrance of different therapeutic agents such as small molecules, glycosaminoglycans, nucleic acids and proteins. After drug loading, the erythrocytes are returned to an isotonic environment, causing the membrane to reseal and trap the therapeutic agents inside.

Therapeutic enzymes encapsulated in Red Blood Cells: the examples of ANSII, GS and ADA

Therapeutic enzymes are among the most promising candidates to be encapsulated in RBC, offering innovative solutions for enzyme replacement therapy, prolonging the half-life of the enzyme while maintaining the therapeutic blood levels. Some of the most prominent examples are the encapsulation of asparaginase for the treatment of acute lymphoblastic leukemia (ALL), glutamine synthetase (GS) for ammonia detoxification and adenosine deaminase (ADA) for the treatment of immunodeficiency. Other examples of erythrocyte-encapsulated enzymes and small molecules are discussed in different reviews by Godfrin *et al.*[83]*, Bax*[84] *and* Rossi *et al*[85].

The use of L-asparaginase type-II (ANSII) as a therapeutic agent for the treatment of ALL relies on the fact that leukemic cells are deficient in asparagine synthetase (ASNS), preventing them from synthesizing L-asparagine. ANSII is an enzyme that catalyzes the conversion of L-asparagine (L-Asn) into L-aspartate (L-Asp), as shown in Figure 6, hence,

killing the cell by depriving it from the metabolite. However, L-asparaginase treatment of ALL has been shown to produce several side effects, such as hypersensitivity reactions, coagulation disorders and hyperglycemia, particularly in patients with pre-existing diabetes. Encapsulation of L-asparaginase in RBC has then been proposed to decrease the secondary effects by limiting the increase of concentration of in-vivo L-asparaginase[86–88].

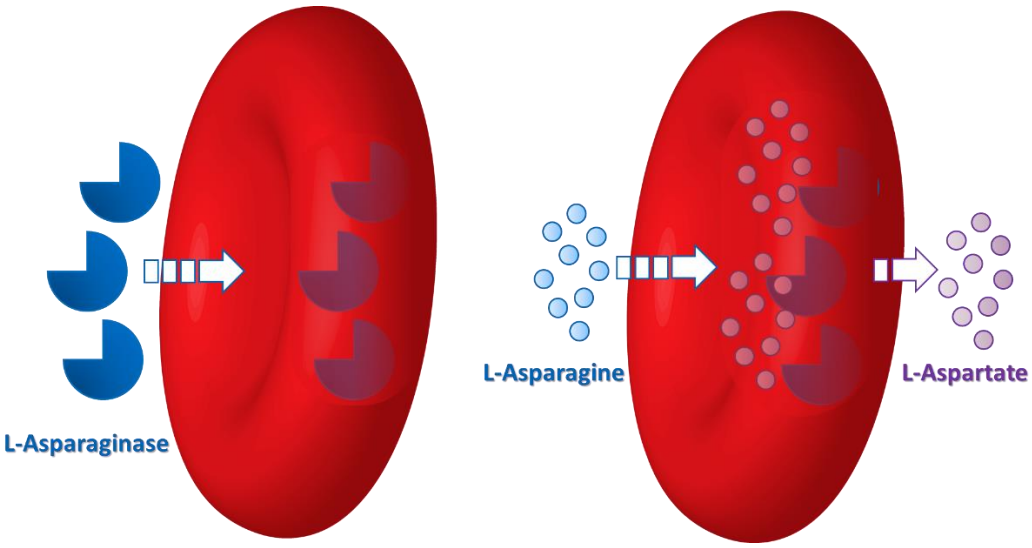


Figure 6. Schematic representation of the mechanism of action of erythrocyte-encapsulated L-asparaginase in the conversion of L-asparagine to L-aspartate for the treatment of acute lymphoblastic leukemia.

Another example of an enzyme encapsulated in RBCs is the glutamine synthetase for the detoxification of ammonia, since high concentration of the metabolite can lead functional disorders in the central nervous system. This approach has been proved to be effective, while keeping most metabolites and metabolic parameters stable[74,75].

Moreover, adenosine deaminase (ADA) deficiency results in the accumulation of 2-deoxyadenosine, which impairs lymphocyte differentiation, proliferation and function, and cause severe immunodeficiency. Treatment of patients with an erythrocyte-encapsulated ADA has been reported with very promising results[91,92].

Current methods for characterization erythrocyte-encapsulated enzymes

Characterization of biologics is essential to ensure their stability and functionality within a pharmaceutical formulation. Several techniques have been used to address the encapsulation of enzymes into red blood cells. Flow cytometry for example, has been used to evaluate the encapsulation efficiency by identifying the proportion of cells that have incorporated the enzyme. In addition, spectrophotometric and HPLC methods are used to quantify the enzyme concentration within the cells, and enzyme-specific assays are generally used to evaluate the functionality of the enzyme after their encapsulation.

Structurally-wise, available methods are restricted to the characterization of the integrity of the RBCs after the encapsulation procedure. Scanning Electron Microscopy (SEM) and Transmission Electron Microscopy (TEM) provide a detailed visualization of the any perturbation on any structural alterations. However, none of these methods are oriented towards the characterization of the preservation of the higher order structure (HOS) encapsulated enzyme. In this thesis work and manuscript (still in preparation), we introduce the assessment of the preservation of the HOS structure of erythrocyte-encapsulated ANSII. This represents the first study addressing the structural integrity of a therapeutic biomolecule encapsulated within red blood cells.

Human Transthyretin: a protein-based drug carrier

Human transthyretin (TTR) is a vital transport protein found in both blood plasma and cerebrospinal fluid, where it plays a key role in shuttling thyroid hormone thyroxine (T4) and retinol (vitamin A) via the retinol-binding protein. Structurally, TTR is composed of four identical subunits, forming a stable tetramer with D2 symmetry and a total molecular mass of 55 kDa. This tetrameric assembly is crucial for its normal function, but mutations in the TTR gene can destabilize the tetramer, leading to its dissociation into monomers. These monomers are prone to misfolding and aggregation, resulting in the formation of amyloid fibrils that deposit in various tissues, causing amyloidosis-related diseases, such as familial amyloid polyneuropathy (FAP) and familial amyloid cardiomyopathy (FAC). TTR's role in

health and disease has made it a significant focus of therapeutic strategies, with treatments aimed at stabilizing the tetramer and preventing amyloid formation.

Tafamidis, a small molecule that binds to the T4 binding sites with nanomolar affinity and preventing tetramer dissociation, is one such therapeutic, offering a promising approach to preventing TTR-related amyloid diseases. The high affinity of this interaction makes tafamidis an ideal anchor for poorly soluble drugs. A similar strategy, using a different high-affinity TTR ligand has been previously used for targeting cancer cells with a cytotoxic molecule.

Solid-state NMR of biomolecules is an emerging research area for the study of the structure and dynamics of biomolecules, particularly in systems where traditional solution NMR methods, X-ray crystallography and cryo-EM are limited. Especially, the capacity of analyzing non-crystalline insoluble samples, at atomic resolution turn solid-state NMR into an essential tool for the study of complex systems.

2. Aim and research topics of this thesis

Nuclear Magnetic Resonance spectroscopy is a powerful tool in structural biology for the study of the three-dimensional structure and dynamic behavior of biomolecules. It provides atomic resolution insights into molecules, both in solution and in solid-state, and allows to capture important structural features involved in essential biological processes.

This thesis focusses on the application of NMR to investigate structural and biophysical properties of sialic acid binding proteins, as well as for the characterization of drug delivery systems with the main goal of uncovering critical structural features that can be translated into therapeutic applications. During the 3 years of the Ph.D., I focused on the study of different biological systems and their interactions, with a particular emphasis on the Siglec proteins and on the study of the preservation of the Higher Order Structure of erythrocyte-encapsulated enzymes with biological interest.

Sialic acid binding proteins have recently become an important target for the treatment of different pathologies, given their role as inhibitory receptors, as well as mediators of host-pathogen interactions. Studying the structural mechanisms by which these proteins bind to their ligands allows us to deepen our understanding of how these interactions influence immune responses. This is crucial for developing targeted therapies in cancer immunotherapy and bacterial infections.

Moreover, the structural characterization of drug delivery systems is of uttermost importance for developing safe and efficient carriers. In particular, NMR is a unique and powerful technique for the characterization of biomolecules embedded in complex matrices such as the erythrocyte environment. The application of RBCs as biological drug delivery systems is rapidly growing due to their low immunogenicity and biocompatibility. Hence, the assessment of the structural integrity of erythrocyte-encapsulated biologics is critical for ensuring the preservation of their folding and stability, directly linked to their biological function.

In particular, the objectives of the research are organized to address the following scientific questions:

- To dissect the structural basis for the molecular recognition of disialylated branched sialoglycans by the Siglec-like Binding Regions N (SLBR-N).
- To determine the binding modes and affinities of different cancer-related gangliosides to Siglec-7, as well as to elucidate the molecular mechanisms of *Fusobacterium nucelatum* recognition.
- To demonstrate the potential of both solution and solid-state NMR for the characterization of biomolecules embedded in different matrices such as protein-macrocycle frameworks and the erythrocyte environment.
- To develop a protocol for the assessment of the preservation of the Higher Order Structure and semiquantitative analysis of therapeutically relevant enzymes encapsulated in red blood cells.

3. Methodological aspects

Protein NMR assignment

NMR is a widely used methodology in structural siology for the study of proteins and nucleic acids due to the advantages described in the introduction of this thesis. However, the resulting NMR spectra are complex maps of peaks, representing specific nuclei of the protein. These peaks provide key information about the structure and dynamics of a biomolecule and knowing the identity of each of them within the protein sequence is critical. Most of the time, the first step when analyzing proteins or protein interactions through NMR is, hence, the NMR assignment of its resonances.

The complexity on the NMR spectra arises from the fact the proteins are big molecules (generally larger than 7 KDa), and the large number of nuclei contributing to the NMR signals results in crowded and unresolved 1D NMR spectra. Multidimensional heteronuclear NMR is key for resolving those signals and extracting atomic-resolution information. It allows to increase the spectral resolution and to resolve overlapping signals along the different dimensions by correlating different nuclei through heteronuclear scalar couplings. However, NMR active nuclei of biologically relevant atoms other than hydrogen, such as carbon, oxygen and nitrogen, are characterized by a lower sensitivity due to a lower gyromagnetic ratio and natural isotopic abundance. In consequence, isotopic enrichment of the protein with ^{13}C and ^{15}N is required to overcome these limitations. Moreover, for relatively large proteins with a molecular weight higher than 30 KDa, deuteration is often applied for the acquisition of 3D NMR experiments for assignment.

Resonance assignment on an NMR spectrum usually involves the acquisition of a series of multidimensional NMR experiments, designed to correlate the resonances of ^{1}H, ^{13}C and ^{15}N in the backbone of the protein. In particular, the 3D NMR experiments used for a complete backbone resonance assignment in the projects reported in this thesis are 3D HNCA, 3D HN(CO)CA, 3D CBCA(CO)NH, 3D HNCACB, 3D HNCO and 3D HN(CA)CO. In these experiments, resonances from the H_N, C_α, C_β, and C_O are detected for a specific

nucleus (*i*) and its predecessor (*i-1*), allowing to establish a sequential connection from one residue to the next one, by identifying the matching resonances.

Upon backbone assignment, side-chain resonances assignment follows. This is achieved using additional 3D NMR experiments, such as HBHA(CO)NH, H(CCO)NH and CC(CO)NH which correlate the side chain ^{1}H and ^{13}C nuclei with the backbone amide signals.

NMR line shape analysis

The study of protein-ligand interactions through NMR is strongly dependent on the affinity of the interaction, and in general, are classified into three possible outcomes (considering a two-state model): slow ($K_{ex} \ll \Delta\nu$), intermediate ($K_{ex} \approx \Delta\nu$), and fast exchange ($K_{ex} \gg \Delta\nu$)[93].

In the slow exchange regime, signals from both free and bound states, are observed with a particular chemical shift, intensity and linewidth for each of them. In this case, the intensity of the peak directly reports on the population of the species and typically corresponds to affinities in the nanomolar and low micromolar range.

The intermediate exchange rate is generally characterized by a line broadening, that often results in coalescence of the peaks into a single signal during the titration, until more than half of the stoichiometry is reached and the linewidth corresponding to the signal of the bound state sharpens. This, however, might not be true for all interactions at an intermediate exchange and the NMR signal of the bound component might remain broadened beyond detection. This behavior is typically observed for interactions in the micromolar range.

At the fast exchange usually one signal is observed with a chemical shift that reports on the population of each of the states as $\delta_{obs} = P_{free}\delta_{free} + P_{bound}\delta_{bound}$. This is averaging in the chemical shift is due to the rapid interconversion in between the different states[94,95].

Chemical shift perturbation (CSP) analysis is a widely used NMR-based approach for the study of molecular interactions and conformational changes in biomolecules for systems on the fast to intermediate exchange rates. This is an invaluable source of information provided by changes in the chemical environment of a specific nuclei in response to a

binding event or a structural re-arrangement. The analysis of the line-broadening and signal intensity of a peak becomes significant for interactions that occur on the intermediate to slow exchange regimes. A global analysis involving chemical shift perturbations and intensity of the NMR signal is successfully used to map the binding region of a protein, to determine protein residues involved in an interaction, as well as to extract kinetic and thermodynamic parameters.

Saturation Transfer Difference NMR

Saturation Transfer Difference (STD) is one of the most popular ligand-based NMR experiments for the study of protein-ligand interactions. Even though it's a widely used NMR-based technique for ligand screening, STD NMR has also been applied to the characterization of several systems, ranging from protein-glycan interactions to amyloid fibrils and hydrogels. It relies upon nuclear Overhauser effect (NOE), in which the signal intensity of a particular nuclei is perturbed due to cross-relaxation with another perturbed nuclei close-by in space[96–98].

The technique consists in the acquisition of a spectra on the complex with a saturation pulse, irradiating specifically at protein frequencies (on-resonance). The selective saturation is then transferred to the whole protein via spin diffusion. In case of an interaction, the saturation of the magnetization is transferred to the ligand's nuclei through dipolar cross-relaxation, resulting in the saturation of the ligand's signals arising from the nuclei located in closer proximity to the protein. A second spectra is acquired with a saturation pulse irradiated far away from any protein or ligand frequency (off-resonance), which results in a regular NMR spectrum of the ligand. By subtracting the off- and on-resonance spectra ($I_0 - I_{sat}$), the STD NMR spectrum is obtained, in which only the proton signals of the ligand's nuclei located in close proximity to the protein are observed, and any signal from a non-binder cancelled out.

Solid-state NMR

Solid-state NMR is a powerful tool for the study of the structural properties at atomic resolution of materials and biomolecules in the solid-state. Unlike solution NMR, the linewidth of the NMR signal is independent on the molecular weight of the system and allows the study of large biomolecules.[99–101]

The absence of molecular tumbling in the sample results in broad spectral line-widths due to Chemical Shift Anisotropy and dipolar couplings. Magic angle spinning (MAS) is the most common method to overcome these limitations. In MAS, the sample is spined at a 'magic angle' of 54.7° relative to the external magnetic field. This cancels out the anisotropic interactions, resulting in narrower and more resolved peaks.

Proton-detected experiments in solid-state NMR are usually not acquired under low MAS rates (usually lower than 60 kHz) due to the strong dipolar interactions with each other given their high gyromagnetic ratio. MAS rate is physically limited by the size of the rotor. Typically, 3.2 mm rotors, such as the ones used in the studies reported in this thesis are limited to a maximum spinning rate of 24 kHz, and hence the acquisition of ^{1}H-detected experiments becomes challenging. To overcome the sensitivity limitations of ^{13}C, ^{13}C-detected experiments are usually acquired combined with a cross-polarization (CP) scheme, which enhances ^{13}C signal intensity by transferring the polarization from the abundant, sensitive ^{1}H nuclei to ^{13}C, significantly improving the signal-to-noise ratio. Combined CP-MAS, provides high resolution ^{13}C spectra in solids, making it one the most common ^{13}C-detected techniques in ssNMR[102,103].

4. Experimental Section

This section contains my contributions to each of the papers and manuscripts attached in the section of "results and discussion".

4.1 Biophysical characterization of carbohydrate binding proteins

Expression and purification of Siglec-7 CRD

The plasmid containing the gene coding for Siglec-7 CRD (UniProt entry: Q9Y286) was transformed into BL21 (DE3) Codon Plus RIPL cells with chloramphenicol for antibiotic selection and cultured at 37 °C, 180 rpm in LB medium or M9 culture medium, accordingly supplemented with 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 0.2 % V/V solution Q, 1 µg/mL thiamine, 1 µg/mL biotin, 1 g $(^{15}NH_4)_2SO_4$ and 3 g $^{13}C_6H_{12}O_6$ for isotopic labeling. Overexpression was induced at OD_{600} = 0.8 with 1 mM IPTG, and further incubated for 24 hours at 25 °C and 180 rpm. The cells were harvested at 7500 rpm for 20 minutes, using a Sorvall RC 6 Plus Superspeed Centrifuge (Thermo Scientific) and then collected and resuspended in lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH = 7.4). The resuspended pellet was subsequently sonicated for 10 cycles (30 seconds ON, 2.5 minutes OFF, 70 % amplitude), using a Vibra-Cell sonicator (Sonics & Materials). The lysate was then ultracentrifuged for 40 minutes at 40000 rpm on an Optima LE-80K Ultracentrifuge (Beckman). At this point, Siglec-7 CRD was found incorporated into inclusion bodies, which were resuspended in 8 M urea lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM Imidazole, pH=7.4) and sonicated for 10 cycles as previously described.

The soluble fraction, containing unfolded Siglec-7CRD was purified using Immobilized Metal Affinity Chromatography (IMAC) on a HisTrap FF column (5 mL, GE Healthcare), previously equilibrated with 20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH = 7.4. The column was eluted at a flow rate of 4 mL/min, with a buffer high in imidazole (20 mM potassium phosphate, 500 mM NaCl, 500 mM imidazole, pH = 7.4). The protein was then refolded using a dialysis protocol in a membrane with 10 KDa MWCO. Each dialysis was performed over night on a 2 L container with the following buffers: 1) 20 mM potassium phosphate, 300 mM NaCl, 4 M urea;

2) 20 mM phosphate buffer, 300 mM NaCl, 2 M urea; 3) 20 mM potassium phosphate, 150 mM NaCl. Refolded Siglec-7 was collected from the dialysis membrane and ultracentrifuged at 40000 rpm for 40 minutes. The supernatant was filtered on a 0.2 µm micropore membrane and purified using Size Exclusion Chromatography (SEC) on a HiLoad 26/60 Superdex 75 pg (GE Healthcare) coupled on an AKTA Go FPLC system. The column was previously equilibrated with 20 mM potassium phosphate, 50 mM NaCl, pH = 7.4.

NMR experiments for backbone resonance assignment of Siglec-7

NMR experiments for backbone resonance assignment were acquired on a sample of [U-^{15}N-^{13}C] Siglec-7 at a concentration of 400 µM in 600 µL of aqueous buffered solution (20 mM potassium phosphate pH = 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM protease inhibitors, 10 % D_2O) in a 5 mM NMR tube. Triple resonance experiments for protein backbone NMR assignment HNCA, HNCACB, CACB(co)NH and HNCO were acquired at 298 K on a Bruker's Avance™ NEO 900 MHz spectrometer, equipped with TCI cryo-probe. 3D HNcaCO was recorded at 298 K on a Bruker's Avance™ NEO 500 MHz spectrometer equipped with an TCI cryoprobe. 93 % of the amino acid sequence from Y26 to T147 was assigned, excluding the 5 proline residues. Data acquisition and processing was performed on TOPSPIN 4.1.1 software and spectra were analyzed by using Computed Aided Resonance Assignment (CARA) software.

NMR titration of Siglec-7 with GD3, DGb3-α3 and DGb3-α6

2D ^{1}H-^{15}N HSQC NMR experiments were recorded on samples of [U-^{15}N] Siglec-7CRD at a concentration of 200 µM in 200 µL of aqueous buffered solution (20 mM potassium phosphate, pH = 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM protease inhibitors, 10 % D_2O) in a 3 mm NMR tube. Increasing aliquots of GD3, DGb3-α3 or DGb3-α were added in order to reach ligand concentrations of 12.5, 25, 50, 100, 200, 400, 800 ,1600 and 3200 µM. Experiments were acquired on a Bruker's AVANCE™ NEO spectrometer operating at 900 MHz (^{1}H Larmour frequency), equipped with a triple resonance TCI cryo-probe. The spectra were acquired using 32 scans, 2048 data points in the direct dimension, 128 data points in the indirect dimension, a recycle delay of 1.2 seconds and the temperature was kept at 298 K. Data

acquisition and processing were performed with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA. Chemical Shift Perturbations (CSP) were evaluated with the formula: $\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + (\frac{\Delta\delta_N}{5})^2}$, [93].

NMR titration of Siglec-7 with OPS from *Fusobacterium nucleatum ATCC 10953*

A stock solution of 3.2 mg of OPS from *F. nucleatum ATCC* 10953 (MW = 650 Da / repetitive unit, total MW = 10 KDa – 20 KDa; 15 – 30 repetitive units / OPS molecule) dissolved in 800 µL of milliQ water was prepared and the solution was aliquoted in different fractions to reach 0.075, 0.075, 0.45, 0.6, 1.2 and 1.1 mg respectively after lyophilization. Each of the lyophilized aliquots were subsequently redissolved in 200 µL of a sample of [U-15N] Siglec-7CRD 200 µM (or the previous titration point) to reach final concentrations of ligand of 25, 50, 200, 400, 800 and 1160 µM assuming an average molecular weight of 15000 KDa for the OPS. Details are shown in Table 1. Experiments were acquired on a Bruker's AVANCE™ NEO 900 MHz spectrometer equipped with a triple resonance TCI cryo-probe. The spectra were acquired using 32 scans, 2048 data points in the direct dimension, 128 data point in the indirect dimension, a recycle delay of 1.2 seconds and the temperature was kept at 298 K. Data acquisition and processing were performed with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA. Chemical Shift Perturbations (CSP) were evaluated with the formula: $\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + (\frac{\Delta\delta_N}{5})^2}$, [93].

Table 1. Experimental set for the titration of the OPS from *F. nucleatum ATCC* 10953. The first row shows the number of the sample containing the ligand. The second row shows the volumes aliquoted from the stock solution of 3.5 mg of OPS dissolved in 800 µL of milliQ water. The third row shows the amount of material (in milligrams) obtained after lyophilization on each of the samples. And fourth indicated the final concentration of the OPS after dissolving the material in 200 µL of [U-15N] Siglec-7CRD, 200 µM.

Sample	Volume lyophilized from stock sol. (µL)	Mass after lyophilization (mg)	Final conc. in NMR tube (µM)
1	17.1	0.075	25
2	17.1	0.075	50
3	102.9	0.450	200
4	137.1	0.600	400
5	274.2	1.200	800
6	251.5	1.100	1160

STD NMR experiments

Saturation Transfer Difference (STD) NMR experiments were recorded on a Bruker AVANCE™ NEO operating at 600 MHz (^{1}H Larmour frequency) equipped with a cryo-probe and data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in phosphate saline deuterated buffer (10 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NAcl, 10 mM NaN_3, pH = 7.4) at 298 K. [D4](trimethylsilyl)propionic acid, sodium salt (TSP, 1%) was used as internal reference. A protein concentration of 25 µM was used and protein:ligand molar ratio 1:50 for all mixtures.

STD NMR experiments were acquired with shaped pulse train for saturation on f2 channel alternating between on and off resonance with 20 ms spinlock pulse applied to suppress protein signals. The spectra were acquired with 104 scans and 65 K data points. The protein resonances were selectively irradiated by 40 Gauss pulses with a length of 50 ms using the off-resonance pulse frequency at 40 ppm and on-resonance pulses at 7.5 ppm and 0 ppm. The STD NMR spectra were carried out using a saturation time of 2 seconds. STD NMR effects

were determined using the formula $\frac{I_0 - I_{sat}}{I_0}$, where I_{sat} is the relative intensity of STD NMR signal and I_0 is the peak intensity on the unsaturated reference spectrum (off-resonance). The strongest STD NMR response was set to 100 % while all the other STD signals were normalized to this value to provide ligands epitope maps.

Affinity constant calculations using TITAN-NMR

Dissociation constants for the interactions of Siglec-7 with GD3, DGb3-α3 and DGb3-α6 were calculated by simulating and fitting spectra from an experimental NMR titration using NMR-TITAN software. The data was fitted to a 2-state binding model for an ^{1}H-^{15}N HSQC pulse program. 10 of the most affected residues by the interactions were used for the fitting : Y26, H62, E126, K127, G128, N129, I130, K131, W132 and N133. Bootstrap error analysis was run with 50 replicas.

HADDOCK models

Docking models of the complex were obtained using the HADDOCK 2.4 software. Protein residues affected by CSP of changes in the intensity were selected as active residues. Carbohydrate residues exhibiting an STD NMR effect were selected as active glycan residues. Clusters with the lowest HADDOCK score and highest population were selected as a base for MD simulations performed by the group of Prof. Silipo.

NMR titration of SLBR-N with sTa-Thr and 3'SLn

2D ^{1}H-^{15}N TROSY-HSQC NMR experiments were recorded on samples of [U-^{15}N] SLBR-N at a concentration of 160 µM in 200 µL of aqueous buffered solution (PBS pH = 7.4, 0.01% NaN_3, 1 mM protease inhibitors, 10 % D_2O) in a 3 mm NMR tube. Increasing aliquots of sTa-Thr or 3'SLn were added in order to reach ligand concentrations of 12.5, 25, 50, 100, 200, 400, 800 ,1600 and 3200 µM. Experiments were acquired on a Bruker's AVANCE™ NEO spectrometer operating at 1200 MHz (^{1}H Larmour frequency) equipped with a triple resonance TCI cryo-probe. The spectra were acquired using 32 scans, 2048 data points in the direct dimension, 128 data point in the indirect dimension, a recycle delay of 1.2 seconds and the temperature was kept at 298 K. Data acquisition and processing were performed

with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA. Chemical Shift Perturbations (CSP) were evaluated with the formula: $\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + (\frac{\Delta\delta_N}{5})^2}$, [93].

NMR titration of SLBR-N – sTa-Thr complex with 3'SLn for competition experiments

The interaction of SLBR-N with both ligands was also investigated in a competition experiment. 2D ^{1}H-^{15}N TROSY-HSQC NMR experiments were recorded on sample of the complex formed by 160 µM [U-^{15}N] SLBR-N and 1280 µM sTa-Thr. Increasing aliquots of 3'SLn were added in order to reach ligand concentrations of 40, 80, 160, 320, 640, 1280, 2560 µM. Experiments were acquired on a Bruker's AVANCE™ NEO spectrometer operating at 950 MHz (^{1}H Larmour frequency) equipped with a triple resonance TCI cryo-probe. The spectra were acquired using 32 scans, 2560 data points in the direct dimension, 128 data point in the indirect dimension, a recycle delay of 1.2 seconds and the temperature was kept at 298 K. Data acquisition and processing were performed with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA. Chemical Shift Perturbations (CSP) were evaluated with the formula: $\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + (\frac{\Delta\delta_N}{5})^2}$, [93].

Solid-State NMR sample preparation with the microcrystalline RSL-sclx$_8$ material

The microcystaline RSL-sclx$_8$ framework was prepared by the group of Prof. Crowley and 9.0 mg of the material were packed in a Bruker 3.2 mm rotor by using a ultra-centrifugal device (courtesy of Giotto Biotech s.r.l.) with a Beckman Coulter Optima L80K floor preparative ultracentrifuge equipped with a SW32 swinging bucket at 15,000 rpm and 4 °C. The rotor was sealed with a silicon plug (courtesy of Bruker) and the microcrystals were further packed by MAS spinning at 14 kHz.

The sample devoid of sclx8 was flash-frozen in liquid nitrogen and lyophilized. 3.4 mg of sample were manually filled into a 3.2 mm rotor with the Bruker tools. This sample was rehydrated by multiple additions of milli-Q H_2O until the maximum resolution was achieved in the 1D { 1H}13C cross-polarization spectrum.

Solid-State NMR experiments on the microcrystalline RSL-sclx$_8$ material

All the solid-state NMR spectra were acquired on a Bruker AVANCE III spectrometer operating at 800 MHz (19 T, 201.2 MHz ^{13}C Larmor frequency) and equipped with a Bruker 3.2 mm Efree NCH probe-head. 2D ^{15}N-^{13}C NCA, NCO, ^{13}C-^{13}C DARR (mixing times 15, 50 and 100 ms) and 3D NCACX, NCOCX, CANCO, NCACB, N(CO)CACB were recorded at 14 kHz. 3D CANCOCA and 2D ^{13}C-^{13}C PDSD (mixing times 25, 50, 100, 200 and 400 ms) were recorded at 20 kHz. The temperature was kept at 280 K.

4.2 NMR as a tool for the biophysical characterization of drug delivery systems

Expression and purification of Carbonic Anhydrase II

The gene coding for α-Carbonic Anhydrase II (CAII) into a pCAM vector was transformed into *E. coli* BL21 (DE3) cells, which were subsequently pre-cultured overnight in LB medium containing ampicillin (0.1 mg/mL) and 1% glucose at 37 °C and at 160 rpm. A 1L culture was then incubated at 37 °C, 160 rpm until reaching OD600 of ~0.6 and harvested for 15 minutes at 4000 rpm. The isotopic enrichment was then applied following the Marley method, in which the cell pellet was resuspended in 1L of M9 medium supplemented with $MgSO_4$, $CaCl_2$, 3 g glucose, 1,2 g of $(^{15}NH_4)_2SO_4$, 0.5 mM $ZnSO_4$ and 100 µg/mL ampicillin. The culture was incubated for 30 minutes at 37 °C and at 160 rpm, induced with 1 mM IPTG and incubated overnight at 25 °C and 160 rpm. The cell culture was harvested at 7500 rpm for 20 minutes. The cell pellet was resuspended in 70 mL of 20 mM Tris-SO_4, 0.5 mM $ZnSO_4$ pH = 8.0 and sonicated for 10 cycles of 30 seconds with a resting period of 3 minutes on ice and 60 % amplitude. The lysate was ultracentrifuged at 40000 rom for 40 minutes and the supernatant containing the crude α-CAII protein was recovered and filtered through a 0.45 µm filter. The protein was purified by nickel affinity chromatography using a linear 0-0.5 M imidazole gradiente on a HisTrap 5 mL column (GE Healthcare). The protein was then

subjected to a size-exclusion chromatography (SEC) on a Suoerdex 75pg 26/60 column (Amersham Biosciences) in 50 mM sodium phosphate buffer, pH = 7.0.

Washing of Red Blood Cells

Commercially available Bovine Red Blood Cells Packed 100% from Innovative Research, Inc., were used. RBCs were washed by diluting 5 mL of 100% packed RBC in 5 mL of cold PBS. The sample was centrifuged at 1000 g for 3 min at 4 °C and the supernatant removed. The process was repeated three times.

Erythrocyte-encapsulation of α-CAII

α-CAII, TTR and ANSII were encapsulated using an adaptation of a previously reported hypotonic dialysis protocol[104]. 700 uL of washed packed RBCs were mixed with 300 µL of a concentrated solution of the protein (3 mM [U-^{15}N] α-CAII, 1.2 mM [U-^{15}N] TTR and 0.9 mM [U- ^{2}H, ^{13}C, ^{15}N] ANSII). The 1 mL of cell suspension was placed on a dialysis tube with 12 KDa MWCO and the sample dialyzed against 150 mL of the hypotonic buffer (5 mM KH_2PO_4, 5 mM K_2HPO_4, pH = 7.4) at 4°C for 3 hours to induce swelling of the RBCs. The erythrocytes were then resealed by transferring the dialysis tube into a container holding 150 mL of the isotonic buffer (PBS 1X, 5 mM glucose, 5 mM MgCl2) at 37 °C and dialyzing for 1 hour. Non encapsulated ^{15}N isotopically enriched protein or released components from the erythrocytes were then removed by transferring the sample to a dialysis tube with a MWCO of 1000 KDa and dialyzing against 1L of isotonic buffer for 12 hours at 4 °C for three times. Finally, the dialyzed erythrocytes were washed with an equal volume of cold PBS three times before acquisition of NMR experiments and a 1D ^{15}N-filtered NMR experiments were acquired to confirm the absence of labeled protein in the supernatant.

Control experiments

Two different control experiments were prepared 1) to rule out the possibility of having the protein absorbed on the surface of the RBCs and to confirm the efficacy of the washing procedures after encapsulation, as well as to 2) measure the background signal coming from the natural abundance ^{15}N isotopes of the proteins inside RBCs.

Control 1 was prepared through a modification of the previously described previously described encapsulation protocol in which the initial hypotonic dialysis was substituted by an isotonic dialysis, hence, preventing encapsulation of the protein. Control 2 was prepared following the encapsulation protocol without the addition of the isotopically enriched protein. Both control experiments were performed in parallel with the experiment of encapsulation of the respective isotopically enriched proteins.

NMR sample preparation and NMR experiments on erythrocyte encapsulated proteins

Samples for NMR were prepared with 540 µL of the protein-containing erythrocyte suspension and 60 µL of PBS 1X in D_2O. ^{1}H-^{15}N sofast-HMQC experiments were acquired with an H_N offset of 8.1 ppm, an excitation bandwidth of 4 ppm, 256 scans ,1.5 K datapoints in the direct dimension and 128 datapoints in the indirect dimension. NMR experiments were recorded on a Bruker's AVANCE™ NEO spectrometer operating at 900 MHz (^{1}H Larmor frequency) equipped with a TCI cryo-probe, at 310 K. Data acquisition and processing were performed with TOPSPIN 4.1.1 software.

Solution NMR experiments for backbone assignment of free-TTR

Solution NMR experiments for backbone resonance assignment were acquired on a sample of [U-^{2}H, ^{13}C, ^{15}N] samples of native TTR at a concentration of 500 µM of the monomer in 600 µL of aqueous buffered solution (50 mM MES pH= 6.5, 100 mM NaCl, 5 mM DTT, 0.1 % NaN_3, 1 mM protease inhibitors, 10 % D_2O) in a 5 mM NMR tube. Triple resonance experiments for protein backbone NMR assignment with a TROSY scheme[105] [3D tr-HNCA, tr-HNCACB, tr-HNCO, tr-HN(ca)CO] were acquired at 310 K on a Bruker's ACANCE MHD operating at 950 MHz (^{1}H Larmour frequency) equipped with a triple resonance cryoprobe. For 3D tr-HNCACB and tr-HN(ca)CO nonuniform sampling (NUS) at 57% and 25%, respectively, compressed sensing reconstruction were used. Data acquisition and processing was performed on TOPSPIN 4.1.1 software and spectra were analyzed by using CARA software.

TTR – tafamidis and TTR – Taf-Ptx complex formation monitored through solution NMR

For the titration with Tafamidis, 2D ^{1}H-^{15}N TROSY-HSQC NMR experiments were recorded on samples of [U-^{13}C-^{15}N] TTR at a concentration of 800 µM of the monomer in aqueous buffered solution (50 mM MES pH= 6.5, 100 mM NaCl, 5 mM DTT, 0.1 % NaN_3, 1 mM protease inhibitors, 10 % D_2O). Increasing aliquots of the Tafamidis (solubilized in DMSO-d6) were added until the titration was completed. Same approach was used for the sample complex of TTR with Taf-Ptx, on a sample of [U-^{13}C-^{15}N] TTR at a concentration of 400 µM of the monomer.

After the titrations, the excess of the ligands was removed using PD10 columns and the buffer exchanged to 10 mM MES pH=6.5 and 20 mM NaCl. PEG 1000 (on a 1:10 weight ratio) was added to protect the protein during the lyophilization process. And the samples were freeze-dried.

The solid state NMR spectra of free TTR and TTR in the presence of Tafamidis or Taf-Ptx were collected on a Bruker Avance III spectrometer operating at 800 MHz (18.8 T, 201.2 MHz ^{13}C Larmour frequency) equipped with a Bruker 3.2 mm Efree NCH probe-head and on a Bruker Avance III wide-bore spectrometer operating at 850 MHz (20T, 213.6 MHz ^{13}C Larmour frequency) equipped with a 3.2 mm DVT MAS probe head in triple-resonance mode. The spectra were recorded at 14 KHz MAS frequency and the sample temperature was kept at ~ 290 K. Standard ^{13}C-detected solid-state NMR spectra [2D $^{15}N^{13}C$ NCA, $^{15}N^{13}C$ NCO and $^{13}C^{13}C$ DARR; 3D NCACX and NCOCX] were acquired. 3D CANCO was also acquired on the sample of free-TTR. All the spectra were processed with the Bruker TopSpin 3.2 software and analyzed with the CARA software.

Solid-state NMR sample preparation of the TTR in complexes

The material obtained from lyophilization was used to pack 3.2 mm zirconia thick wall rotors. The materials were then rehydrated by multiple addition of MilliQ H_2O until the resolution of the 1D [^{1}H]^{13}C CP solid-state NMR spectra stopped improving. Silicon plugs placed below the turbine cap were used to close the rotor and preserve hydration. A sample

of 25 mg of [U-^{13}C-^{15}N] free TTR lyophilized in the presence of PEG1000 (1:10 weight ratio) was also analyzed in 3.2 mm zirconia rotor after rehydration, as reference for the solid-state NMR experiments.

5. Results and Discussion

5.1 Biophysical characterization of carbohydrate binding proteins

Siglec-7 NMR assignment and binding specificities to GD3 and its derivatives for new immunomodulatory strategies against cancer

*Article in preparation

Siglec-7 NMR assignment and binding specificities to GD3 and its derivatives for new immunomodulatory strategies against cancer

Cristina Di Carluccio[a,b], Marta Tiemblo-Martìn[a], Alessandro Antonio Masi[a], Rosario Oliva[a], Antonio Molinaro[a,b], Pompea Del Vecchio[a], Roberta Marchetti[a], Alba Silipo[a,b]

Luis Padilla-Cortés[c], Giulia Roxana Gheorghita[c,d], Linda Cerofolini[c], Marco Fragai[c]

Celeste Abreu[e], Ondřej Vaněk[e]

Chun-Cheng Lin[f]

[a] Department of Chemical Sciences and Task Force for Microbiome Studies, University of Naples Federico II, Via Cinthia 4, 80126, Naples, Italy

[b] CEINGE-Biotecnologie Avanzate Franco Salvatore, Via Gaetano Salvatore 486, 80145, Napoli, Italy

[c] Magnetic Resonance Centre (CERM), CIRMMP and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

[d] Giotto Biotech s.r.l., Sesto Fiorentino, Italy

[e] Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, 12800, Prague, Czech Republic

[f] Department of Medicinal and Applied Chemistry, Kaohsiung Medical University, Kaohsiung, 80708, Taiwan.

Abstract

Gangliosides are sialylated glycosphingolipids widely distributed in human cells and tissues and play crucial roles in cellular processes, such as neurotransmission, interaction with regulatory proteins of the nervous system, cell-cell recognition and modulation of signal transduction pathways. However, altered expression of gangliosides has been correlated with pathological processes, including cancer, inflammatory disorders and autoimmune diseases. Notably, gangliosides are ligands for Siglecs, I-type lectins found on most white blood cells of the immune system that specifically recognize sialic acid–containing glycans. Indeed, certain Siglecs are engaged by endogenous gangliosides to trigger important physiological and pathophysiological signaling events. For example, sialoglycans expressed on cancer cells surface can engage Siglec-7 on natural killer (NK) cells, leading to the inhibition of immune responses. Moreover, the presence of ligands for Siglec-7 and Siglec-9 has been found to be high in various cancer types, such as pancreatic cancer and melanoma. We here present a comprehensive analysis of the structure, conformation, and interactions of Siglec-7 with the sugar moieties of α2,8-linked gangliosides, including GD3 and Gb3 (globotriaosylceramide) derivatives, named DSGb3α3 and DSGb3α6. Understanding these interactions could provide valuable insights into disease mechanisms and facilitate the use of gangliosides as potential diagnostic and therapeutic strategies for cancer treatment. To this aim we combined structural biology approaches, NMR techniques, physico-chemical studies, and *in silico* methods to provide information on the binding affinities and 3D models of the complexes.

Introduction

Sialic acids are a family of negatively charged nonulosonic acids[1] attached to the terminal portion of N-glycans, O-glycans and glycosphingolipids via the enzymatic activity of membrane-bound sialyltransferases (STs). Depending on the enzyme, these STs can add one or more sialic acids (Sias) to a sugar residue of the glycan through α2-3, α2-6 and/or α2-8 glycosidic linkages, resulting in an immense diversity among the sialoglycans. Moreover, the sialoglycan recognition can be translated into a specific biological response when they are recognized by receptors like lectins[2]. Among these, Siglecs, sialic acid-binding immunoglobulin-like receptors, are I-type lectins that share an N-terminal Ig domain that recognizes sialic acid–containing glycans[3]. They are mostly expressed by immune cells, including B-cells, NK cells, eosinophils, neutrophils, dendritic cells and play a crucial role in distinguishing between self and non-self via glycan recognition[4]. Gangliosides, glycosphingolipids containing sialic acid (Scheme 1), are endogenous Siglec ligands[5], structurally composed of an extra-cellular carbohydrate moiety which is linked to a ceramide, a hydrophobic lipid portion embedded in the membrane. Their biosynthetic pathway starts in the endoplasmatic reticulum where glycosphingolipids are formed by specific glycosyltransferases which add sugars to the ceramide; the elongation of glycan structure, for example with the addition of one or more sialic acids to lactosylceramide (LacCer), the precursor of most gangliosides, ends in the Golgi. When complete, gangliosides are transferred to the extracellular environment by vesicular delivery. Based on the number of sialic acids on their glycan structure, gangliosides are classified as a-, b- and c-series corresponding to one, two or three Sias, respectively[6,7].

Recently, the study of gangliosides has made significant strides, particularly with respect to their role in pathological events. Even though they are broadly distributed, gangliosides predominate in the brain[8], and in particular, disialyl gangliosides play crucial roles in cellular processes such as neurotransmission, interaction with regulatory proteins of the nervous system, cell-cell recognition and modulation of signal transduction pathways[9-12]. In the nervous system, GD1a and GT1b, containing α2-3-linked sialic acids, are recognized by Siglec-4 (or myelin-associated glycoprotein MAG) to support axon-myelin interactions essential for long term axonal survival. Changes in the glycan structures of gangliosides are also associated to malignancies and cancer, such as the

increase of GD3 or GM2 in melanoma, of GD2 in neuroblastoma; certain cancers also produce and shed gangliosides that have immunosuppressive effects. Disialylated gangliosides in cancers affect cell behavior such as proliferation, migration, invasion, adhesion, and angiogenesis, as well as tumour immunosuppression[13-15]. The abnormal increase of gangliosides in serum has also been closely linked to cancer, presenting the possibility of predicting prognosis and recurrence in cancer patients[16,17].

Ligands for Siglec-7 and Siglec-9 have been found to be high in various cancer types, such as pancreatic cancer and melanoma[18]. Siglec-7 is an inhibitory immune receptor expressed on human natural killer (NK) cells and subsets of myeloid and dendritic cells[19,20]. Sialoglycans expressed on cancer cells surface can engage Siglec-7 and therefore inhibit NK cell-mediated killing, this making Siglec-7 and cognate ligands novel glyco-immune checkpoints for cancer immunotherapy[21-23]. Siglec-7 preferentially binds internally branched α2,6-linked disialylated gangliosides, such as DSGb5, disialosyl Lc4 (DSLc4), and α2,8-linked gangliosides such as GD2, GD3, and GT1b[24,25]. In particular, GD3 (NeuAcα2-8NeuAcα2-3Galβ1-4Glcβ1-1′Cer) plays essential roles in brain development but its levels are decreased in adults[26]. Conversely, GD3 is upregulated in pathological conditions, such as cancers and neurodegenerative disorders[27]. Once sialoglycans on cancer cells surface target Siglec-7 on NK cells, they evade immune detection and replicate [28,29]. Recent findings indicate that cell surface GD3 with regular ceramide supports Siglec-7 binding, whereas GD3 with ceramide alterations, including an extra hydroxyl group on sphingosine C4 (phytoceramide) or a 2-hydroxyl group on the fatty acid amide, does not effectively engage with Siglec-7. Furthermore, it was observed that dehydroxylated ceramide-containing GD3 reduced the sensitivity to NK cells, indicating that ceramide structures in glycosphingolipids not only influence Siglec-7 binding but also impact the biological effects triggered by their interactions[5,30]. This raises the intriguing speculation that glycan binding, specifically of the Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc of ganglioside GD3 enhances engagement with the hydrophobic ceramide. Understanding the dynamics of these interactions holds great promise for providing insights into disease mechanisms and potentially opening the door to the development of diagnostic and therapeutic strategies[31]. This can have significant implications in the field of immunotherapy, where targeting these interactions may lead to novel therapeutic strategies against cancer, which often exploits sialic acid-containing molecules in immune evasion mechanisms[32,33]. To deeply understand how Siglec-7 can bind closely-related

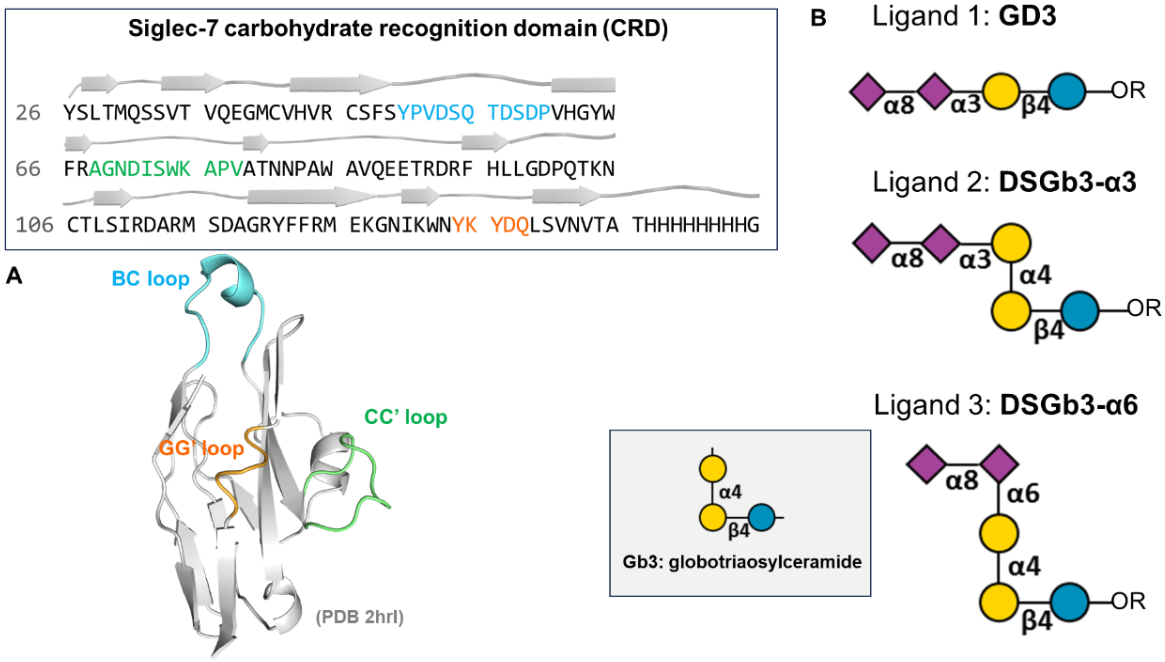
sialoglycan structures, and with the long-term goal of developing potential ligands within cancer immunotherapy, we undertook a comprehensive study of Siglec-7 recognition and binding to the sugar portions of GD3[34] and two Gb3 (globotriaosylceramide) derivatives[35], such as DSGb3α3 and DSGb3α6 (Scheme 1). For this purpose, we employed a combination of multidisciplinary and complementary methods, consisting of fluorescence, high-resolution ligand- and protein-based NMR experiments, isothermal titration calorimetry and computational approaches, including docking, molecular dynamic simulations and RedMat 3D structure evaluation program[36] to achieve information about binding preferences, affinities and 3D molecular features of such protein-ligand complexes. Comprehending these interactions is instrumental in elucidating their roles in immunology, disease pathology, and potential therapeutic applications.

Results

The molecular binding between the glycosylated Siglec-7 and gangliosides (ligands **1**, **2** and **3,** Figure S1 and Scheme 1) was investigated using a combination of biophysical, NMR and computational methods.

NMR assignment of the backbone ^{1}H and ^{15}N resonances of Siglec-7 CRD

The N-terminal Ig-like C-type carbohydrate recognition domain (CRD) of Siglec-7 (Scheme 1) has been expressed in E. coli in M9 culture medium containing ^{15}N NH_4Cl and ^{13}C-glucose. The protein was recovered from inclusion bodies and after a refolding protocol, highly pure Siglec-7 CRD was obtained isotopically enriched. ^{1}H, ^{15}N and ^{13}C backbone resonances were assigned through solution state NMR. 2D ^{1}H-^{15}N HSQC NMR experiments showed well-dispersed resonances, which were used as a confirmation of the folding state of the protein (Figure 1, Table S1). 3D triple-resonance experiments HNCA, HNCACB, HNCO and CBCAcoNH were acquired at 900 MHz, HNcaCO a spectrometer operating at 500 MHz. 93 % of the amino acid sequence from Y26 to T147 was assigned, excluding the 5 proline residues and the His-tag tail. The recombinant Siglec-7 CRD was then used in protein-based NMR experiments for binding studies with gangliosides **1-3** (see below).



Scheme 1. *A) Amino acids sequence of the two different forms of Siglec-7 studied in this work, the carbohydrate recognition domain (CRD) and the full extracellular domain (FED). The PDB structure of Siglec-7-CRD (2hrl) is also shown with BC, CC' and GG' loops highlighted. B) Ligands studied in this work: GD3 and Gb3 (globotriaosylceramide) derivatives DSGb3-α3 and DSGb3-α6.*

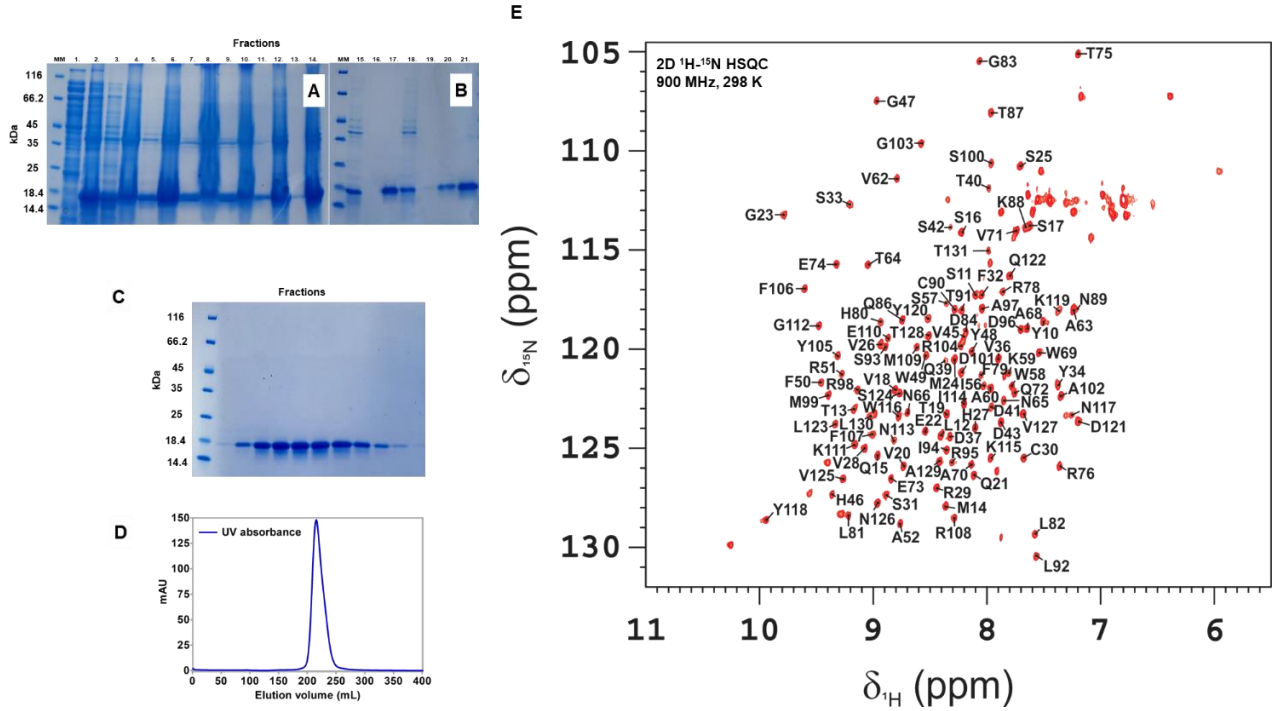


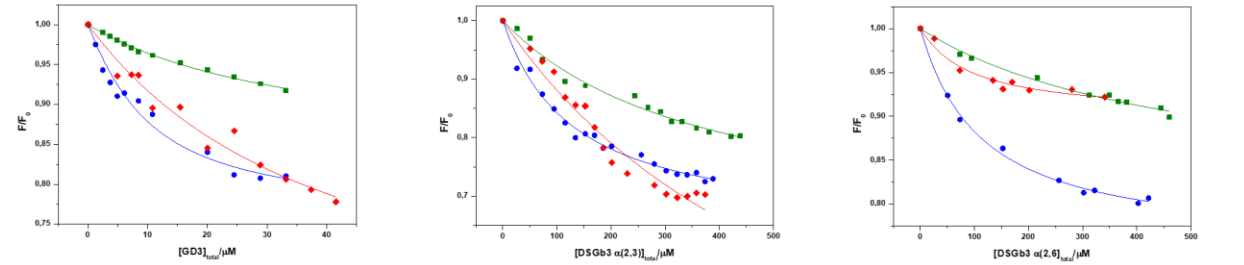
Figure 1. *Siglec-7 production and backbone assignment by NMR. A) SDS-PAGE representation: Lysis step - (1) supernatant, (2) cell debris; Solubilization of Inclusion bodies - (3) supernatant, (4) cell debris; Solubilization of Inclusion bodies - (5) supernatant, (6) cell debris; Solubilization of Inclusion bodies - (7) supernatant, (8) cell debris; Solubilization of Inclusion bodies - (9) supernatant, (10) cell debris; Solubilization of Inclusion bodies - (11) supernatant, (12) cell debris; Solubilization of Inclusion bodies - (13) supernatant, (14) cell debris. B) SDS-PAGE representation: HisTrap purification, 1-st cycle - Flow-through (15), Column wash (16), Elution (17); HisTrap purification, 2-nd cycle - Flow-through (18), Column wash (19), Elution (20), 1-st Elution and 2-nd Elution (21). C) SDS-PAGE representation: Size-exclusion chromatography - collected fraction of Siglec-7 CRD. D) Size-exclusion chromatogram representation of Siglec-7 CRD as single peak. E) 2D ^{1}H-^{15}N HSQC NMR spectrum of the apo Siglec-7 CRD in 20 mM KPi, 50 mM NaCl, pH 7.4 acquired on a spectrometer operating at 900 MHz at 298 K. NH amino acid resonances obtained by the protein assignment are reported in the spectrum.*

Binding specificities

Concentration dependent reduction in fluorescence intensity upon ganglioside binding was used to monitor the interaction and the binding affinities were determined by non-linear regression analysis[37]. In details, fluorescence titrations of increasing amounts of sialoglycans into a fixed concentration of the proteins showed that the tryptophan residues of Siglec-7 were quenched by ligand addition, therefore proving the complex formation[38]. These measurements allowed to calculate the binding affinities (K_b) for all systems. The results showed the ability of Siglec-7 to

recognize all gangliosides; while a stronger preference for ligand **1** was observed, the binding constants were all in the micromolar range (Figure 2). To understand if Siglec-7-ganglioside binding was enthalpically and/or entropically driven, fluorescence analyses of all systems were performed at different temperatures (10°C, 25°C and 35°C) and the corresponding association constants were derived accordingly. By assuming that ΔH° was temperature-independent, the van't Hoff equation was applied. In all cases, the enthalpy values were negative, meaning that the binding was exothermic and therefore, dominated by hydrogen bonds (Figure S2A). Slight variations among the systems were attributed to the entropy (ΔS), although it was negative in all cases. It is worth mentioning that the thermodynamic values (ΔH°, ΔS and ΔG) of the Siglec-7 - ligand **1** system were also calculated by isothermal titration calorimetry (ITC), which confirmed the results obtained by the application of the van't Hoff equation (details in Methods section, Figure S2B) [39]. Overall, we can affirm that Siglec-7 interactions with GD3 and DSGb3 containing α2,3 and α2,6 linkages (DSGb3α3 and DSGb3α6, respectively) have affinities in the low micromolar range (Figures 2 and S2).

Binding affinities were also studied using NMR lineshape analysis of 2D ^{1}H-^{15}N HSQC experiments. Solution NMR is a powerful tool for the study of biomolecular interactions, and changes on 2D NMR spectra of a free protein upon titration with a given ligand is a rich source of information capable of characterizing the thermodynamic features of an interaction. Dissociation constants for the interactions between Siglec-7 and ligands 1, 2 and 3, were estimated using TITAN (TITration ANalysis)[40], a software that uses quantum mechanical simulation and fitting of whole 2D spectra to specific binding models, depending on the system. The data was predicted and fitted for HSQC spectra and a 2-state binding model for all three interactions. Supplementary figures S, show the fitting obtained for representative residues involved in the interaction. Dissociation constants in the high micromolar range were obtained for all three interactions.



Ligand 1: GD3

Temperature	System	K_h / M^{-1}	K_D / µM
10	Sig7(CDR)/GD3	(1,94±0,7)·10^5	5,1546
25	Sig7(CDR)/GD3	(6,29±0,1)·10^4	15,898
35	Sig7(CDR)/GD3	(4,07±0,1)·10^4	24,5700

Ligand 2: DSGb3 α(2,3)

Temperature	System	K_h / M^{-1}	K_D / µM
10	Sig7(CDR)/DSGb3a-(2,3)	(8,07±0,1)·10^3	123,9157
25	Sig7(CDR)/DSGb3a-(2,3)	(2,66±0,2)·10^3	375,9398
35	Sig7(CDR)/DSGb3a-(2,3)	(1,40±0,2)·10^3	716,8459

Ligand 3: DSGb3 α(2,6)

Temperature	System	K_h / M^{-1}	K_D / µM
10	Sig7(CDR)/DSGb3a-(2,6)	(1,54±0,6)·10^4	64,95615
25	Sig7(CDR)/DSGb3a-(2,6)	(1,60±0,4)·10^3	626,9592
35	Sig7(CDR)/DSGb3a-(2,6)	(1,25±0,1)·10^3	801,9246

***Figure 2**. Binding isotherms measured by fluorescence spectroscopy following the formation of the complex between Siglec-7 and (A) GD3 (B) DSGb3α3 and (C) DSGb3α6 in phosphate buffer saline (PBS) buffer, pH 7.4, at temperatures of 10°C (blue circles), 25°C (green squares) and 35°C (red rhombuses). The solid lines represent the best fit of the experimental data according to a 1:1 binding model.*

Molecular binding between Siglec-7 and ligand 1

The interaction between Siglec-7 and GD3 (ligand **1**) was used to assess the functionality of the recombinant protein, to describe the binding mode and to dissect the 3D complex.

The molecular details of the interaction between Siglec-7 and its natural ganglioside GD3 were unveiled by both ligand- and protein-based NMR spectroscopy, as well as computational approaches[41] (MM and MD simulations) (Figure 3). Saturation Transfer Difference (STD) NMR experiments were performed on the mixture to map the recognized epitope of GD3, therefore, to determine the protons closest to the receptor surface (Figure 3A). The different profiles between the reference (in black, Figure 3A) and the STD NMR (in red, Figure 3A) spectra were a clear indication of selective binding between the protein and ligand **1**. Furthermore, the STD spectrum did not contain signals from Gal (**D**) and Glc (**B**) residues; the main STD signals belonged to the terminal sialic acid (**N**), indicating its primary involvement in the interaction with Siglec-7, while residue **K** contributed mainly with H7, H4 and its acetyl group, and at lower extent with H5-H6. The highest STD NMR response was attributed to the acetyl group of the terminal sialic acid (AcN) and was set to 100%; the %STD effects of the magnetization transferred from Siglec-7 to the other protons of GD3 were derived accordingly.

The interaction of Siglec-7 with GD3 was further investigated by protein-based NMR experiments (Figure 3B). Aliquots of GD3 were sequentially added to [U-^{15}N] labelled Siglec-7 CRD and ^{1}H-^{15}N-HSQC NMR spectra were acquired. As expected, during the NMR titration, some protein

signals experienced aa decrease in signal intensity and/or chemical shift perturbations (CSP) (Figure 3B and 4A-B), These changes were particularly noticeable in signals located near Arg124, the canonical residue in the Siglec-7 binding site . In particular, the signals affected by the largest CSP were assigned to Lys127, Lys131 and Trp132, followed by Glu126, Gly128, Ile130, as well as Tyr134 and Lys135 of GG' loop, supporting the anticipated interaction involving Arg124. Other CSP effects were observed for Tyr26, Ser27, Arg111, and for His62 of BC loop and Ile72, Ser73 of CC' loop. On the other hand, the signals affected by the largest decreases in intensity were assigned to Asn133, that even disappeared with the addition of the ligand, and Lys127 and Lys131, the same residues that also experienced CSP located close to Arg124. To a lesser extent, other intensity decreases were also observed for Thr56, Gly63, Phe95, Met125.

The findings from both the NMR titration and STD experiments were key in constructing the binding site (Figure 4A and 4B). The α2-8 linkage connecting the two sialic acids of GD3 shows high levels of conformational flexibility due to the five torsion angles by which it is characterized: F (C1-C2-O-C8′), Y (C2-O-C8′-C7′), ω_9 (O9′-C9′-C8′-O), ω_8 (O8′-C8′-C7′-O7′), and ω_7 (O7′-C7′-C6′-O6′). Once the conformational behavior of GD3 in the free state, previously evaluated[42], was here further studied via MD simulations, (Figure S3A), the ligand in its representative pose was then modeled into Siglec-7. Depending on the φ (C1-C2-O-C'3) torsion angle around the Neu5Ac-α-(2,3)-Gal glycosidic linkage, ligand **1** could populate two different conformations, namely *–g* and *t*, corresponding to φ values of -60° and 180°, respectively. In the bound state, a preference for φ = -60° was observed from MD simulations (Figure S3B), also supported by tr-NOESY experiments (Figure S3C). Overall, the spectral shifts observed in Siglec-7 upon the addition of ligand **1** were in agreement with a protein-ligand binding affinity at a low micromolar level. The combination of ligand- and protein-based NMR experiments allowed the modelling of **1** into the binding site. The complex was subjected to MD simulation (Figure S3B) and validity of the 3D models were evaluated and validated with RedMat (Figure 4E), a program that calculates the predicted STD-NMR intensities from MD trajectories and compares them to the measured STD effects. Notably, the calculated STDs, matched well with those experimentally observed, therefore confirming reliability of the MD simulation on Siglec-7 and GD3 (ligand **1**) and thus of the proposed 3D model in Figure 4. As shown in Figures 4C and 5, the carboxylate group of the terminal sialic acid established the key salt bridge with the Arg124. Also, a significant H-bond was detected between the H5N of terminal sialic acid **N** and Lys131, one of the amino acid most

perturbed by both CSP and intensity decrease (Figures 3B and 4A-B). At lesser extent, H-bonds between O8 of glycerol chain of **N** with Asn133 were observed, the most perturbed amino acid in signal intensity decrease. Regarding the internal Sia residue **K**, an H-bond between O4 of sialic acid and Trp74 was recurrent along the simulation. The analysis also revealed the presence of hydrophobic amino acids in the binding pocket involving the aromatic components of Trp132 and Trp74, this latter situated within the CC' loop of the protein, in proximity to the residues **N** and **K,** respectively (Figure 4C). In particular, in our 3D model, Trp132 established a π-π interaction with Tyr26, both amino acids affected by CSP variations during the protein NMR titration (Figures 3B and 4B).

Overall, the 3D model of Siglec-7 and GD3 (ligand **1**) here proposed matched NMR binding studies, indicating the terminal sialic acid of the ganglioside as the main residue involved in the interaction with Siglec-7 and the internal sialic acid partially contributing to the recognition, while the other residues were solvent exposed.

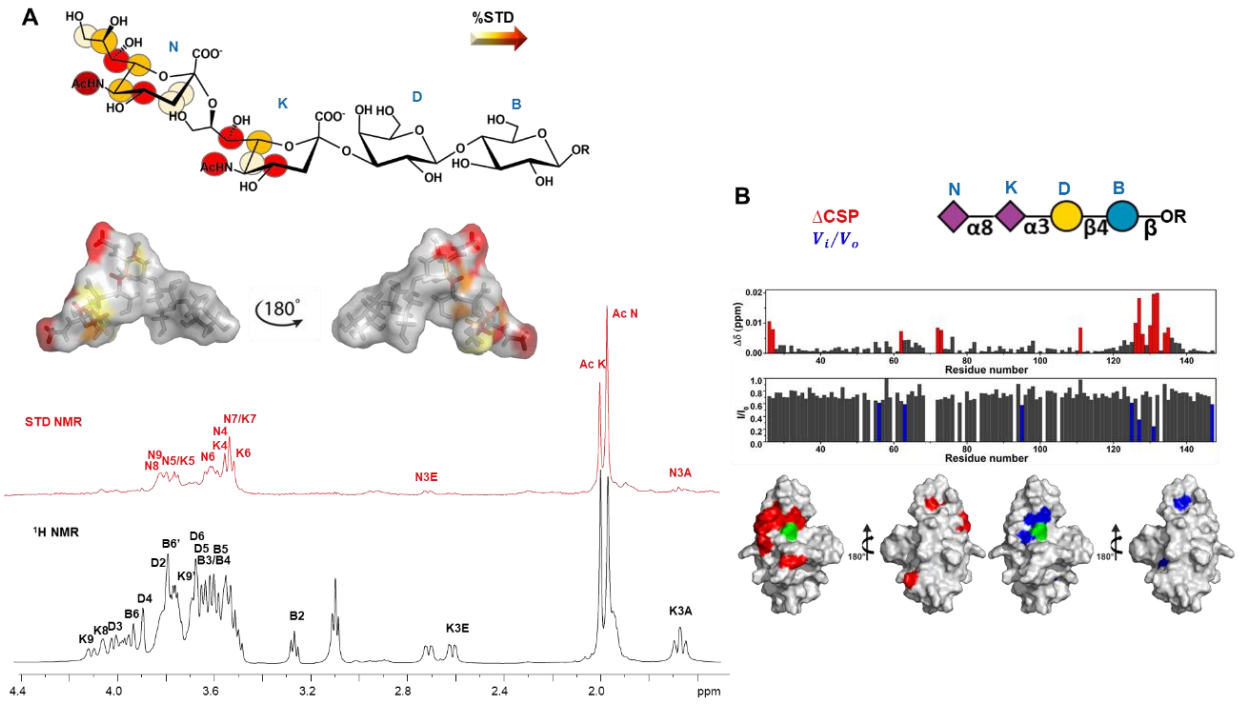


***Figure 3**. Ligand- and Protein-based NMR analysis of ligand **1** in interaction with Siglec-7. A) Epitope mapping of ligand **1** highlights the protons most involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMR spectrum (red) and I0 is the peak intensity of the off-resonance spectrum (black). Despite some isochronous signals between N and K, the isolated signals from K as K8 as well as the diastereotopic K9 and K3 (axial and equatorial) protons were not recognized by the protein. However, an STD enhancement referring to the acetyl group (AcK) was observed, meaning that a partial contribution of the internal sialic acid (K) occurred. Bioactive conformation of ligand*

__1__, the surface was colored according to the STD effects. B) Plots representing chemical shift perturbation (CPS) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensity in blue and CSP in red in the presence of 200 μM of ligand __1__. Arg124 is colored green.

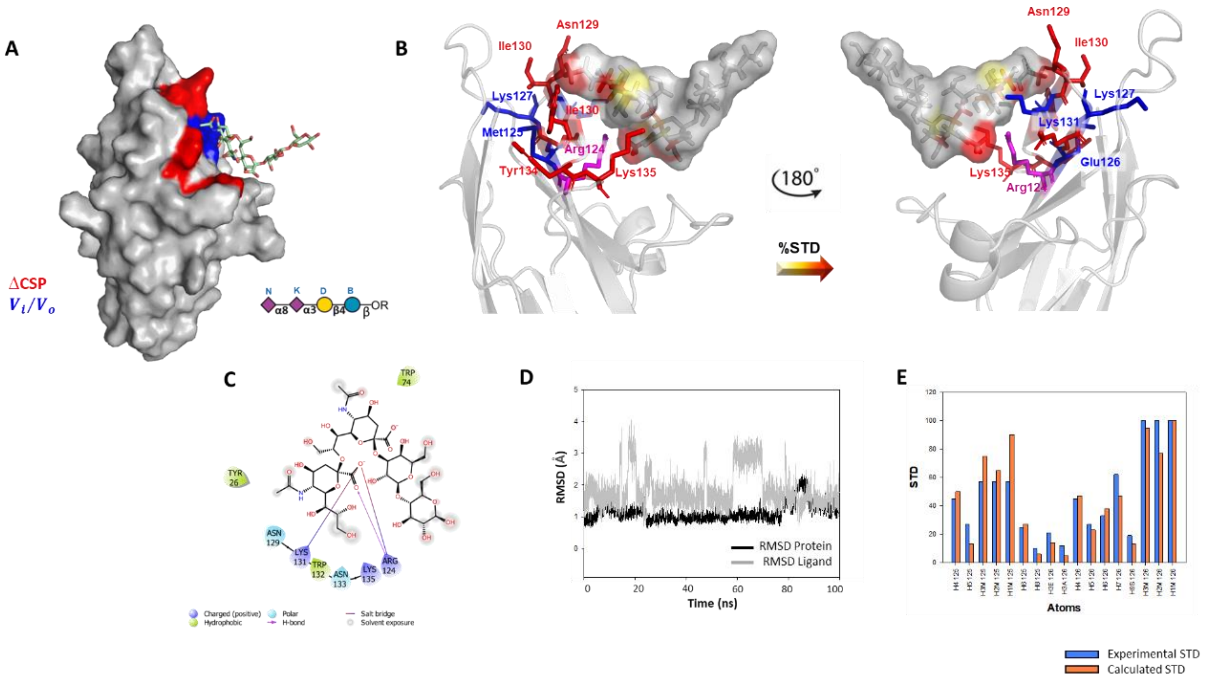


***Figure 4**. 3D model of Siglec-7CRD -ligand __1__ complex. A) 3D complex with the protein surface colored according to the chemical shift perturbation (in red) and intensity decreased (in blue) detected by protein-based NMR titration. B) 3D views of the Siglec-7−__1__ complex: the amino acids involve in the interactions as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code. C) 2D plot of the interactions establishing at the protein-ligand interface. D) RMSD along MD simulation. E) Comparison between experimental (blue) and calculated %STD for protons of ligand __1__ by RedMat analysis. A NOE R-factor of 0.2 was calculated.*

Molecular binding between Siglec-7 and ligand 2

The molecular recognition of DSGb3α3 (ligand **2**) by Siglec-7 was analogously investigated by NMR and computational studies. STD NMR experiments indicated that Siglec-7 mainly accommodates sialic acid residues, primarily the terminal Sia (**N**) (Figure 6A). An inspection of the STD spectra highlighted the above, for instance K3 signals were based in the STD spectrum, the acetyl group and N4 STD signals were higher for Sia **N** then for the internal Sia **K**. Furthermore, other sugar residues (galactoses, **D** and **C**, and glucose, **B**) were mostly excluded from recognition by Siglec-7 (Figure 6A).

Siglec-7 titration with ligand **2** was then performed in order to have a deeper insight into the 3D complex; both CSP and decreases in signal intensity were observed and allowed to map the region of the protein involved in the interaction with DSGb3α3, especially surrounding the key Arg124 (Figures 6B and 7). In this region, the most significative CSPs were attributed to Gly128, Ile130, Lys131, Trp132 and Tyr134, while Lys131, Trp132, Asn133 and all the amino acids from Lys135 to Gln138 of the GG' loop experienced a decrease in signal intensity. The residues of GG' loop (Tyr134 by CSP and Lys135, Tyr136, Asp137 and Glu138 by intensity decreases) were indeed particularly influenced by the presence of ligand **2**; we also observed a CSP of Ala76 of the BC loop and decreases in signal intensity of Tyr50 and Val52 belonging to the BC loop. Besides Lys131, Trp132 that were affected by both kinds of variations, we also found Tyr26, Ser27 and His62, close in space to the binding site (Figure 7).

The bioactive conformation adopted by ligand **2** upon binding to Siglec-7 was then explored by tr-NOESY experiments (Figure S6) coupled to MD simulation. In the free state, ligand **2** could populate *–g* and *t* conformations according to the φ (C1-C2-O-C'3) torsion angle around the Neu5Ac-α-(2,3)-Gal glycosidic linkage (-60° and 180° respectively, Figure S4). Therefore, an MD simulation in the bound state was first carried out with ligand **2** in the *t* conformation (Figure S5A), maintained for approximately 20 ns after which the φ torsion angle turned to -60° (change from *t* to *-g*) and maintained until the end of the MD simulation. In addition, a second MD simulation in the bound state was undertaken considering ligand **2** in *-g* conformation as starting pose (Figure S5B): for the entire simulation time, the φ torsion angle populated the value at -60°. Therefore, in both MD simulations of the bound states, the conformation of the φ torsion angle around Neu5Ac-α-(2,3)-Gal of ligand **2** converged to -60º, suggesting the preference for the -g conformation. Analogously, a detailed analysis of key tr-NOE contacts indicated a propensity of **2** for adopting the -g conformation upon binding to Siglec-7. This was supported by a decrease in NOE intensity corresponding to $K3_{ax}$-D3, and the appearance of a weak NOE corresponding to K9-D3, denoting a preference for φ at -60°. Regarding the torsion angles around Neu5Ac-α-(2,8)-Neu5Ac linkage, φ (C1-C2-O8-C8) can populate different conformational states between -60° and -180° in the free state (Figure S4), while φ at-60° was favored in the bound state (Figure S5B). In addition, tr-NOE contacts between the terminal sialic acid (**N**) and the glucose residue (**B**) further suggested a bent conformation for **2** upon binding with Siglec-7, well matching with a φ torsion angle around Neu5Ac-α-(2,8)-Neu5Ac at -60° (Figure 5-6). Notably these torsions remained stable along the

MD simulation (Figure S5). Moreover, the MD simulation trajectories were analyzed through RedMat algorithm: the frame more in accordance with the experimental data was extrapolated by the MD simulation and a 3D model of Siglec-7 with DSGb3α3 was proposed (Figures 7 and 8). The results obtained (Figure 8), correlating with the previous analysis of ganglioside GD3 (ligand **1**), illustrate the key role of the carboxylate group of the terminal sialic acid **N** in establishing a salt bridge with Arg124. Additionally, hydrogen bonds were identified again between the H5N of the terminal sialic acid **N** and Lys131, affected by both CSP and decrease of signal intensity, as well the previous interactions with Asn133, which showed a strong reduction in signal intensity during the protein NMR titration (Figure 6B). Within the internal Sia residue **K,** this residue was facing toward the aromaticTrp74 belonging to the CC' loop by which a hydrogen bond was also established (Figures 7C and 8). Other aromatic amino acids located in the binding site included Tyr26 and Trp132, which were affected by both CSP and intensity decreases and established π-π interactions with each other.

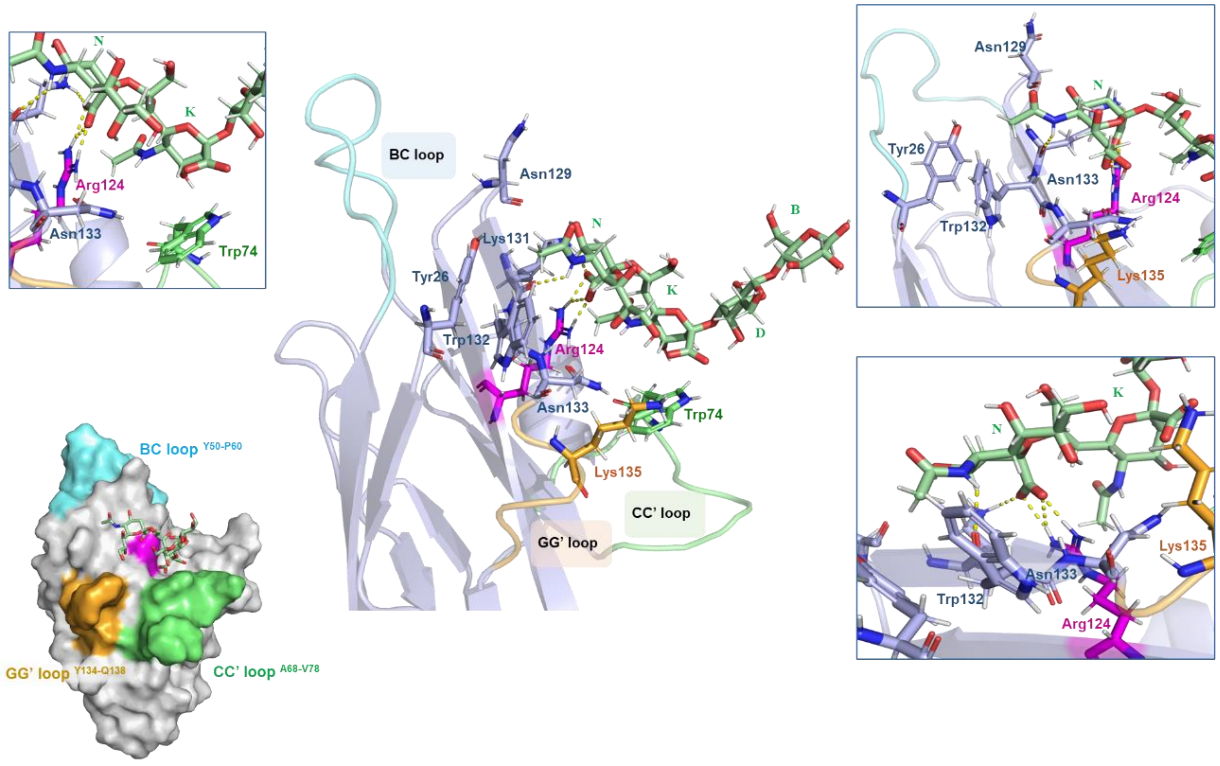


Figure 5. *Representation of the Siglec-7−**1** complex with the BC, CC' and GG' loops highlighted in cyan, green and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend).*

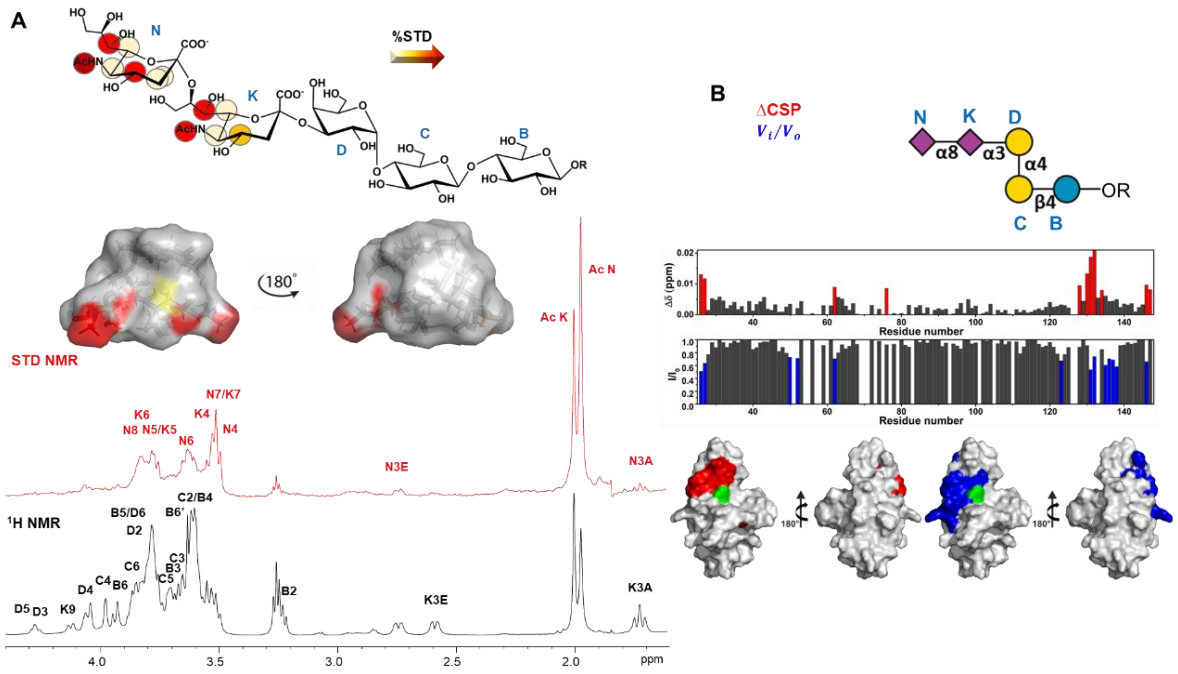


Figure 6. *Ligand- and Protein-based NMR analysis of ligand **2** in interaction with Siglec-7. A) Epitope mapping of ligand **2** highlights the protons more involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMR spectrum (red) and I0 is the peak intensity of the off-resonance spectrum (black). Bioactive conformation of ligand **2**, the surface was colored according to the STD effects. B) Plots representing chemical shift perturbation (CPS) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensity in blue and CSP in red in the presence of 200 μM of ligand **2**. Arg124 is colored green.*

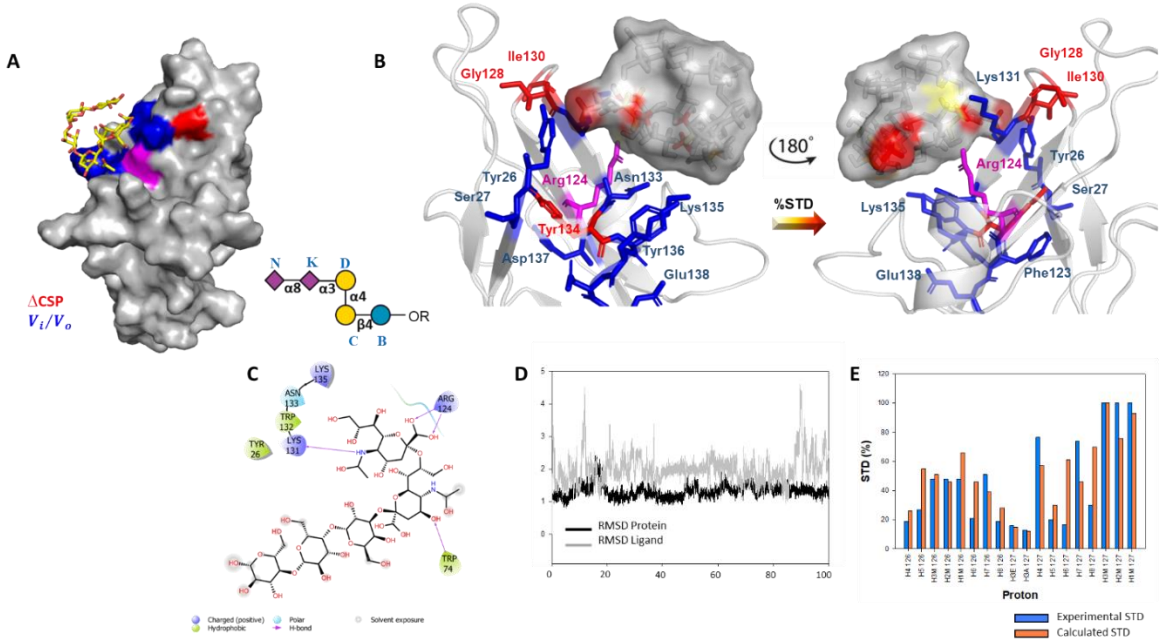


***Figure** 7. 3D model of Siglec-7CRD -ligand **2** complex. A) 3D complex with the protein surface colored according to the chemical shift perturbation (in red) and intensity decreased (in blue) detected by protein-based NMR titration. B) 3D views of the Siglec-7−**2** complex: the amino acids involve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

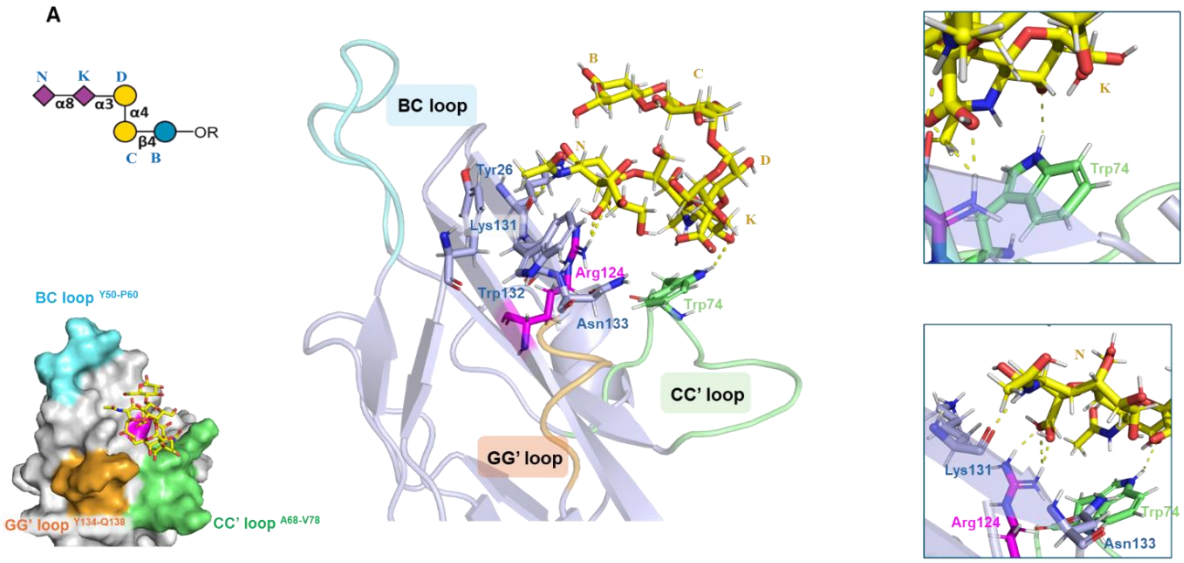


Figure 8. *Representation of the Siglec-7−**2** complex with the BC, CC' and GG' loops highlighted in cyan, green and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend).*

Molecular binding between Siglec-7 and ligand 3

NMR and computational studies have been also performed to study the molecular recognition of ligand **3** by Siglec-7. Similarly to ligand **2**, STD NMR experiments of ligand **3** showed that the terminal Sia residue (**N**) was crucial in the interaction (Figure 9A). Interestingly, as for the diasterotopic methylene signals of Sia, a low STD enhancement was only noticed for N3E, while the other protons of Sia **N** contributed from N4 to N9, including the acetyl group that gave the 100% of STD effect. Regarding Sia **K**, the main STD responses were attributed to K4 and K7 (both partially overlapped to N7), K6 and the acetyl group, while no effects were shown for K8 and K9, resonating in a more isolated region of the spectrum. Similarly to ligands **1** and **2**, the other residues were excluded from interaction with Siglec-7, as indicated by the absence of corresponding STD NMR signals.

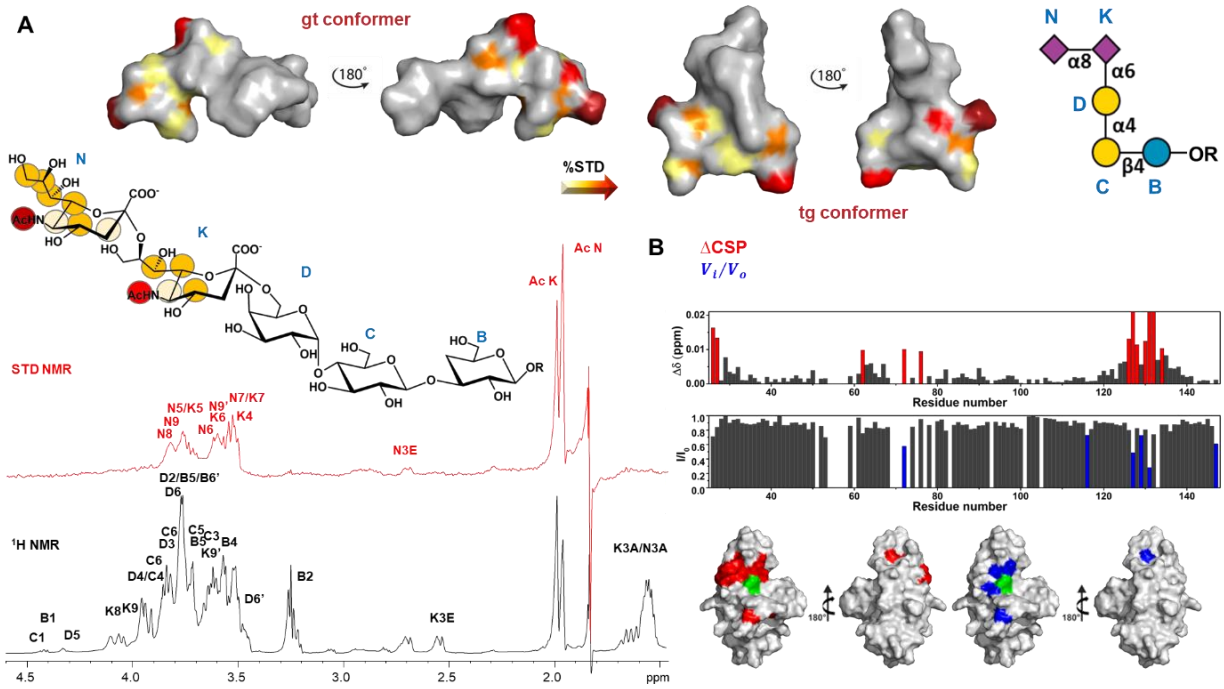


Figure 9*. Ligand- and Protein-based NMR analysis of ligand **3** in interaction with Siglec-7. A) Epitope mapping of ligand **3** highlights the protons more involved in interaction with Siglec-7. %STD values are obtained by the ratio (I0 − Isat)/I0, where (I0 − Isat) is the signal intensity in the STD-NMR spectrum (red) and I0 is the peak intensity of the off-resonance spectrum (black). Bioactive conformation of ligand **3**, the surface was colored according to the STD effects. B) Plots representing chemical shift*

*perturbation (CPS) in red and decreases in signal intensity in blue. Surface representation of a model of Siglec-7 with the residues experiencing the largest decrease in intensity in blue and CSP in red in the presence of 200 μM of ligand **3**. Arg124 is colored green.*

The molecular binding between Siglec-7 and ligand **3** was also investigated by protein-based NMR binding experiments (Figure 9B). Both CSP and decreases in signal intensity were observed during the titrations, which allowed the definition of the binding site of Siglec-7 (Figures 9B and 10). Therefore, residues from Glu126 to Tyr134, in the vicinity of Arg124, exhibited consistent perturbations; in particular, Lys127, Lys131 and Trp132 experienced the largest CSP, followed by Glu126, Gly128, Ile130, and Tyr134; the remaining amino acids were instead affected by intensity decrease (Lys127, Asn129, Lys131 and Asn133). Notably, Lys127 and Lys131 were affected by both CSP and intensity decrease. Moreover, the amino acids in proximity to this latter region included Tyr26, Ser27, which gave only CSP effects. On the other hand, lower perturbations were evidenced for residues His62 of the BC loop and Ala76 of the CC loop influenced by CSP and for Ile72 of the CC loop affected by both CSP and intensity decrease (Figures 9B and 10).

TrNOESY experiments and computational studies allowed to depict the 3D model. Thus, Neu5Acα2-6-Gal linkage was affected not only by φ (C1-C2-OC6′) and ψ (C2-O-C6′-C5′), but by an additional ω torsion angle around the C5′-C6′ (O-C6′-C5′-O5′). Previous computational studies in the free state (Figure S7A) showed the ψ angle mainly acquired values around 180°[43,44], also maintained in the bound state (Figure S7B). As for the φ torsion angle, the absence of NOE contacts between the diastereotopic protons at position 3 of sialic acid (K3A and K3E) and protons at position 6 of galactose (D residue) could be an indication of a preference for the -g conformation (Figure S8A). MD simulations (Figure S7) confirmed the strong preference for φ = -60° (-g conformer). Additionally, the ω torsion angle could populate three different values in the free state, 60°/180°/-60°, corresponding to gt/tg/gg conformations, respectively (see Newman projection in Figure S8B)[45,46]. Overall, both MD simulations (Figure S7A-B) and tr-NOESY NMR experiments (Figure S8A) revealed no conformational differences between free and bound state, and that both tg and gt conformations were likely populated (Figure 10 and S9). Indeed, ω at -60° was less energetically favored and, the different multiplicity of the diastereotopic protons at position 6 of the galactose residue D observed on the $^{1}H^{13}C$-HSQC spectrum confirmed the exclusion of the gg conformation (Figure S8B). As for the other two torsions, no significant NOE contacts were found for ligand **3** to discriminate in the bound state between the two conformers

(Figure S7B). The coexistence of both gt and tg conformers was further supported by protein-based NMR experiments (Figures 11 and 12), which showed that amino acids located in the CC' loop of the protein, such as Ile72 and Trp74, were affected in both cases. This also occurred with amino acids nearby or part of the GG' loop such as Lys131, Trp132, Asn133, Tyr134 or Lys135. The interactions of the key Arg124 with the terminal sialic acid **K** of the ligand occurred in the gt and tg conformations, hydrogen bonds could also be identified between Lys131 and the terminal sialic acid, in particular the NH of the acetyl group. Particularly, in the tg conformation the H-bonds between the terminal sialic acid (**N**) O8 and (at lesser extent) O9 of glycerol chain and Asn133 were also observed. Within the internal Sia **K** residue, a hydrogen bond with Trp74 was observed in both conformers (Figures 11 and 12).

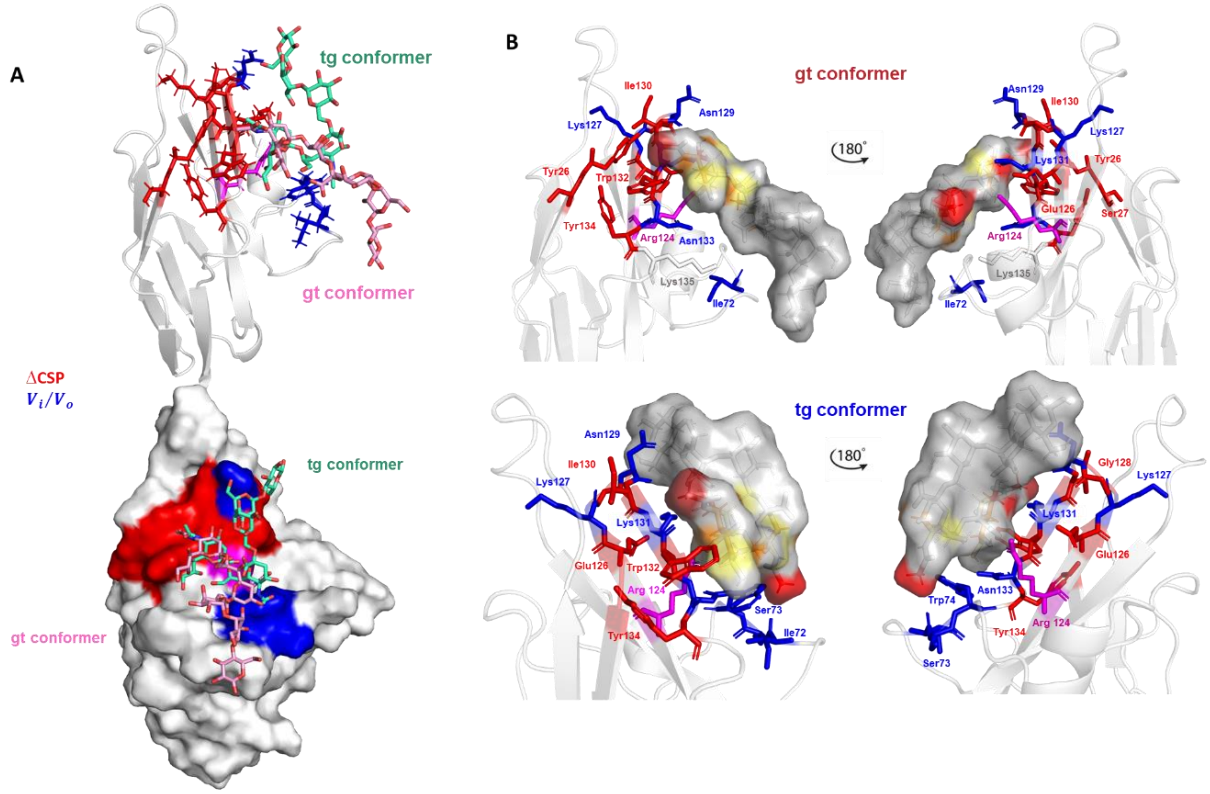


***Figure 10**. A) Superimpositon of 3D models of Siglec-7CRD bound to ligand **3** in tg and gt. B) 3D views of the Siglec-7−**3** complex: the amino acids involve in the interactions as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code.*

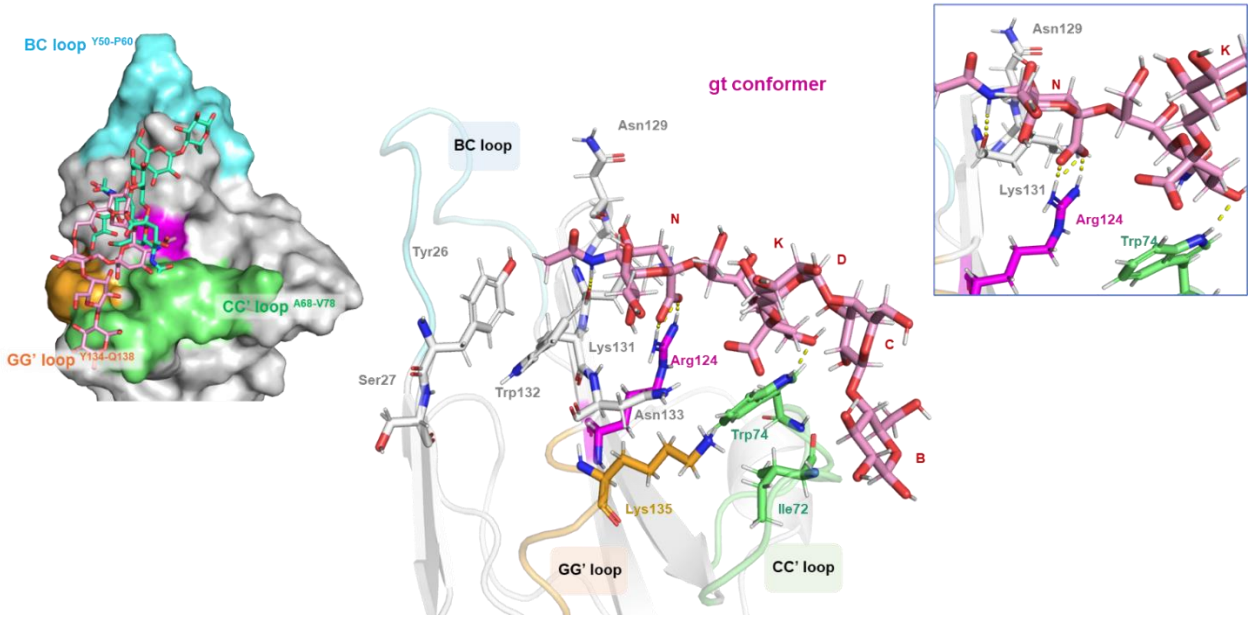


Figure 11. *Representation of the Siglec-7−**3** complex in the gt conformation with the BC, CC' and GG' loops highlighted in cyan, green and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend).*

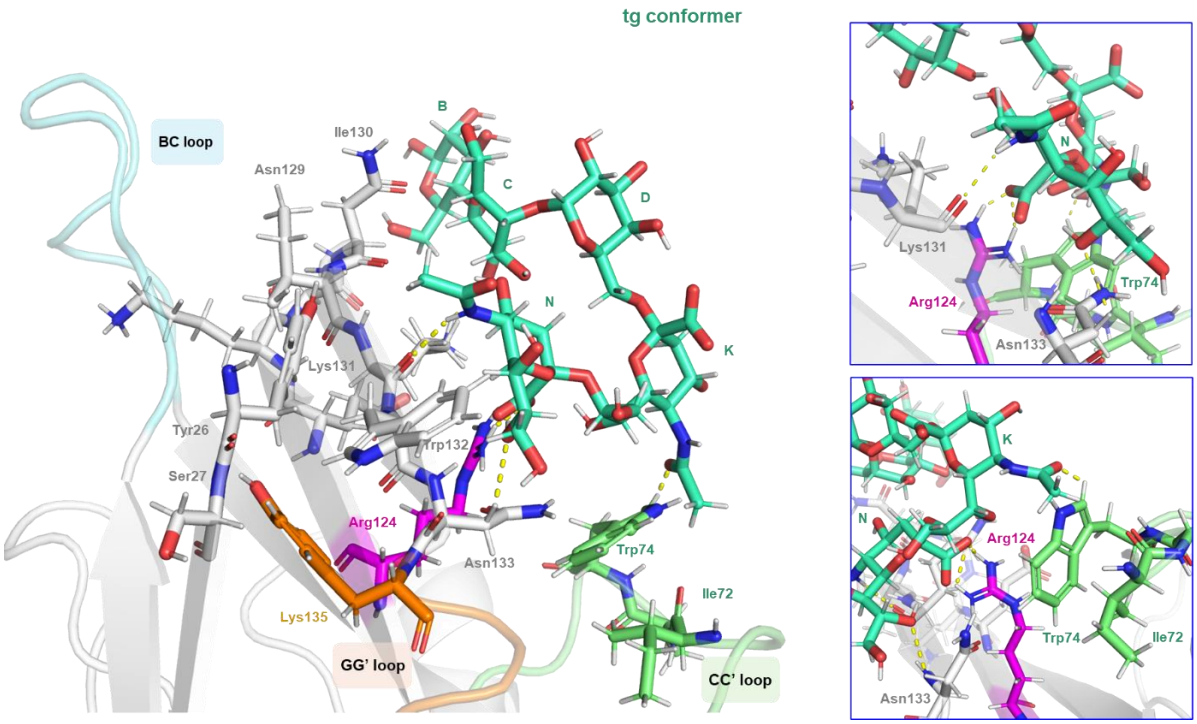


Figure 12. *Representation of the Siglec-7−**3** complex in the tg conformation with the BC, CC' and GG' loops highlighted in cyan, green and orange, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops legend).*

Discussion

Gangliosides are sialylated glycosphingolipids ubiquitous in all human cells, but most abundant in the brain and nervous system. Based on their location in the plasma membrane, these sialylated glycosphingolipids play several biological roles under healthy conditions, such as cell-cell and cell-matrix interactions, modulation of signal transduction, coordination of signal events with other cells.

However, ganglioside composition could be altered during malignant transformation and high levels of ganglioside expression contribute to pathological conditions, such as GD3 and GD2 in melanoma and neuroblastoma respectively, leading to tumor initiation and progression. Moreover, the presence of gangliosides was detected in the blood circulation of certain cancer patients, making them potential biomarker for tumor diagnosis and/or prognosis[47] or targets for antibody therapies[48,49]. Several *in vitro* studies have shown that most gangliosides containing one or two sialic acids have suppressive effects against the activation of immune cells, such as B cells, NK cells, dendritic cells, monocytes, promoting tumor progression, while structures with more than two sialic acids stimulate T cell activity and pro-inflammatory cytokines. Thus, certain gangliosides, for example GD2, GD1b and GT1b, play dual role in immunity, depending on their composition, sialylation pattern and specific tumor type [50],[i] but the mechanism is still unclear. GM3, GM2 and GD3, present in the nervous system, could inhibit NK cell activity [51] [ii] and one possible mechanism by which gangliosides can regulate NK cell cytotoxicity is through their recognition by Siglecs (sialic acid–binding immunoglobulin-like lectins). Indeed, GD3 strongly engages Siglec-7 on NK cells, leading to immunosuppressive response *in vitro* and *in vivo*[52].

Siglec-7 is a glycan-binding immune receptor that is emerging as a significant target of interest for cancer immunotherapy. Siglec-7 ligands offer valuable insights into the motifs and key elements essential for engaging Siglec-7 and evading immune recognition. This interaction establishes Siglec-7 and its corresponding ligands as innovative glyco-immune checkpoints in the realm of cancer treatment. Indeed, the structure of the N-terminal V-set domain of Siglec-7 interacting with a synthetic analog of GT1b was solved[53].

In the context of Siglecs-gangliosides axis of the modulation of immune response and tumour microenvironment, antibody therapies have been investigated in pre-clinical and clinical studies[54]

and synthetic mimetics of carbohydrate portions of GD2 and GD3 gangliosides have been tested as potential vaccines[55]. However, the antibody therapies only partially succeed, and ganglioside mimetic vaccines need improvement to be effective against a wider range of tumours.

Therefore, molecular level understanding of gangliosides interaction with Siglec-7 not only could shed light on the mechanisms underlying immune evasion in cancer cells but it could also help the development of targeted strategies to modulate these interactions. To this aim, we combined structural biology methodologies, NMR techniques, chemical-physical studies, and computational approaches to provide information on binding affinities and 3D models of Siglec-7 interacting with its preferred ganglioside, GD3, and structurally related derivatives.

We here expressed the full extracellular domain (FED) in human cell lines (HEK293) and, for the first time, we assigned the carbohydrate recognition domain (CRD) of Siglec-7 using NMR spectroscopy.

The combination of STD NMR, tr-NOESY and protein-based NMR with MD simulations and Redmat calculations allowed to achieve 3D models of the Siglec-7-ganglioside complexes, highlighting the interactions occurring at the interface.

GD3 was used as the main ganglioside ligand for Siglec-7. To investigate whether the glycosidic linkages of gangliosides could affect their binding accommodation, we decided to further study the molecular binding with modified GD3 structures, which maintained the disialylated Neu5Ac-α-(2-8)-Neu5Ac portion, varying the linkage of the internal sialic acid to galactose, via α-(2-3) and α-(2-6) glycosidic linkages. Variations in elongation and in type of glycosidic linkage between the internal sialic acid (**K**) and the galactose unit (**D**) of Gb3-derivatives (DSGb3α3 and DSGb3α6) allowed to obtain more detailed information on gangliosides recognition from Siglec-7. Indeed, all these complexes show affinity in the micromolar range, stronger for GD3 as expected, and their formation is enthalpically and entropically favored in all cases, with similar negative Gibbs free energy ($\Delta G < 0$). The reason is that the sialic acid moieties of GD3, DSGb3α3 and DSGb3α6 are the only residues involved in the interaction with Siglec-7, with a bias for the terminal sialic acid of the Neu5Ac-α-(2-8)-Neu5Ac glycosidic linkage, while the other sugars result solvent exposed and, as they make any contacts with the protein, consequently, represent the more flexible portions of the ligands. Furthermore, in order to define the binding site of Siglec-7, protein-based NMR experiments were performed by adding increasing amounts of each ganglioside in solution. The results showed some similarities on the effects of binding site residues. The amino acids more

affected by the presence of gangliosides correspond to those located in the vicinity of Arg124, which is crucial for sialoglycan binding, exhibiting consistent perturbations. For example, Gly128, as well as the region close to and partially including GG' loop, the residues from Ile130 to Tyr134$^{GG'}$, were subjected to CSP (Gly128, Ile130, Lys131, Trp132$^{GG'}$, Tyr134$^{GG'}$) and decrease of signal intensity (Lys131 and Asn133$^{GG'}$) in all the three complexes. These results support the importance of the binding site neighboring Arg124 that always establishes key salt bridge with the carboxylate group of the terminal sialic acid. Additionally, other CSPs in common to the three complexes were shown for Tyr26, Ser27 and His62^{BC}. In particular, the terminal sialic acid (**N**) of all gangliosides is involved in making hydrogen bonds mainly with its carboxylate group, NH of the acetyl group and glycerol chain, but we also observed the vicinity of this residue to the aromatic Trp132$^{GG'}$ that is stabilized by π-π interactions with Tyr26, both amino acids always affected by CSPs. Moreover, two different lysine residues are also perturbed in the binding site of Siglec-7: a CSP of Lys135$^{GG'}$, observed in protein complex with both GD3 and DSGb3α3 (ligands **1** and **2**, respectively), and signal intensity decrease of Lys127, detected in protein complex with both GD3 and DSGb3α6 (ligands **1** and **3**, respectively). Overall, the numerous chemical shift perturbations and the lower effects of signal intensity decrease of the Siglec-7 binding site agreed with the binding affinities calculated by fluorescence analyses, resulting in the micromolar range.

These findings, combined with the results of NMR assay and STD experiments, are instrumental in delineating binding site characteristics influenced by interactions with ligands **1**, **2**, and **3**. By incorporating the unique effects of each ligand on the protein structure and dynamics, a deep understanding of ligand-protein interactions and their impact on binding site architecture can be achieved for further optimization and drug development.

Therefore, exploring the structural diversity of gangliosides and their dynamic interplay with Siglec-7 provides understanding of the intricate cellular communication processes, offering potential avenues for therapeutic interventions aimed at restoring immune recognition and enhancing anti-cancer immune responses.

Acknowledgements

AS, MTM, MF, LPC, GRG acknowledge H2020-MSCA-ITN-2020 (contract n° 956758). AS acknowledges Ministry of Education, Universities and Research, PRIN MUR 2022 (2022ZEZS45) and PRIN MUR PNRR 2022 (P2022M457Z). AS, RM acknowledge PNRR, Missione 4 – Componente 2 – NextGenerationEU - Partenariato Esteso INF-ACT - One Health Basic and Translational Research Actions Addressing Unmet Needs on Emerging Infectious Diseases MUR: PE00000007. The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy centre. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 re-search and innovation program under grant agreement No 851356 to R.M. This work benefited from STSM funding to CDC, OV and AS by COST Action (CA18103 INNOGLY).

Methods

Sequence design of the carbohydrate recognition domain (CRD) of Siglec-7

The carbohydrate recognition domain (CRD) of Siglec-7 was added to pET29b(+) vector in between NdeI and XhoI restriction sites. The vector sequence was designed as previously published [56] [iii]. In brief, the coding region containing Glu18-His148 sequence from the full-length mature protein was codon optimized by Twist Bioscience. An additional modification to the lectin domain was done by replacing Cys41 with a serine residue, to avoid unspecific disulphide bridge formation.

Expression and purification of the carbohydrate recognition domain (CRD) of Siglec-7

The Siglec-7 CRD carrying-plasmid was transformed into BL21(DE3) Codon Plus RIPL or Rosetta 2 cells with the specific antibiotic selection and cultured at 37°C, at 180 rpm, in either LB or M9 culture medium, correspondingly supplemented with 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 0.2 % v/v solution Q, 1 µg/mL thiamine, 1 µg/mL biotin, 1 g $(^{15}NH_4)_2SO_4$ and 3 g $^{13}C_6H_{12}O_6$ for isotopic labeling. The overexpression was induced at OD_{600} = 0.8 with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and incubated further for 24 h at 25 °C under 180 rpm. The cells were harvested at 7500 rpm, for 20 minutes, using a Sorvall RC 6 Plus Superspeed Centrifuge (Thermo Scientific) and then collected and resuspended in the lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and subsequently sonicated for 10 cycles (30 seconds ON, 2.5 minutes OFF) at 70 % amplitude, using a Vibra-Cell sonicator (Sonics & Materials). The lysate has been ultracentrifuged for 40 minutes at 40 000 rpm with Optima LE-80K Ultracentrifuge (Beckman). Siglec-7 CRD is then found incorporated into inclusion bodies. While supernatant was discarded, the inclusion bodies were resuspended in 8 M urea lysis buffer (20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and re-sonicated. The solubilization of the inclusion bodies has been repeated for improved recovery yield of the protein of interest. The soluble, unfolded fraction containing Siglec-7 CRD has been further subjected to IMAC purification on a HisTrap FF (5 ml, GE Healthcare). The column was previously equilibrated with 20 mM potassium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4 and the protein was loaded at a flow rate of 3 mL/min, in closed cycle, for 2 h. The column was eluted at a 4 mL/min flow

rate, with high-concentration imidazole buffer (20 mM potassium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4) (Figure 1A-B). To proceed with the refolding of the unfolded, purified protein of interest, a series of dialysis were set in place. The protein was placed in a 10 kDa cut-off membrane and each dialysis was performed overnight by changing the buffer in a 2 L tank, as it follows: 1st dialysis) 20 mM potassium phosphate, 300 mM NaCl, 4 M urea; 2nd dialysis) 20 mM potassium phosphate, 300 mM NaCl, 2 M urea; 3rd dialysis) 20 mM potassium phosphate, 150 mM NaCl. Siglec-7 CRD, now refolded, is collected from the membrane and ultracentrifuged out from the formed aggregates at 40 000 rpm, for 40 minutes and then, filtered on a 0.2 µm micropore membrane. The final purification was achieved by a size-exclusion chromatography on a HiLoad 26/60 Superdex 75 pg (GE Healthcare) coupled on an AKTA Go FPLC system, equilibrated with 20 mM potassium phosphate, 50 mM NaCl, pH 7.4 (Figure 1C-D).

Backbone resonance assignment of Siglec-7

NMR protein experiments for backbone assignment were acquired on a sample of [U-^{15}N-^{13}C] Siglec-7 at the concentration of 400 µL in 600 µL of aqueous buffered solution (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM of protease inhibitors, 10% D_2O, in a 5 mm NMR tube. Triple resonance experiments for protein NMR assignment HNCA, HNCACB, HNCO and CBCAcoNH were recorded at 298 K on a Bruker's Avance™ NEO 900 MHz spectrometer, equipped with TCI cryo-probe. 3D HNcaCO was recorded at 298 K on a Bruker's Avance™ 500 MHz spectrometer. 93 % of the amino acid sequence from Y26 to T147 was assigned, excluding the 5 proline residues. Data acquisition and processing was performed on TOPSPIN 4.1.1 software and spectra were analyzed by using CARA (Computer Aided Resonance Assignment) software) http://wiki.cara.nmr.ch/FrontPage.

Protein-based NMR experiments

For ligand binding studies, 2D ^{1}H-^{15}N HSQC NMR experiments were recorded on samples of 200 µM [U-^{15}N] Siglec-7 CRD in 200 µL of aqueous buffered solution (20 mM potassium phosphate, pH 7.4, 50 mM NaCl, 0.01% NaN_3, 1 mM of protease inhibitors, 10% D_2O in a 3 mm NMR tube. Experiments were acquired on a Bruker's Avance™ NEO 900 MHz spectrometer, equipped with triple resonance TCI cryo-probe. The spectra were acquired using 32 scans with 2048 data point in t2, 128 increments in the indirect dimension (t1), a recycle delay of 1.2 seconds, and the

temperature was kept at 298 K. The interaction of Siglec-7 CRDwith the ligands was investigated by adding increasing amounts ligands 1, 2 or 3, to three different tubes containing Siglec-7 CRD, to reach ligand concentrations of 12.5, 25, 50, 100, 200, 400, 800, 1600 and 3200 μM. 2D ^{1}H-^{15}N HSQC spectra were added upon the addition of each ligand aliquot. Data acquisition and processing were performed with TOPSPIN 4.1.1 software and the spectra were analyzed using CARA. Chemical Shift Perturbations (CSP) were evaluated with the formula: $\Delta\delta = \frac{1}{2}\sqrt{\Delta\delta_{H^2} + (\frac{\Delta\delta_N}{5})^2}$ [iv].

Dissociation constant calculation

Dissociation constants for the interactions of Siglec-7 with ligands **1**, **2** and **3** were calculated by simulating and fitting spectra from an experimental NMR titration using NMR-TITAN software. Data was fitted to a 2-state binding model for an HSQC pulse program. 10 of the most affected residues by the interactions were used for the fitting: Y26, H62, E126, K127, G128, N129, I130, K131, W132 and N133. Bootstrap error analysis was run, and the results are reported on supporting table on the supplementary materials.

Ligand-based NMR experiments.

Saturation Transfer Difference (STD) and transferred-NOESY NMR experiments were recorded on a Bruker AVANCE NEO 600-MHz equipped with a cryo-probe and data acquisition and processing were performed with TOPSPIN 4.1.1 software. Samples were prepared in phosphate saline deuterated buffer (10 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 10 mM NaN_3, pH 7.4) at 298 K. [D4] (trimethylsilyl)propionic acid, sodium salt (TSP, 1 %) was used as internal reference. A protein concentration of 25 μM was use and protein:ligand molar ratio was 1:50 for all the mixtures.

STD NMR experiments were acquired with shaped pulse train for saturation on f2 channel alternating between on and off resonance with 20 ms spinlock pulse applied to suppress protein signals. The acquisition was set with 65 k data points and 104 number of scans. The protein resonances were selectively irradiated by 40 Gauss pulses with a length of 50 ms, using the off-resonance pulse frequency at 40 ppm and on-resonance pulses at 7.5 ppm and 0 ppm. The STD NMR spectra were carried out using a saturation time of 2 s. STD NMR effects were determined

using the formula $(I_0 - I_{sat})/I_0$, where I_{sat} is the relative intensity of the STD NMR signal and I_0 the peak intensity of an unsaturated reference spectrum (off-resonance). The strongest STD NMR response was set to 100% while all the other STD signals were normalized to this value to provide ligands' epitope maps[58].

2D ^{1}H-^{1}H NOESY experiments were acquired by using data sets of 4096x800 points and 100-400 ms as mixing times. Proton – proton cross relaxation rates were calculated by integration of cross peaks normalized against the corresponding diagonal peak. ^{1}H-^{1}H distances were calculated using the following equation: $r_{ij} = r_{ref}\sqrt[6]{\sigma_{ref}/\sigma_{ij}}$, where r_{ij} is the unknown distance to be estimated, r_{ref} is the reference interproton distance, σ_{ref} is the cross-relaxation rate of the NOE cross peak of interest and σ_{ij} is the cross-relaxation rate of reference[59].

Fluorescence analysis

Stationary-state fluorescence spectra were collected on a Fluoromax-4 spectrophotometer (Horiba, Edison, NJ, USA). Emission spectra were recorded in the range of 310 to 450 nm with excitation at 295 nm. The slit widths were set to 5 nm for excitation and to 5 nm for emission wavelength. A quartz cuvette with a pathlength of 1 cm and a volume of 0.2 mL was used. The starting solution of Siglec-7 CRD protein, prepared at a fixed concentration of 4 μM in PBS buffer (pH 7.4), was titrated by adding small aliquots of each ligand (**1**-**3**). The experiments were carried out at 10°C, 25°C and 35°C and triplicate at each temperature was considered. For all ligands analyzed, the fluorescence intensity was quenched. The binding curve was obtained by fitting the data using nonlinear regression with a 1:1 binding model using the function described by Ribeiro *et al* [60].[v] (plotting the ratio between the fluorescence intensity at each addition of ligand (F) and the fluorescence intensity of the protein in absence of ligand (F_0) against the to the total ligand concentration [L] in μM.

Isothermal titration calorimetry

The ITC experiment was carried out by means of a Nano ITC-III from TA Instruments (New Castle, DE, USA). Briefly, a 500 μM solution of GD3 was placed in the calorimetry vessel. Then, a solution of Siglec-7 CRD, at the concentration of 62.5 μM was injected (injection volume of 15 μL) into the calorimetry vessel. The heat of dilution of the protein was evaluated by performing

injection in the calorimetry vessel containing buffer only. The temperature was set to 25 °C and the stirring speed to 250 rpm. The ΔHb° was calculated by integration of the peaks, then subtraction of the heat of dilution, and normalizing them by the amount of injected protein[41].

MD simulations

Glycans were generated on GLYCAM website[61]. and MD simulations were performed by using AMBER 18[62,63]. The prmtop and inpcrd files were generated with the tLEaP module. The force fields were GLYCAM06j-1 for carbohydrates parameters and protein.ff14SB for SLBR-N. Complexes were properly prepared for the MD simulations by adding counter ions to neutralize the systems using the Leap module and placing them in octahedral boxes with explicit TIP3P water molecules. MD simulations were run by using the CUDA implementation of PMEMD in AMBER18. Minimization steps of all complexes were performed using Sander. The smooth particle mesh Ewald method was employed to describe the electrostatic attractions in the system while applying periodic boundary constraints and the grid spacing was 1 Å. The system underwent the first annealing gradually and gently over a 25-ps period from 100 °K to 300 °K. Throughout 50 ps, a steady temperature of 300 °K was maintained with progressive energy minimizations. The MD coordinates were gathered to acquire 10000 frames of the progression of the dynamics. Using the K-mean algorithm implemented in the ptraj module of the AMBER18 software, the trajectories were submitted to cluster analyses, to obtain the main representative poses. MD simulations were visualized by using VMD program[64].

RedMat

The analysis was employes considering the STD NMR effects and the best poses of the complexes between Siglec-7 and ligands **1** and **2**. The parameters were set as follows: NMR spectrometer frequency at 600 MHz, the concentration of protein at 25 μM, the concentration of ligand at 1000 μM, the cut-off distance at 10 Å and the dissociation constants as derived from the fluorescence analysis. The goodness of the complexes and the accordance between the theoretical and experimental data were evaluated from the R-factor (R-NOE) values, 0.2 and 0.3 for ligands **1** and **2**, respectively, and calculated using the following formula:

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

Where STDexp,i is the experimental STD value for a generic proton I, while STDcalc,i is the STD value simulated with the RedMat algorithm.

References

1. Schauer, R. & Kamerling, J. P. Exploration of the Sialic Acid World. in Advances in Carbohydrate Chemistry and Biochemistry vol. 75 1–213 (Academic Press Inc., 2018).
2. Di Carluccio, C. et al. Exploring the fascinating world of sialoglycans in the interplay with Siglecs. Carbohydr Chem 44, 31–55 (2021).
3. Varki, A. et al. Essentials of Glycobiology FOURTH EDITION. https://www.ncbi.nlm.nih.gov/books/NBK310274/ (2017).
4. Klaas, M. & Crocker, P. R. Sialoadhesin in recognition of self and non-self. in Seminars in Immunopathology vol. 34 353–364 (2012).
5. Schnaar, R. L. Gangliosides as Siglec ligands. Glycoconjugate Journal vol. 40 159–167 Preprint at https://doi.org/10.1007/s10719-023-10101-2 (2023).
6. Groux-Degroote, S., Guerardel, Y., & Delannoy, P. (2017). Gangliosides:Structures, biosynthesis, analysis, and roles in cancer. ChemBioChem, 18(13), 1146–1154.
7. Schnaar, R. L., Sandhoff, R., Tiemeyer, M., & Kinoshita, T. (2022). Glycosphingolipids. In th, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, & P. H. Seeberger (Eds.), Essentials of Glycobiology (pp. 129-140).
8. Schnaar, R. L. Gangliosides of the Vertebrate Nervous System. Journal of Molecular Biology vol. 428 3325–3336 Preprint at https://doi.org/10.1016/j.jmb.2016.05.020 (2016).
9. Ohmi, Y. et al. Gangliosides are essential in the protection of inflammation and neurodegeneration via maintenance of lipid rafts: Elucidation by a series of ganglioside-deficient mutant mice. J Neurochem 116, 926–935 (2011).
10. Regina Todeschini, A. & Hakomori, S. itiroh. Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains. Biochimica et Biophysica Acta - General Subjects vol. 1780 421–433 Preprint at https://doi.org/10.1016/j.bbagen.2007.10.008 (2008).
11. Abad-Rodriguez, J. & Robotti, A. Regulation of axonal development by plasma membrane gangliosides. Journal of neurochemistry vol. 103 Suppl 1 47–55 Preprint at https://doi.org/10.1111/j.1471-4159.2007.04713.x (2007).
12. Schnaar, R. L., Gerardy-Schahn, R. & Hildebrandt, H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. Physiol Rev 94, 461–518 (2014).

13. Yue, J., Huang, R., Lan, Z., Xiao, B. & Luo, Z. Abnormal glycosylation in glioma: related changes in biology, biomarkers and targeted therapy. Biomarker Research vol. 11 Preprint at https://doi.org/10.1186/s40364-023-00491-8 (2023).
14. Webb, T. J. et al. Molecular identification of GD3 as a suppressor of the innate immune response in ovarian cancer. Cancer Res 72, 3744–3752 (2012).
15. Furukawa, K., Hamamura, K., Ohkawa, Y., Ohmi, Y. & Furukawa, K. Disialyl gangliosides enhance tumor phenotypes with differential modalities. Glycoconj J 29, 579–584 (2012).
16. Sasaki, N., Toyoda, M. & Ishiwata, T. Gangliosides as signaling regulators in cancer. International Journal of Molecular Sciences vol. 22 Preprint at https://doi.org/10.3390/ijms22105076 (2021).
17. Torres Demichelis, V., Vilcaes, A. A., Iglesias-Bartolomé, R., Ruggiero, F. M. & Daniotti, J. L. Targeted Delivery of Immunotoxin by Antibody to Ganglioside GD3: A Novel Drug Delivery Route for Tumor Cells. PLoS One 8, (2013).
18. Boelaars, K. & van Kooyk, Y. Targeting myeloid cells for cancer immunotherapy: Siglec-7/9/10/15 and their ligands. Trends in Cancer vol. 10 230–241 Preprint at https://doi.org/10.1016/j.trecan.2023.11.009 (2024).
19. Bordoloi, D. et al. Siglec-7 Glyco-Immune Binding MAbs or NK Cell Engager Biologics Induce Potent Antitumor Immunity against Ovarian Cancers. https://www.science.org (2023).
20. Rodrigues, E. & Macauley, M. S. Hypersialylation in cancer: Modulation of inflammation and therapeutic opportunities. Cancers vol. 10 Preprint at https://doi.org/10.3390/cancers10060207 (2018).
21. van de Wall, S., Santegoets, K. C. M., van Houtum, E. J. H., Büll, C. & Adema, G. J. Sialoglycans and Siglecs Can Shape the Tumor Immune Microenvironment. Trends in Immunology vol. 41 274–285 Preprint at https://doi.org/10.1016/j.it.2020.02.001 (2020).
22. Theruvath, J. et al. Anti-GD2 synergizes with CD47 blockade to mediate tumor eradication. Nat Med 28, 333–344 (2022).
23. Krengel, U. & Bousquet, P. A. Molecular recognition of gangliosides and their potential for cancer immunotherapies. Frontiers in Immunology vol. 5 Preprint at https://doi.org/10.3389/fimmu.2014.00325 (2014).
24. Yamaji, T., Teranishi, T., Alphey, M. S., Crocker, P. R. & Hashimoto, Y. A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to α2,8-disialyl and branched α2,6-sialyl residues. A comparison with Siglec-9. Journal of Biological Chemistry 277, 6324–6332 (2002).

25. Ito, A., Handa, K., Withers, D. A., Satoh, M. & Hakomori, S. itiroh. Binding specificity of siglec7 to disialogangliosides of renal cell carcinoma: Possible role of disialogangliosides in tumor progression. FEBS Lett 504, 82–86 (2001).
26. Ramos, R. I. et al. Upregulation of cell surface GD3 ganglioside phenotype is associated with human melanoma brain metastasis. Mol Oncol 14, 1760–1778 (2020).
27. Malisan, F. & Testi, R. GD3 Ganglioside and Apoptosis. www.bba-direct.com.
28. Bärenwaldt, A. & Läubli, H. The sialoglycan-Siglec glyco-immune checkpoint–a target for improving innate and adaptive anti-cancer immunity. Expert Opinion on Therapeutic Targets vol. 23 839–853 Preprint at https://doi.org/10.1080/14728222.2019.1667977 (2019).
29. Daly, J., Carlsten, M. & O'Dwyer, M. Sugar free: Novel immunotherapeutic approaches targeting siglecs and sialic acids to enhance natural killer cell cytotoxicity against cancer. Frontiers in Immunology vol. 10 Preprint at https://doi.org/10.3389/fimmu.2019.01047 (2019).
30. Hashimoto, N. et al. The ceramide moiety of disialoganglioside (GD3) is essential for GD3 recognition by the sialic acid– binding lectin SIGLEC7 on the cell surface. Journal of Biological Chemistry 294, 10833–10845 (2019).
31. Liu, J. et al. Ganglioside GD3 synthase (GD3S), a novel cancer drug target. Acta Pharmaceutica Sinica B vol. 8 713–720 Preprint at https://doi.org/10.1016/j.apsb.2018.07.009 (2018).
32. Minasian, L. M. et al. A Phase I Study of Anti-GD3 Ganglioside Monoclonal Antibody R24 and Recombinant Human Macrophage-Colony Stimulating Factor in Patients with Metastatic Melanoma.
33. Houghton, A. N. et al. Mouse Monoclonal IgG3 Antibody Detecting GD3 Ganglioside: A Phase I Trial in Patients with Malignant Melanoma (Complement Activation/Inflammation/Immunotherapy). Medical Sciences vol. 82 https://www.pnas.org (1985).
34. Li, P. et al. Chemoenzymatic Synthesis of DSGb5 and Sialylated Globo-series Glycans. Angewandte Chemie 131, 11395–11400 (2019).
35. Kim, CH. (2020). Glycosylation. In: Ganglioside Biochemistry. Springer, Singapore. https://doi.org/10.1007/978-981-15-5815-3_1
36. Nepravishta, R. et al. Fast Quantitative Validation of 3D Models of Low-Affinity Protein–Ligand Complexes by STD NMR Spectroscopy. J Med Chem (2024) doi:10.1021/acs.jmedchem.4c00204.
37. Di Carluccio, C. et al. Molecular recognition of sialoglycans by streptococcal Siglec-like adhesins: Toward the shape of specific inhibitors. RSC Chem Biol 2, 1618–1630 (2021).
38. Gonzalez-Gil, A. & Schnaar, R. L. Siglec ligands. Cells vol. 10 Preprint at https://doi.org/10.3390/cells10051260 (2021).
39. R. Oliva, et al. J. Phys. Chem. B 2021, 125, 2, 539–546

40. Waudby, C., Ramos, A., Cabrita, L. et al. Two-Dimensional NMR Lineshape Analysis. Sci Rep 6, 24826 (2016). doi: 10.1038/srep24826
41. Nieto-Fabregat, F.; Lenza, M. P.; Marseglia, A.; Di Carluccio, C.; Molinaro, A.; Silipo, A.; Marchetti, R. Beilstein J. Org. Chem. 2024, 20, 2084–2107
42. Soares, C. O., Grosso, A. S., Ereño-Orbea, J., Coelho, H. & Marcelo, F. Molecular Recognition Insights of Sialic Acid Glycans by Distinct Receptors Unveiled by NMR and Molecular Modeling. Frontiers in Molecular Biosciences vol. 8 Preprint at https://doi.org/10.3389/fmolb.2021.727847 (2021).
43. Forgione, R. E. et al. Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. iScience 23, (2020).
44. Di Carluccio, C. et al. Characterisation of the Dynamic Interactions between Complex N-Glycans and Human CD22. ChemBioChem 21, 129–140 (2020).
45. Forgione, R. E. et al. Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. iScience 23, (2020).
46. Forgione, R. E. et al. Structural basis for Glycan-receptor binding by mumps virus hemagglutinin-neuraminidase. Sci Rep 10, (2020).
47. Valentino, L., Moss, T., Olson, E., Wang, H., Elashoff, R., & Ladisch, S. (1990). Shed tumor gangliosides and progression of human neuroblastoma. *Blood, 75*(7), 1564–1567.
48. Groux-Degroote, S., & Delannoy, P. (2021). Cancer-associated glycosphingolipids as tumor markers and targets for cancer immunotherapy. International Journal of Molecular Sciences, 22(11), 6145.
49. Li, Q., Sun, M., Yu, M., Fu, Q., Jiang, H., Yu, G., & Li, G. (2019). Gangliosides profiling in serum of breast cancer patient: GM3 as a potential diagnostic biomarker. Glycoconjugate Journal, 36(5),419–428.
50. Van Vliet et al, 2023
51. Cutillo, G., Saariaho, A.-H., & Meri, S. (2020). Physiology of gangliosides and the role of antiganglioside antibodies in human diseases. Cellular & Molecular Immunology, 17(4), 313–322.
52. Iltis, C., Seguin, L., Cervera, L., Duret, L., Hamidouche, T., Kunz, S., Croce, O., Delannoy, C., Gueŕardel, Y., Allain, F., Moudombi, L., Hofman, P., Cosson, E., Guglielmi, J., Pourcher, T., Braud, V. M., Shkreli, M., Michallet, M.-C., Feral, C. C., Cherfils-Vicini, J. (2021). A ganglioside-based senescence-associated immune checkpoint. bioRxiv, 2021.2004.2023.440408.
53. Attrill, H. et al. Siglec-7 undergoes a major conformational change when complexed with the α(2,8)-disialylganglioside GT1b. Journal of Biological Chemistry 281, 32774–32783 (2006).

54. Cavdarli, S., Delannoy, P., & Groux-Degroote, S. (2020). O-acetylated gangliosides as targets for cancer immunotherapy. Cells, 9(3), 741.
55. Tong, W., Maira, M., Roychoudhury, R., Galan, A., Brahimi, F., Gilbert, M., Cunningham, A. M., Josephy, S., Pirvulescu, I., Moffett, S., & Saragovi, H. U. (2019). Vaccination with tumorganglioside glycomimetics activates a selective immunity that affords cancer therapy. Cell Chemical Biology, 26(7), 1013-1026 e1014.
56. Pröpster JM, Yang F, Ernst B, Allain FHT, Schubert M. Functional Siglec lectin domains from soluble expression in the cytoplasm of Escherichia coli. Protein Expr Purif. 2015; 109:14-22. doi:10.1016/j.pep.2015.01.005
57. Williamson, M. Chemical shift perturbation. Modern Magnetic Resonance 2018, 995-1012. Giuntini, S.; Balducci, E.; Cerofolini, L.; Ravera, E.; Fragai, M.; Berti, F.; Luchinat, C. Characterization of the conjugation pattern in large polysaccharide–protein conjugates by NMR spectroscopy. Angewandte Chemie 2017, 129 (47), 15193-15197.
58. Mayer, M.; Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. journal of the american chemical society 2001, 123 (25).
59. Meyer, B.; Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angewandte Chemie International Edition 2003, 42 (8), 864-890.
60. Ribeiro, M.M.B.; Franquelim, H.G.; Castanho, M.A.R.B.; Veiga, A.S. Molecular interaction studies of peptides using steady-state fluorescence intensity. Static (de)quenching revisited. J. Pept. Sci. 2008, 14, 401–406.
61. (16) https://glycam.org/
62. Case, D. A.; Betz, R.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N. AMBER 2016
63. Nieto-Fabregat F., Lenza MP, Marseglia A., Di Carluccio C., Molinaro A., Silipo A., Marchetti R. Computational toolbox for the analysis of protein–glycan interactions. Beilstein J. Org. Chem. 2024, 20, 2084–2107.
64. Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. Journal of chemical theory and computation 2013, 9 (7), 3084-3095.

Molecular dissection of *Fusobacterium nucleatum* strain 10953 interaction with Siglec-7

*Article in preparation

Molecular dissection of *Fusobacterium nucleatum* strain 10953 interaction with Siglec-7

Cristina Di Carluccio[a,b] & Ferran Nieto-Fabregat[a], Alessandro Antonio Masi[a], Antonio Molinaro[a,b]…, Roberta Marchetti[a], Alba Silipo[a,b]

Giulia Roxana Gheorghita[c,d], Luis Padilla-Cortés[c], Linda Cerofolini[c], Marco Fragai[c],

Celeste Abreu[e], Ondřej Vaněk[e],

Nathalie Juge[f]

[a] Department of Chemical Sciences and Task Force for Microbiome Studies, University of Naples Federico II, Via Cinthia 4, 80126, Naples, Italy

[b] CEINGE-Biotecnologie Avanzate Franco Salvatore, Via Gaetano Salvatore 486, 80145, Napoli, Italy

[c] Magnetic Resonance Centre (CERM), CIRMMP and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

[d] Giotto Biotech s.r.l., Sesto Fiorentino, Italy

[e] Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, 12800, Prague (Czech Republic)

[f] The Gut Microbiome and Health and Food Safety Institute Strategic Programme, Norwich Research Park, Quadram Institute Bioscience, Norwich, UK

Introduction

Fusobacterium nucleatum (*Fn*) is a Gram-negative obligate anaerobe, among the most abundant species residing in the human oral cavity and rarely found in the lower gastrointestinal tract of healthy individuals. *Fn* has been identified as one of the pathobionts that outgrow during dysbiosis that precedes periodontal disease[1,2]. *F*n can disseminate systemically in different body sites, such as genitourinary tracts, and overcome placental and blood-brain barriers. A large number of studies suggest that *Fn* is closely related to the developments of various systemic diseases, including rheumatoid arthritis, Alzheimer's disease, adverse pregnancy outcomes and gastrointestinal disorders such as IBD[3]. In addition, *Fn* has gained attention as emerging cancer-associated bacterium, overabundant in various types, including colorectal (CRC), pancreatic, esophageal, and breast cancers[4-6], and associated with shorter patients survival. In particular, CRC is accompanied by dysbiosis of the *intestinal* microbiota with a reduction of protective or beneficial bacteria such as *Clostridium* and *Faecalibacterium* and an increase in cancer-associated polyketide synthase *E. coli*, enterotoxigenic *Bacteroides fragilis*, and as most abundant species *Fusobacterium nucleatum*[2, 7-10]. Indeed, *Fn* promotes colorectal cancer progression by several mechanisms, including inhibition of host antitumor immunity, innate immune cell modulation, activation of cell proliferation, promotion of cellular invasion, induction of chronic inflammation and immune evasion, mobilization of immune cells in the tumor microenvironment[10-12].

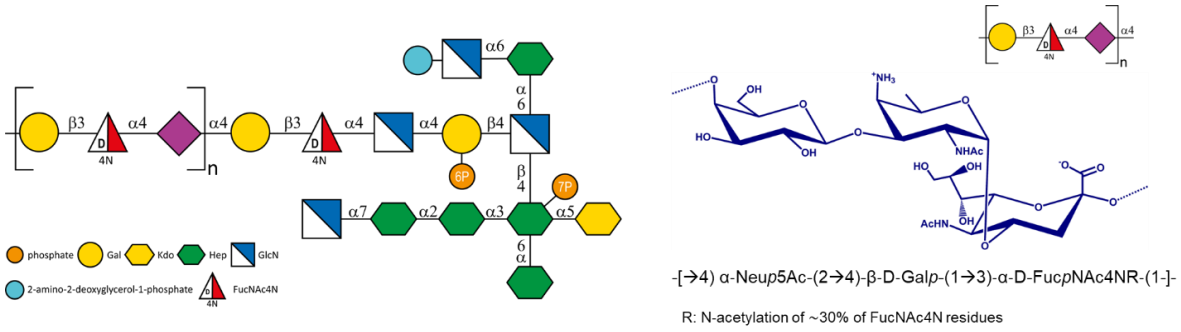
A possible virulence factor in *Fn* is the lipopolysaccharide (LPS), a microbe-associated molecular pattern located on the outer membrane of Gram-negative bacteria. LPS is an amphiphilic molecule with a general structural architecture composed by i) a glycolipid moiety named lipid A and ii) a heteropolysaccharide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

The sialic acid-binding immunoglobulin-type lectins (Siglecs) comprise a family of 15 cell surface proteins, mostly inhibitory receptors, which can be divided in two different classes according to their sequence and evolutionary similarity: the main subset of CD33-related Siglecs and the conserved Siglecs' subgroup (Scheme S1). Siglecs recognize sialic acid (Sia) containing glycans,

by means of a key conserved arginine residue present in the N-terminal domain that establishes a salt bridge with the carboxylate group of Sia (Scheme S1). Therefore, engagement of sialylated glycoprotein and glycolipids on all mammalian cells by Siglecs induces tolerance to self-antigens and prevents unwanted autoimmune responses. However, hypersialylation, frequent trait of tumor cells, increases Sialoglycans-Siglecs interactions to evade immune surveillance in the favourable immunosuppressive tumour microenvironments. This makes Siglecs novel emerging immune checkpoint for cancer therapy[13]. In addition, a number of bacterial pathogens developed cell surface Sia mimicry to engage Siglecs as an immune evasion strategy[14-18]. Indeed, *E. coli* K1 and *N. meningitidis*, important agents of bacterial meningitis, group B *streptococcus*, *N. gonorrhea*, *C. jejuni* possess sialylated capsules and/or LPS to evade immunosurveillance.

Siglec-7 (CD328), a member of the CD33-related Siglecs subgroup, is primarily expressed on the surface of innate lymphoid human NK cells, and also found on T cells and dendritic cells [19]. It exhibits a binding preference for α2,8-linked disialylated ligands and internally branched α2,6-sialyl residues. Siglec-7 consists of an extracellular region, containing an N-terminal V-set domain representing the carbohydrate binding region and two C2-type Ig-like domains, and an immunoreceptor tyrosine-based inhibitory (ITIM) motif in the cytosolic region (Figure S1)[20]. Siglec-7 *trans* interactions with cognate ligands drives phosphorylation of its cytoplasmatic ITIM domain motif that triggers the inhibition of NK cell pathways. Tumour cells and bacteria/pathogens can exploit this mechanism to evade immune recognition and facilitate their migration through the circulatory system. For example, group B *streptococcus* binds Siglec-7 on NK cells, inhibiting pyroptosis and suppressing NK cell sentinel activity [21]. The interaction between the LPS of *Campylobacter jejuni* and Siglec-7, which occurs through GQ1b-like epitopes and potentially contribute to Guillain-Barré syndrome, has also been reported[22]. We have already shown that *Fn ssp. animalis* ATTC19919, the most predominant in CRC, interacts with Siglec-7 expressed on innate immune cells, observed using either bacterial outer membrane vesicles and LPS. This interaction induced immunosuppression, thus likely contributing to the bacterial carcinogenic behaviour[23], and unveiled LPS O-antigen as potential ligand for Siglec-7. Furthermore, while the direct causal link between *Fn* and CRC has been highlighted, the molecular basis of Siglec-7 based immune subversion by cancer upon engagement by *Fn* surface-exposed glycans has not been fully disclosed yet. To contribute in filling this gap, we here isolated the LPS from *Fn ssp polymorphum* ATCC 10953, abundant in CRC tissues, whose LPS O-antigen composition consists in a peculiar

repeating unit of [→4)-α-Neu*p*5Ac-(2→4)-β-Gal*p*-(1→3)-α-Fuc*p*NAc4NAc-(1→], and combined complementary approaches, including NMR spectroscopy from both protein and ligand perspectives, computational methods, biochemical and biophysical techniques to demonstrated its binding to Siglec-7 and to reveal a novel binding mode.



Scheme 1. *SNFG schematic representation of Fusobacterium nucletum ATCC 10953 structure. A) OPS containing the o-antigen repeating unit and the core oligosaccharide[43].[43] B) O-antigen trisaccharide repeating unit: →4)-β-D-Galp-(1→3)-α-D-FucpNAc4NR-(1→4)-α-Neup5Ac-(2→; R=H (70%), R=Ac (30%).*

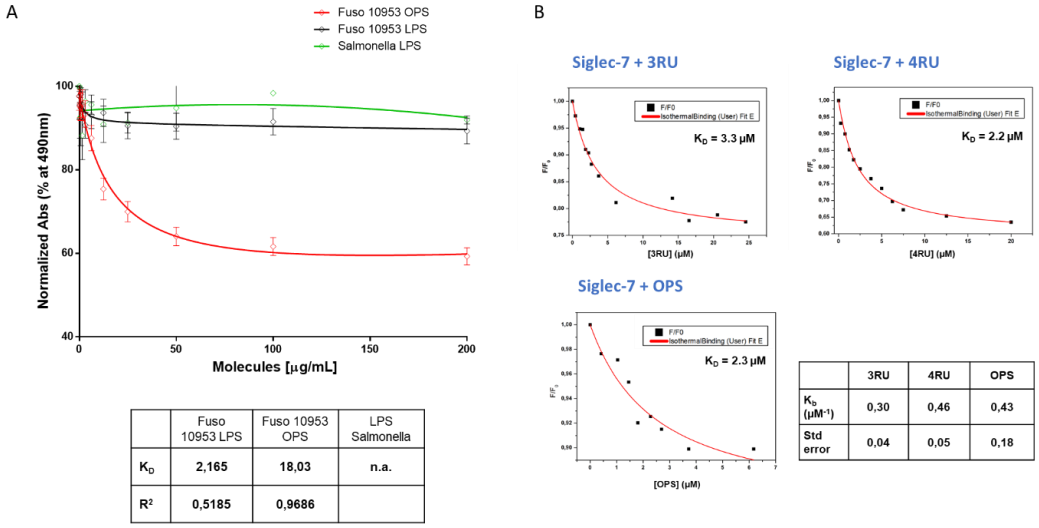
Results

The full extracellular domain of Siglec-7 was expressed in HEK293 GnTI- (human embryonic kidney) cells. The LPS from *Fn* ATCC 10953 (referred to as *Fn*10953) was extracted and purified[24] and the presence of a trisaccharide repeating unit [4)-α-Neu*p*5Ac-(2→4)-β-D-Gal*p*-(1→3)-α-D-Fuc*p*NAc4N*Ac*-]$_n$ was confirmed (Figure S2). Notably, the acetylation of Fuc*p*NAc4N (AAT) unit at N4 was around 25% [25]. Furthermore, oligomers containing an increasing number of O-antigen (OPS) repeating units were obtained by performing a mild acid hydrolysis which exploited the lability of sialic acid linkage. Structural insights on recognition and binding of *Fn*10953 O-antigen to Siglec-7 was then evaluated as follows (Figure S2).

***Fn*10953 LPS binds Siglec-7 and differentially activate blood cells**

To assess the recognition of *Fn*10953 LPS by Siglec-7, we evaluated the binding choosing a cell model overexpressing Siglec-7. To this aim several haematopoietic cell lines were tested, and finally RPMI8226 cells, a model of multiple myeloma, were identified as those most strongly expressing Siglec-7 on their surface (Figure S3). The cells were titrated with increasing concentrations of *Fn*10953 LPS and the isolated O-antigen and subsequently the fluorescence

signal associated with Siglec-7 was assessed. As shown in Figure 1A, both *Fn*10953 LPS and its O-antigen bound Siglec-7-expressing RPMI8226 cells (affinity of approximately 2 µg/ml for LPS and 18 µg/ml for OPS). We did not detect binding with *Salmonella* LPS, used as a negative control.



	Fuso 10953 LPS	Fuso 10953 OPS	LPS Salmonella
K_D	2,165	18,03	n.a.
R^2	0,5185	0,9686	

	3RU	4RU	OPS
K_b (μM^{-1})	0,30	0,46	0,43
Std error	0,04	0,05	0,18

Figure 1. A) *ELISA binding assay of Salmonella LPS (green line), Fn10953 LPS (black line) and Fn10953 full polysaccharide (OPS) (red line) with RPMI8226 cell line. In table are reported the K_d and R^2 value. The worse binding quality associated with the use of LPS could be due to the steric cluster generated by this molecule, which could cover the binding sites and generate a worse signal than OPS* ***B)*** *Fluorescence analysis of Siglec-7 with different oligomers from Fn10953. The quenching fluorescence titration of Siglec-7 with Fn10953 3RU, Fn10953 4RU and Fn10953 full polysaccharide (OPS) were fitted in a non-linear regression with One Site- Specific Binding model for the determination of the association constants K_b and the corresponding dissociation constants K_D.*

In addition, steady-state fluorescence analysis was undertaken to provide the binding affinities of Siglec-7 for *Fn*10953 O-antigen. A concentration dependent reduction in fluorescence intensity of Siglec-7 upon glycan addition was used to monitor the interaction, and the binding constants (K_b) were determined by non-linear regression analysis (Figure 1B) [26,27]. The trisaccharide containing a single repeating unit (1-Mer) showed no changes in fluorescence intensity, thus suggesting that it was not sufficient for the recognition by the protein (Figure S4A). On the contrary, the titration of longer oligomers containing up to four repeating units (4-Mer) quenched tryptophane residues of Siglec-7, therefore proving complex formation. In all cases, fluorescence results provided similar K_D values in the low micromolar range, fully comparable to the value obtained for the entire LPS O-antigen.

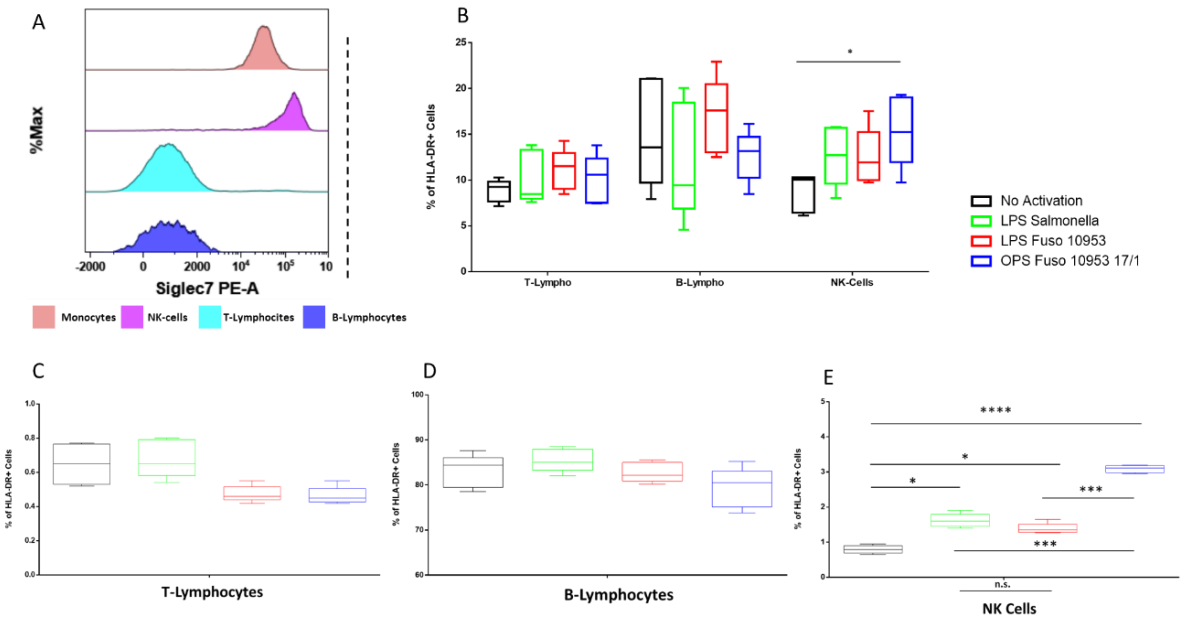


Figure 2. *Siglec-7 protein in blood cells.* ***A)*** *Cytofuorimetric analyses of Siglec-7 protein levels, in Monocytes (pink), NK (magenta), T-Lymphocytes (light blue) and B-Lymphocytes (blue). Percentage of HLA-DR+ cells in PBMC derived from 5 healthy donor (B) and in purified B-lymphocytes (C), T-Lymphocytes (D) and NK-cells (E) from the same subjects. Black bars represent no activated cells. Light green bar represents LPS Salmonella treated cells. Red bars represent Fn10953 LPS treated cells. Blue bars represent O-antigen Fn10953 LPS treated cells. *= p-value<0,05. ***= p-value< 0,001. **** p-value < 0,0001. n.s.= not significant. Paired t-test.*

To test the effect of *Fn*10953 LPS on blood cells, the level of surface expression of Siglec-7 was firstly assessed on all blood cells, resulting in highest expression levels on NK cells (Figure 2A and S5). The ability of *Fn*10953 LPS to activate individual blood components was investigated. PBMCs from 5 healthy donors were processed, and the activation levels of individual cells were assessed by measuring HLA-DR expression levels. T-lymphocytes, B-lymphocytes, and NK cells were identified, and HLA-DR expression levels on the gated cells were simultaneously assessed (as described in the gating strategy in Figure S6). As shown in Figure 2B, the increase in HLA-DR positive cells, compared to untreated cells, was significant only in the case of activation with *Fn*10953 O-antigen and only specific for NK cells. All other variations were not significant (Table S1). The individual components of PBMCs (T lymphocytes, B lymphocytes, and NK cells) were purified, and subsequently activated with *Fn*10953. No significant changes were observed in the degree of activation of T lymphocytes (Figure 2C), and B lymphocytes (Figure 2D). In the case of NK cells, in contrast to the general analysis above, significant changes were observed compared

to untreated cells in the case of both *Fusobacterium* and *Salmonella* activation with LPS. Activation of NK cells with *Fn*10953 O-antigen results in a significant increase in cell activation levels compared to both untreated and treated cells with both LPSs used (Figure 2E). These data, although preliminary, suggested an involvement of Siglec-7 in the activation process of NK cells in the presence of *Fn.*

Molecular details of Siglec-7 and *Fn*10953 O-antigen oligomers

Ligand- and protein-based NMR binding studies

The molecular recognition features of *Fn*10953 by Siglec-7 were unveiled by a combination of NMR and MD simulations. Firstly, NMR approaches, including STD NMR, tr-NOESY and relaxation experiments (Table S2), were employed to investigate the molecular details of the interaction between Siglec-7 and *Fn*10953 oligomers, elucidating the ligands binding epitopes [28] [28] and their conformational behaviour. In accordance to the fluorescence results (Figure S4A), no STD NMR effects, and therefore no binding, were detected on the system formed by Siglec-7 and 1-Mer (data not shown), confirming that a single repeating unit of *Fn*10953 O-antigen was not sufficient for recognition and binding by Siglec-7. On the other hand, NMR binding experiments conducted on the core-OPS fraction showed a selective binding of Siglec-7 to the O-antigen portion, while no STD NMR effects were observed from protons belonging to the core region. (Figure S4B).

With the aim to achieve accurate details on the minimal epitope required for O-antigen binding to Siglec-7, STD NMR experiments were performed on the mixture with shorter OPS oligomers. The STD NMR analysis carried out on the isolated 2-Mer (Figure 3A) in the presence of Siglec-7 shed light on the recognized epitope, revealing that the recognition primarily occurred through the internal sialic acid (**N'**) and the AAT (Fuc*p*NAc4N) (**A'**) residues of the OPS (Figure 2A). The protons at positions 4, 5 and 7 of **N'** exhibited clear STD NMR enhancements, differently from the reducing sialic acid unit **N**. Additionally, a slight STD NMR enhancement relative to **N3$_{eq}$'** was further indicative of the protein preference for the internal sialic acid residue. Moreover, while the intensity of STD signals belonging to the acetyl groups of **N**, **A** and **A'** residues was quite low, a strong STD response was detected for the acetyl group of **N'**. Protons at position 6 and 9 of **N'** also received a good magnetization transfer, further confirming that the **N'** unit strongly contributed to the interaction with Siglec-7. Regarding the AAT, we observed a strong STD NMR

signal only for the anomeric proton of **A'** (the internal AAT), whereas **A** residue gave no STD signals. Similarly, proton signals at position 2 and 6 of **A'** exhibited medium STD effects, while the remaining protons of **A'** gave STD signals below 20 %. Therefore, the above confirmed the involvement of the internal AAT in the binding event. Remarkably, the galactose units were not involved in the binding process and revealed *no STD NMR signals* (Figure 3A), thus were oriented further from the protein binding pocket. Therefore, the *Fn*10953 2-Mer preferentially accommodated the internal sialic acid and AAT residues into Siglec-7 binding site.

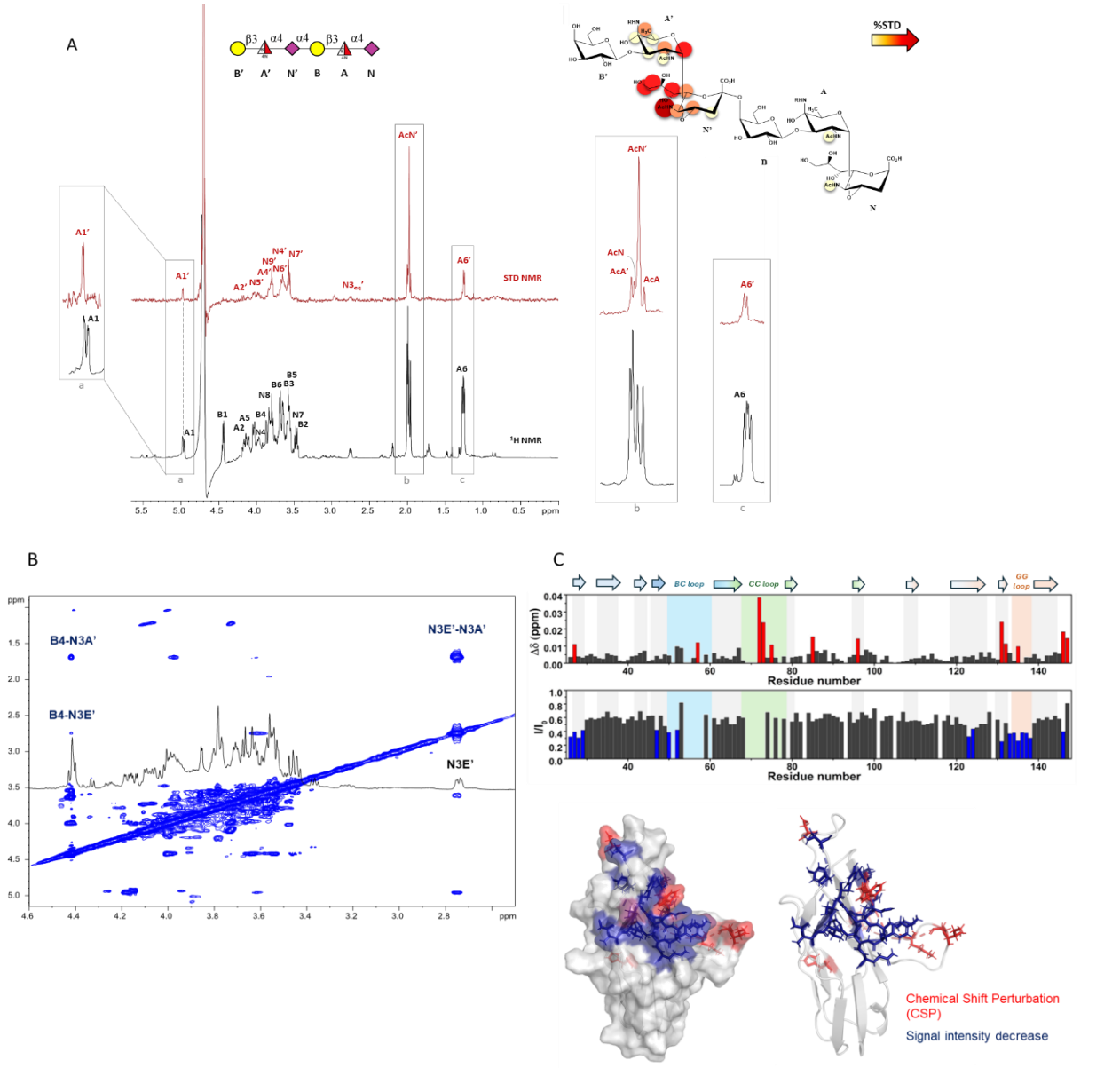


Figure 3. *NMR studies on the interaction between 2RU and Siglec-7.* ***A)*** *Schematic structure of the 2RU ligand depicted following the Symbol Nomenclature for Glycans. STD NMR spectrum composed by* ^{1}*H NMR reference spectrum (bottom) and 1D STD NMR*

spectrum (top), of the Siglec-7:2RU mixture at a 1:50 ratio. 2D representation illustrating the interacting epitope map derived from STD NMR data, showcasing the interactions between 2RU and Siglec-7. ***B)*** *Spectrum of the tr-NOESY NMR study conducted on the Siglec-7:2RU mixture at a 1:50 ratio. The key NOE contacts shown in the spectrum were indicative of extended bioactive conformation of 2RU.* ***C)*** *Diagrams of the CSP (top) and decrease in signal intensity (bottom) of the amino acids of 200 μM Siglec-7 in the presence of 4 mM of 3RU from Fn-10953. The CSP effect were evaluated with the formula* $\Delta\delta = \frac{1}{2}\sqrt{{\Delta\delta_H}^2 + (\Delta\delta_N/5)^2}$. *The residues experiencing the largest CSP (S27, D57, I72, S73, K75, W85, H96, K131, W132, K135, L146 and T147) have been highlighted in red; the residues experiencing the largest decrease in signal intensity (Y26, S27, L28, T29, S47, Y50, V52, F123, R124, K131, N133, Y134, K135, Y136, D137, Q138 and L146) have been highlighted in blue. On the right the 3D surface of Siglec-7 with the amino acids colored according to the CSP (red) and decrease in signal intensity (blue) effects.*

To gain further insights into the binding process, tr-NOESY NMR experiments were carried out on the 2-Mer. Additionally, computational studies were conducted to investigate the binding mode and propose a 3D model of the complex between Siglec-7 and the 2-Mer. Comparing NOESY (Figure S7) and tr-NOESY spectra (Figure 3B), no differences neither in the NOE contacts nor in the interproton distances were observed, indicating that the hexasaccharide containing two repeating units underwent no substantial conformational changes upon binding. Notably, the presence of key NOEs, as the contact of **B4-N'3** (axial and equatorial) and **B1-A2** (Figure 3B and S7), as well as the absence of the NOE between **N'3** with **A5'** and **A6'** identified a ligand conformational preference toward the *t* conformation (φ =180°). To depict a 3D complex, the uncommon AAT unit was parametrized using AMBER18 package [29] and then, the 2-Mer was built according to the energetic minima calculated through the adiabatic maps by molecular mechanics calculation (see details in supporting information, Figure S8A) and subjected to MD simulation (Figure S8B). The ligand conformation in the free state mainly changed with respect to the φ (C1-C2-O-C4') torsion angle around Neu5Ac-α-(2,4)-Gal glycosidic linkage (Figure S8B). Indeed, while ψ (C1-O-C4'-H4') torsion angle always assumed a value around 0°, φ dihedral angle could populate two different values, corresponding to 180° (conformer *t*) and -60° (conformer *g*), which lead to an extended and folded shape, respectively (Figure S8C). The MD simulations (Figure 3) agreed with NOE-based results, since the cluster analysis in the free state showed that the ligand remained in *t* conformation for the 70 % of time, thus indicating that 2-Mer preferentially adopted a more extended conformation in the free state. Similarly, and according to tr-NOESY experiments, the 2-Mer adopted an extended conformation upon binding. An almost comparable binding epitope and bioactive conformation was detected for 3-Mer, resulting in a similar accommodation and recognition from Siglec-7 (Figure S9).

Protein-based NMR experiments were also employed to evaluate the interaction from the protein viewpoint (Figure 3C). The NMR titration was performed between [U-^{15}N] Siglec-7 and increasing

amounts of 3-Mer that allowed to define the binding pocket of the protein following chemical shift perturbation (CSP) and/or signal intensity of the cross-peaks in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum. The addition of the ligand mainly induced a decrease of signal intensity, and the residues experiencing the largest intensity decrease were Tyr26, Ser27, Leu28, Thr29, Ser47, F123, R124, Lys131 and Asn133 as well as Tyr50 and Val52 of BC loop and the amino acids Tyr134, Lys135, Tyr136, Asp137 and Gln138 of GG' loop. Otherwise, Ile72, Ser73 and Lys75 of CC loop were affected by CSP, together with Ser27, Trp85, His96, Lys131, Trp132, Leu146, Thr147 and Asp57 of BC loop as well as Lys135 of GG' loop. These results indicated that the key Arg124 and the neighbouring amino acids from Lys131 to Gln138 formed the main binding site of Siglec-7, with residues from BC and CC' loops partially involved in the binding with 3RU.

3D complex

The extended, bioactive conformation of 2-Mer was modelled in the V-set Siglec-7 binding site (crystal structure pdb 2HRL [30]) and once optimized with Maestro Schrödinger, the complex was subjected to MD simulations (Figure 4) [31]. The results demonstrated the stability of the complex (Figure 4D) and confirmed that 2-Mer preferentially accommodated in the bound state with an extended conformation. MD results indicated that 3D complex of Siglec-7 and *Fn*10953 2-Mer exhibited significant contacts primarily mediated by residues **N'** and **A'**. These findings closely aligned with the experimental results (Figure 4). Particularly, the carboxylate moiety of **N'** played a pivotal role by establishing a stable salt bridge with the crucial Arg124, which emerged as the most robust interaction throughout the MD simulation (Figure 4) anchoring and orienting the glycan into the binding site. Further interactions were engaged by **N'** with Asn133 and Lys131 residues, both located in the GG' loop, through the hydroxyl group at position 8 of the glycerol chain and the NAc moiety, respectively. Additionally, the hydroxyl group at position 9 interacted with both Asn133 and Lys135 residues. On the other hand, Lys131 simultaneously stabilized **A'** residue through interacting with its acetyl group. Notably, the interaction between Siglec-7 and the AAT residue (**A'**) further involved Asn129, which formed polar contacts with the positively charged amino group at position 4. Furthermore, Asn70 and Trp74 of the CC' loop established transient hydrogen bonds interactions with the NAc moiety of the reducing **N** residue. Overall, computational studies combined with STD and tr-NOESY NMR results, indicated that the 2-Mer

adopted an extended conformation, with residues **N'** and **A'** playing crucial roles in the interaction with Siglec-7, engaged in polar, charged and hydrophobic interactions.

With the aim to propose a 3D complex of Siglec-7 and 3-Mer, the nonasaccharide was modelled into the receptor binding pocket, considering that, according to the experimental results, the reducing sialic acid unit (**Sia-1**) did not interact with the protein and therefore was oriented far from the protein surface. Conversely, both Sias units (**Sia-2** and **Sia-3**) could interact with the residue of Arg124. On the one hand, the engagement of **Sia-3** in the interaction with of Arg124 resulted in the elongation of the ligand towards the CC' loop. On the other hand, the accommodation of 3-Mer **Sia-2** interacting with Arg124, extended the saccharide chain towards the BC loop rather than only the CC' loop. This maintained the potential interactions with the CC' loop while also adding the possibility of interacting with the BC loop. In addition, MD simulations were performed by using different ligand conformations (both *g* and *t* conformations) and orientation in the binding site. The results revealed that, as expected, the ligand preferentially adopted the *t* conformation around Sia-Gal linkage, with all the other torsion angles adopting values in accordance with the exo-syn anomeric conformation, and that the most stable interactions throughout the dynamics involved the internal trisaccharide epitope (Figure 5A-E). Indeed, the carboxylate group of **Sia-2** acted as an anchor in the interaction with Arg124, making this the most stable interaction along the MD simulation. In addition to Arg124, Asn133 and Lys131 stabilized the hydroxyl group at the position 8 and the NAc group of **Sia-2**, respectively. Furthermore, **AAT-2** established hydrogen bonds with Lys131 and Asn129 which interacted respectively with the acetyl and positively charged amino groups, as already seen in Siglec-7-2-Mer complex. Moreover, in Siglec7-3RU complex, the ammonium ion at N4 of **AAT-2** also interacted with Asp53, part of the BC loop that acted as hydrogen bond acceptor. Moreover, the analysis of RMSF (Root Mean Square Fluctuation) during the molecular dynamics revealed a high degree of flexibility of BC and CC' loops also in the bound state, unlike the GG' loop that exhibit a decrease in flexibility (Figure 5F). This could be a consequence of the GG' loop residues being close to the binding pocket cavity, where the interaction with the internal trisaccharide epitope occurs. Thus, while the internal repeating unit remained stable along the MD, creating polar interactions with Arg124, Asn129, Lys131 and Asn133, the BC and CC' loop act as wings allowing the establishment of polar and hydrophobic interactions with the terminal and reducing trisaccharide respectively (Figure 5 and S9).

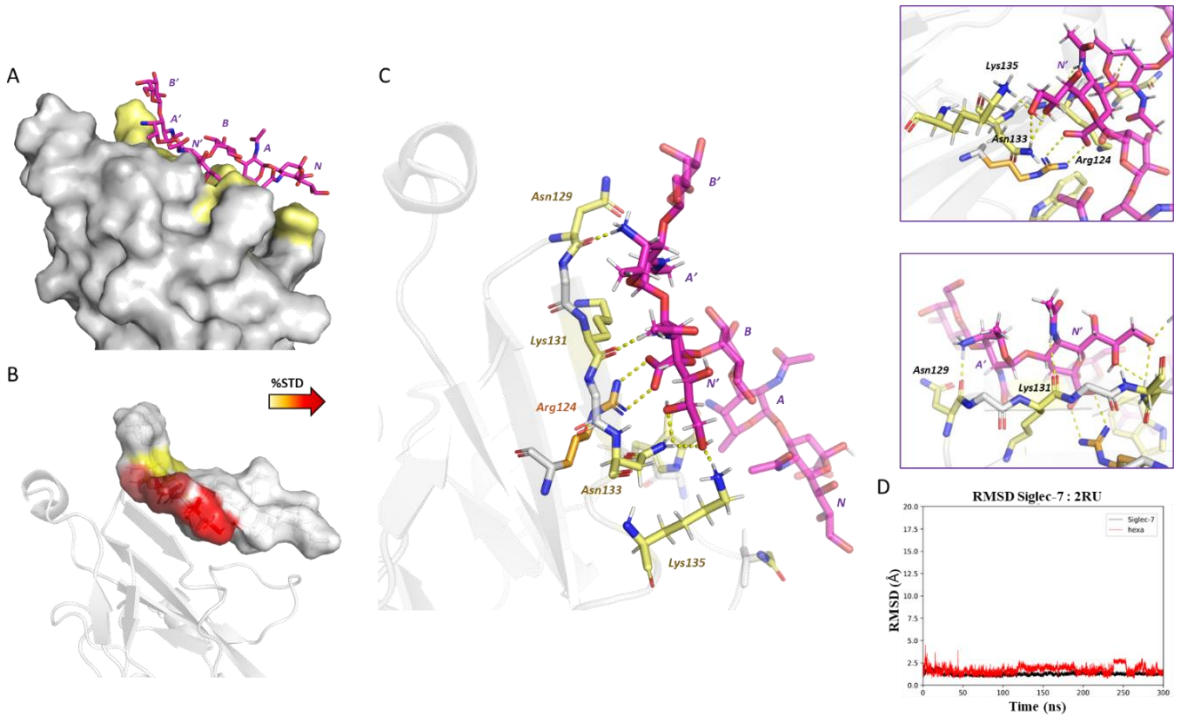


Figure 4. *Analysis of MD Simulations for the Siglec-7:2RU Complex:* ***A)*** *A 3D model of the Siglec-7:2RU complex, showing the protein surface with the interacting amino acids highlighted in yellow.* ***B)*** *3D depiction of the interaction, with the ligand surface coloured based on STD NMR and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

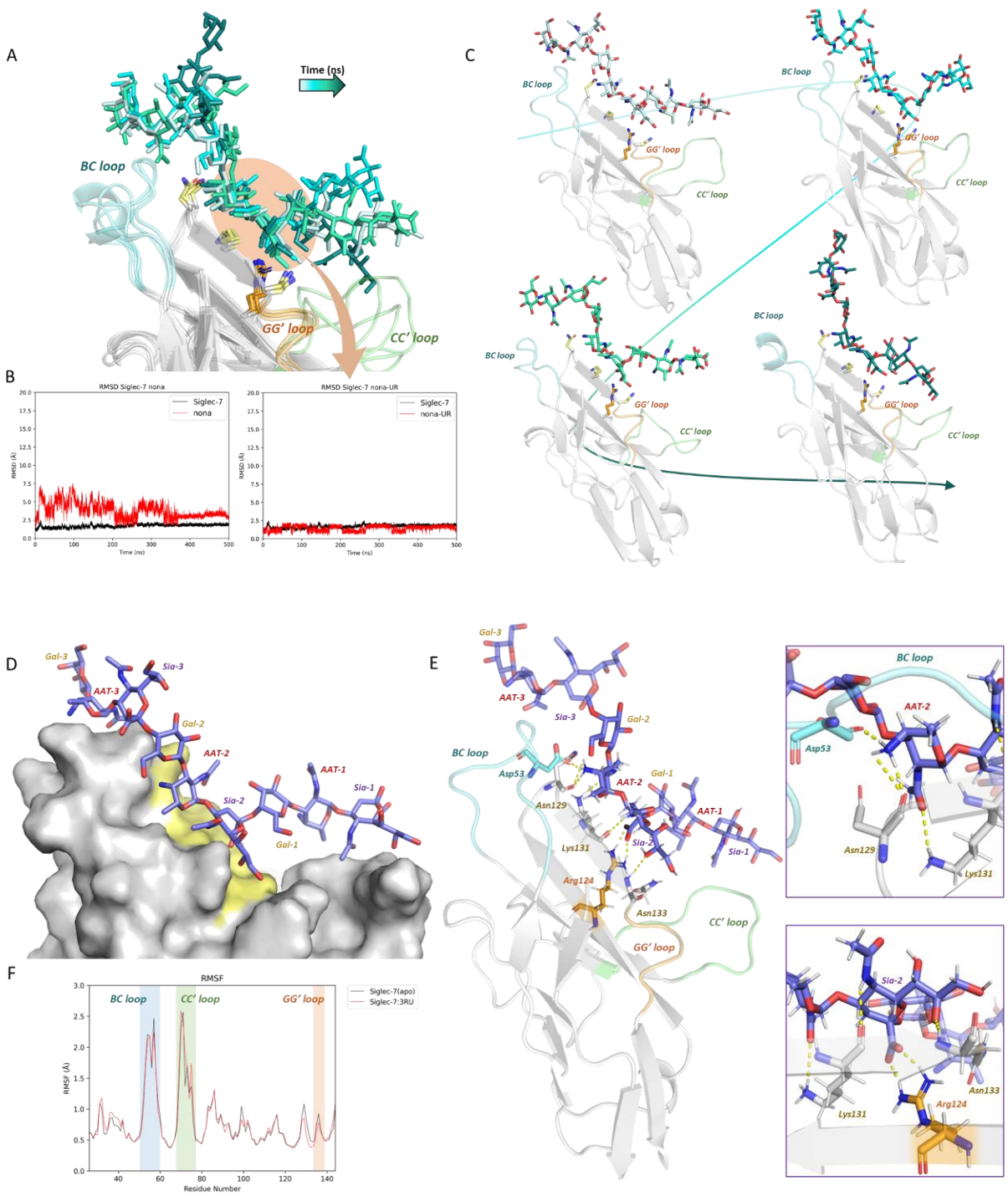


Figure 5. *Dynamics and stability of the Siglec-7:3RU complex across MD simulations.* ***A)*** *3D visualization of the Siglec-7:3RU complex through the overlay of various representative poses (from light to dark blue) to depict the mobility of the complex during MD, with the 3RU ligand shaded in varying tones of blue over time. The BC, CC', and GG' loops are highlighted in cyan, green, and orange, respectively. Amino acids involved in 3RU recognition are marked in yellow, with the key* ***Arg124*** *in orange.* ***B)*** *Left: RMSD analysis of the Siglec-7:3RU system; Right: RMSD of the inner trisaccharide (Gal2-AAT2-Sia2), illustrating the internal trisaccharide's stability during MD simulations, in contrast to the 3RU's variability due to the movements of terminal and reducing*

trisaccharides. ***C)*** *Temporal representation showcasing simultaneous movements of both terminal and reducing trisaccharides along with the BC and CC' loops, highlighting the anchored internal trisaccharide amidst 'wing-like' movements.* ***D)*** *A 3D model of the Siglec-7:3RU complex shows the protein surface with interacting amino acids in yellow.* ***E)*** *3D view of a representative pose from the most common MD simulation cluster, with the amino acids involved in the binding colored in grey, Arg124 in orange, and Asp53 from the BC loop in cyan. BC, CC', and GG' loops are again depicted in cyan, green, and orange, respectively, with dashed yellow lines indicating hydrogen bond interactions. A closer inspection reveals the interactions between Siglec-7 and the sialic acid (****Sia-2****, bottom) and AAT (****AAT-2****, top).* ***F)*** *RMSF results showing protein flexibility in free (black) and bound (red) state. A slight decreasement of the flexibility is observed only for the GG' loop.*

Discussion

Fusobacterium nucleatum is an oral oncobiont mostly associated with the development of periodontitis, and highly in various diseases, including gastrointestinal disorders and colorectal cancer. Recent research has demonstrated its involvement in tumour growth, metastasis, and modulation of immune responses. Current evidence suggests that tumour localization by *Fn* is dictated by glycan-lectin interactions; FadA and Fap2 adhesins[32] have been widely accepted to contribute to the pathogenicity of *Fn,* once it enters the bloodstream during transient bacteremia, frequent in periodontal disease. Surface exposed Fap2 lectin recognizes Gal-GalNAc epitope (the Thomsen–Friedenreich antigen), abundantly overexpressed in colorectal and breast cancer cells [4]. In addition, we have already shown the interaction between *Fn* strains and Siglec-7 on immune cells, although the molecular basis behind this recognition process is not fully explored. In this frame, shedding light on the recognition mechanisms underlying CRC tumorigenesis and understanding Siglec-7 *Fn*10953 interactions may provide insights into the development of targeted therapies for colorectal cancer and other related diseases. Thus, we here investigated the interaction between *Fn* ATCC 10953 and Siglec-7 through the application of multidisciplinary approached that include fluorescence, NMR, computational studies, as well as biological assays. Therefore, the LPS was extracted and oligomers containing increasing number of repeating units were isolated and evaluated in their interaction with Siglec-7.

Our study demonstrates that the *Fn10953* LPS, along with its isolated O-antigen, specifically binds to Siglec-7 on RPMI8226 cells, a multiple myeloma model. *Salmonella* LPS, used as a negative control, did not show binding, confirming specificity.

Siglec-7 is highly expressed on NK cells among blood cells, aligning with its known role in modulating NK cell functions. When PBMCs from healthy donors were exposed to *Fn10953* LPS

and O-antigen, only the O-antigen significantly increased HLA-DR expression in NK cells. T-lymphocytes and B-lymphocytes did not show significant activation.

Purified NK cells exhibited significant activation responses to both *Fn*10953 LPS and O-antigen, unlike T-lymphocytes and B-lymphocytes. This suggests a direct influence on NK cell activation through Siglec-7 binding. *Salmonella* LPS also activated NK cells but to a lesser degree, indicating a broader response pathway for *Salmonella* LPS. The higher activation induced by *Fn*10953 O-antigen in NK cells suggests a potent immunostimulatory role mediated by specific Siglec-7 engagement, making *Fn* a microorganism that could impact the tumour microenvironment.

Our data revealed that the core region of *Fn*10953 LPS did not interacted with Siglec-7; indeed, the equilibrium affinity constants measured for the full O-antigen portion and for the oligosaccharides containing various number of repeating units were comparable, all in the micromolar range, according to the protein-based NMR experiments that showed numerous intensity decreases of the most perturbed chemical shifts (Figures 3C and 6). In addition, a single repeating unit (1-Mer), including a reducing sialic acid, was not interacting with the protein, suggesting the necessity of a longer glycan chain to induce recognition process. Therefore, the key contributors to the receptor binding were the internal sialic acid **N'** and the AAT residue **A'**, while the galactose unit did not significantly interact with the receptor. Overall, these data indicated that at least two repeating units were required for stable and long-lasting interactions with Siglec-7. Computational studies of the 2-Mer alone and interacting with Siglec-7 provided a model of the 3D complex (Figure 4). The 2-Mer preferentially adopted an extended shape in the free state that is also maintained in its bioactive conformation (Figures 3B and S7). Extensive MD simulations confirmed the ligand's preference for the *t* conformation around the Sia glycosidic linkage (φ=180°). The internal Sia residue **N'** was accommodated in the binding pocket establishing the crucial salt bridge between its carboxylate group and the key Arg124 residue, as well as hydrogen bonds with its NHAc and the glycerol chain and Lys131, Asn133 and Lys135 residues, respectively. Additionally, significant interactions were observed between the positively charged amino group at N4 of AAT **A'** and the carbonyl group of Asn129 backbone. The above lead to the overall stabilization of the protein GG' and CC' loops upon binding. The ligand containing three repeating units (3-Mer) likely showed a preference for **Sia-2** to interact with Arg124 over the unstable reducing **Sia-1** and the **Sia-3**. The internal **Sia-2** and **AAT-2** residues emerged as the key actors of the recognition, causing their presence a chemical shift perturbation and an intensity

decrease of the amino acids near these sugars as evidenced by protein-based NMR experiments and corroborated by their stability in the RMSD analysis. Indeed, protein-based NMR experiments showed that the main protein residues affected upon ligand binding corresponded to those monitored along the MD simulation. Among these, Arg124 and the contiguous/adjacent amino acids comprising Lys131, Asn133, Lys135, were affected by an intensity decrease of their chemical shifts and were also the main residues in contact with **Sia-2** and **AAT-2** of the 3-Mer oligosaccharide. Notably, Lys135, together with Tyr134, Tyr136, Asp137 and Glu138 were all affected by intensity decrease upon binding and belonged to the GG' loop, which was slightly stabilized in the bound state, as monitored by the RMSF (Figure 5F). Furthermore, despite of the computational viewpoint in which a "wing-like" movement of the BC and CC' loops alternately approaches the terminal and reducing portion of the O-antigen ligand, the BC loop was predicted to establish more stable interactions compared to the CC' loop that, instead, is involved in more transient interactions. Indeed, from protein-based NMR results, BC loop contains Tyr50 and Val52 affected by intensity decrease and Asp57 affected by chemical shift perturbation (CSP); instead, CC' loop is only affected by CSP effect detected for Ile72 and Lys76, thus involved in a faster exchange with the ligand. In this context, the comparison between GT1b and the 2-Mer and 3-Mer ligands (Figure 7), reveals that while the GT1b ligand interacts more extensively with the GG' and CC' loops due to its terminal sialic acid, the 2-Mer and 3-Mer, which feature an internal sialic acid, occupy different spatial regions within the binding site. This leads to a distinct distribution of interactions, particularly involving the BC loop. These differences in spatial distribution suggest that the internal sialic acid of the 2-Mer and 3-Mer induces a different engagement of the binding site compared to GT1b. This alternative engagement of BC and CC' loops by the ligand opens avenues for understanding the mechanistic underpinnings of Siglec-7 engagement by *Fn* envelope glycans and therefore, the role of bacterial glycome in immune modulation and cancer progression. In conclusion, our comprehensive investigation into the interaction between *Fn* ATCC 10953 and Siglec-7 highlighted the essential role of the internal sialic acid residue, in forming stable complexes through a key salt bridge with Arg124, from the 2-Mer to more elongated structures that simulate the OPS. These findings not only advance our understanding of the intricate mechanisms governing immune cell recognition and bacterial evasion but also underscore the potential of targeting Siglec-7 for therapeutic strategies against colorectal cancer and other related diseases.

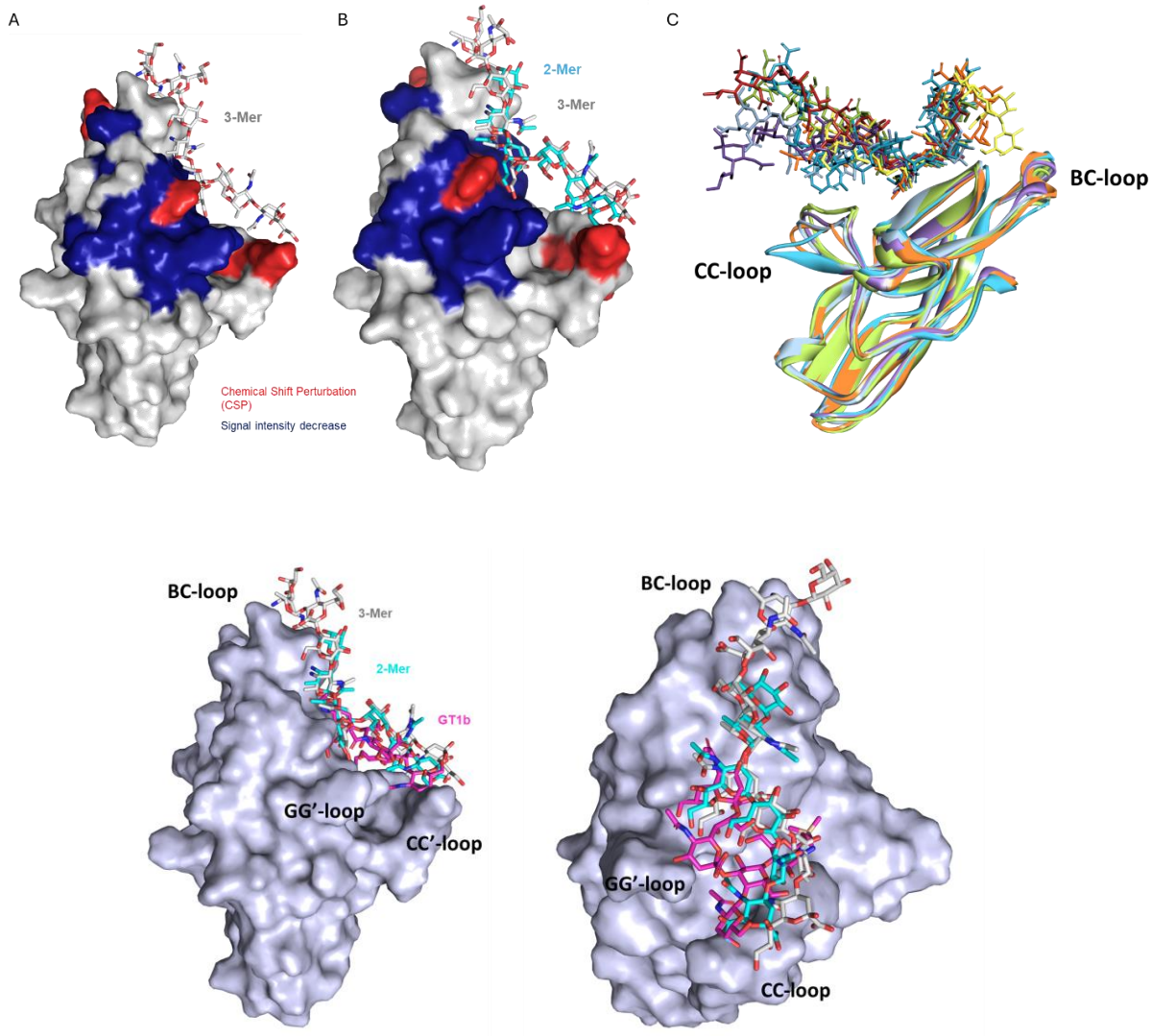


Figure 8. *3D views of Siglec-7 in complex with 2-Mer and 3-Mer. A) Best 3D model of Siglec-7 which amino acids are colored according to protein-based NMR titration with 3-Mer. B) Superimposition of the complexes of Siglec-7 in interaction with 2-Mer (cyan) and 3-Mer (grey); the amino acids of Siglec-7 are colored according to protein-based NMR titration. C) Superimposition of the different poses of Siglec-7 in interaction with 3-Mer obtained by MD simulation.*

Acknowledgments

AS, MTM, MF, LPC, GRG acknowledge H2020-MSCA-ITN-2020 (contract n° 956758). AS acknowledges Ministry of Education, Universities and Research, PRIN MUR 2022 (2022ZEZS45) and PRIN MUR PNRR 2022 (P2022M457Z). F.N-F. acknowledge Italian MUR, PRIN 2020, Project SEA-WAVE 2020BKK3W9. AS, RM acknowledge PNRR, Missione 4 – Componente 2 – NextGenerationEU - Partenariato Esteso INF-ACT - One Health Basic and Translational Research Actions Addressing Unmet Needs on Emerging Infectious Diseases MUR: PE00000007. The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy centre. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 re-search and innovation program under grant agreement No 851356 to R.M.

Methods

Healthy subjects and cell line Healthy subjects in this study were enrolled according to the protocol approved by IRCCS Pascale, Institutional Ethical Review Board (CE: Protocollo n. 4/21, 2021). RPMI 8226 cell line were obtained by IRCCS SYNLAB SDN Biobank (10.5334/ojb.26) and cultured in RPMI supplemented with 10 % fetal bovine serum and 1 % Glutamine. For the interaction study, $3x10^4$ cells/well were seeded in a 96-well plate and incubated at 37 °C and 5 % CO_2. After O/N incubation different concentration of LPS from *Salmonella*, *Fusobacterium nucleatum* and OPS from *Fusobacterium nucleatum* were added (from 0.39 µg/mL to 200 µg/mL) for 1 h at 37 °C. After two PBS washing steps, cells were labelled with PE-conjugated anti-Siglec-7 antibody (#12-5759-42, Invitrogen). After appropriate wash, a minimum of 10 000 events were recorded on a Cytoflex cytofluorimeter (Beckman coulter). For the interaction analysis, the normalized (respect to the untreated sample) mean of anti-Siglec-7 antibody were reported with respect to the concentrations of LPS and OPS. Experiment was repeated three times with similar results. The kDs were calculated using GraphPad 6 software (One-site Total equation). For the blood cells separation, T-Lymphocytes, B-Lymphocytes and NK-cells were purified from 5 healthy subjects using Easy SepTM kit (#17961, #19044, #17995, STEMCELL). For Cytofluorimetric analyses, PBMC were labelled using CD45-KO, CD3-APC700, CD14-PC5.5, CD19-PC7, CD56-APC750, CD45-56-19-3 tetrachrome and HLADR-PC5 antibodies (Beckman coulter). Kaluza software (Beckman Coulter) was used for the determination of the percentage of gated cells according to the gating strategy.

Cytokines evaluation For the cytokines expression evaluation, 200 ng/mL of LPS from *Salmonella*, *Fusobacterium nucleatum*, and OPS from *Fusobacterium nucleatum* were added to $5x10^5$ PBMC from 3 healthy donors (male, median age 32 years) for O/N incubation at 37 °C. Cells were harvested and cytokine expression levels were evaluated in the medium using LEGENDplex™ Human Inflammation Panel (13-plex), according to the manufacturer instruction (740118, Biolegend), using Cytoflex (Beckman Coulter).

Protein-based NMR experiments Uniformly ^{15}N-labeled Siglec-7 was resuspended in buffer containing 10 mM Na_2HPO_4 at pH 7.4, 2.7 mM KCl, 137 mM NaCl and 10 mM NaN_3, with a final

concentration of 100 μM. The buffer also included 0.05 mM [D4](trimethylsilyl) propionic acid sodium salt (TPS) as an internal standard for spectra calibration. ^{1}H-^{15}N HSQC experiments were conducted at 298 K. The ^{15}N-Siglec-7 was titrated with increasing equivalents of ligands until full saturation was achieved. The experiment was acquired with 32 scans were recorded, with 256 (t1) × 2048 (t2) complex data points in ^{15}N and ^{1}H, respectively. Data analysis was performed using CARA softare,[33] and the average chemical shift changes were calculated by using the following equation [34]. [34]

$$\delta_{average} = \sqrt{\frac{1}{2}[\Delta\delta_H^2 + (0{,}14\Delta\delta_N^2)^2]}$$

Ligand-based NMR experiments. All NMR experiments were recorded on a Bruker AVANCE NEO 600 MHz spectrometer equipped with a cryo-probe. Data acquisition and processing were performed using TOPSPIN 4.1.1 software. Samples were prepared in phosphate-buffered saline (10 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl, 10 mM NaN_3, pH 7.4) at 298 K, using D4 propionic acid sodium salt (TSP, 0.05 mM) as internal reference. The NMR experiments were conducted with a protein ratio of 1:50. Saturation Transfer Difference (STD) NMR Experiments: STD NMR spectra were acquired using 32k data points, zero-filled to 64k before processing. Protein resonances were selectively irradiated with 40 Gauss pulses lasting 50 ms, with the off-resonance pulse frequency set at 40 ppm and the on-resonance pulses at 0 ppm. To suppress protein signals, a 20 ms spinlock pulse was applied. The acquisition parameters included 65k data points and 112 scans. The STD NMR experiments were performed at a saturation time of 2 seconds. Ligand epitope mapping was achieved by calculating the ratio $(I_0 - I_{sat})/I_0$, where I_{sat} corresponds to the intensity of the STD NMR signal and I_0 to the intensity of the off-resonance spectrum. The strongest STD response was normalized to 100 %, and all other proton STD signals were scaled accordingly. Control experiments were also performed in the absence of the protein to ensure the specificity of the observed STD signals. Transferred NOESY (tr-NOESY) Experiments: Transferred NOESY spectra were recorded with data sets of 2048 x 512 points, using a mixing time of 0.3 seconds. Homonuclear 2D NOESY experiments were performed to analyse the interproton distances within the ligand, with mixing times ranging from 100 to 400 ms. The cross-relaxation rates (σij) were calculated by integrating the NOE cross peaks and normalising against

the corresponding diagonal peaks. Experimental interproton distances (rij) were then determined using the isolated spin pair approximation.

Parametrization The non-parametrised AAT unit was parametrised using a custom protocol with Gaussian 09 [35],[35] performing a Restrained Electrostatic Potential (RESP) charge calculation with a Hartree-Fock method and a 6-31G* basis set. The .prep and .frcmod files were generated by combining Antechamber and xLeap [36]. [36] Trajectory analysis was conducted using the ptraj module in AMBER 18 [29],[29] and the molecular dynamics MD results were visualised with the VMD program [37].[37]

Molecular mechanics calculations were performed using the MM3* force field available in the MacroModel 8.0 software from the Maestro package [38].[38] A dielectric constant of 80 was applied to simulate vacuum conditions, and the disaccharide structures were explored by incrementally varying the Φ and Ψ angles with an 18° grid step. Each (Φ, Ψ) point was optimised using 2000 conjugate gradient steps.

Molecular dynamics simulations were conducted with the AMBER 18 suite, employing specific force fields: ff14SB for the protein[39], Glycam06j-1[40] for the saccharide portion of the ligands and TIP3P for the water molecules. A glycam-adapted force field was prepared for the AAT unit. Proteins were prepared using the Maestro Protein Preparation Wizard, which added missing hydrogens, adjusted the protonation states of ionisable groups, and capped the termini. Systems were hydrated with a truncated-octahedral box of TIP3P water, with a 15 Å buffer, and counterions were added to neutralise the system. Input files were generated using the tleap module of AMBER 18. Initial energy minimisation was performed using the Sander module, while molecular dynamics simulations were carried out with the PMEMD module. Periodic boundary conditions and the particle mesh Ewald method were applied to represent electrostatic interactions, using a grid spacing of 1 Å. The initial minimisation was conducted with the complex fixed, followed by minimisation of the entire system. Subsequently, the system was gradually heated from 0 to 300 K, starting at constant volume and then transitioning to an isobaric ensemble. The temperature was maintained at 300 K for 50 ps with progressive energy minimisations and solute restraints. After the equilibration phase, restraints were removed, and the simulations proceeded in an isothermal-

isobaric ensemble during the production phase. System coordinates were saved every 2 ps, generating a set of 10,000 structures per complex for further analysis.

Trajectories were analysed using the ptraj module in AMBER 18, and the molecular dynamics results were visualised with the VMD program[37]. Cluster analysis based on ligand RMSD was performed using the K-mean algorithm in the ptraj module. The representative structure of the most populated cluster was used to illustrate the complex interactions. Hydrogen bond determination was carried out with the CPPTRAJ module [41], defining a hydrogen bond as having a maximum distance of 3 Å between donor and acceptor atoms and a minimum A-H-D angle of 135°. 3D images were prepared using PyMOL[42], and dihedral conformation analysis was performed using a custom script that generated histograms of the most populated values during the simulation.

References

1. Brennan, C. A.; Garrett, W. S., Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. *Nature Reviews Microbiology* 2019, *17* (3), 156-166.
2. Alon-Maimon, T.; Mandelboim, O.; Bachrach, G., Fusobacterium nucleatum and cancer. *Periodontology 2000* 2022, *89* (1), 166-180.
3. Han, Y. W., Fusobacterium nucleatum: a commensal-turned pathogen. *Current opinion in microbiology* 2015, *23*, 141-7.
4. Parhi, L.; Alon-Maimon, T.; Sol, A.; Nejman, D.; Shhadeh, A.; Fainsod-Levi, T.; Yajuk, O.; Isaacson, B.; Abed, J.; Maalouf, N.; Nissan, A.; Sandbank, J.; Yehuda-Shnaidman, E.; Ponath, F.; Vogel, J.; Mandelboim, O.; Granot, Z.; Straussman, R.; Bachrach, G., Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nature Communications* 2020, *11* (1), 3259.
5. Wu, S.; Rhee, K.-J.; Albesiano, E.; Rabizadeh, S.; Wu, X.; Yen, H.-R.; Huso, D. L.; Brancati, F. L.; Wick, E.; McAllister, F., A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nature medicine* 2009, *15* (9), 1016-1022.
6. Marchesi, J. R.; Dutilh, B. E.; Hall, N.; Peters, W. H. M.; Roelofs, R.; Boleij, A.; Tjalsma, H., Towards the Human Colorectal Cancer Microbiome. *PLOS ONE* 2011, *6* (5), e20447.
7. Sepich-Poore, G. D.; Zitvogel, L.; Straussman, R.; Hasty, J.; Wargo, J. A.; Knight, R., The microbiome and human cancer. *Science* 2021, *371* (6536).
8. Wang, S.; Liu, Y.; Li, J.; Zhao, L.; Yan, W.; Lin, B.; Guo, X.; Wei, Y., Fusobacterium nucleatum Acts as a Pro-carcinogenic Bacterium in Colorectal Cancer: From Association to Causality. *Front. Cell Dev. Biol.* 2021, *9*.
9. Zepeda-Rivera, M.; Minot, S. S.; Bouzek, H.; Wu, H.; Blanco-Míguez, A.; Manghi, P.; Jones, D. S.; LaCourse, K. D.; Wu, Y.; McMahon, E. F.; Park, S.-N.; Lim, Y. K.; Kempchinsky, A. G.; Willis, A. D.; Cotton, S. L.; Yost, S. C.; Sicinska, E.; Kook, J.-K.; Dewhirst, F. E.; Segata, N.; Bullman, S.; Johnston, C. D., A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature* 2024, *628* (8007), 424-432.
10. Casasanta, M. A.; Yoo, C. C.; Udayasuryan, B.; Sanders, B. E.; Umaña, A.; Zhang, Y.; Peng, H.; Duncan, A. J.; Wang, Y.; Li, L.; Verbridge, S. S.; Slade, D. J., Fusobacterium nucleatum host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives colorectal cancer cell migration. *Science signaling* 2020, *13* (641).

11. Wong, S. H.; Yu, J., Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol* 2019, *16* (11), 690-704.
12. Wu, J.; Li, Q.; Fu, X., Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Translational oncology* 2019, *12* (6), 846-851.
13. Stanczak, M. A.; Läubli, H., Siglec receptors as new immune checkpoints in cancer. *Molecular Aspects of Medicine* 2023, *90*, 101112.
14. Angata, T.; Varki, A., SiglecSiglecsInteractions with PathogensPathogens. In *Glycoscience: Biology and Medicine*, Endo, T.; Seeberger, P. H.; Hart, G. W.; Wong, C.-H.; Taniguchi, N., Eds. Springer Japan: Tokyo, 2021; pp 1-9.
15. Kim, C.-H., *Glycobiology of innate immunology*. Springer: 2022.
16. Hsieh, S.-L., *Lectin in host defense against microbial infections*. Springer: 2020; Vol. 1204.
17. Chen, Y.; Chen, H.; Zheng, Q., Siglecs family used by pathogens for immune escape may engaged in immune tolerance in pregnancy. *Journal of Reproductive Immunology* 2023, *159*, 104127.
18. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2022.
19. Pillai, S.; Netravali, I. A.; Cariappa, A.; Mattoo, H., Siglecs and Immune Regulation. *Annual review of immunology* 2012, *30* (1), 357-392.
20. Varki, A.; Angata, T., Siglecs-the major subfamily of I-type lectins. *Glycobiology* 2006, *16* (1), 1r-27r.
21. Fong, J. J.; Tsai, C.-M.; Saha, S.; Nizet, V.; Varki, A.; Bui, J. D., Siglec-7 engagement by GBS β-protein suppresses pyroptotic cell death of natural killer cells. *PNAS* 2018, *115* (41), 10410-10415.
22. Heikema, A. P.; Jacobs, B. C.; Horst-Kreft, D.; Huizinga, R.; Kuijf, M. L.; Endtz, H. P.; Samsom, J. N.; van Wamel, W. J., Siglec-7 specifically recognizes Campylobacter jejuni strains associated with oculomotor weakness in Guillain–Barré syndrome and Miller Fisher syndrome. *Clinical Microbiology Infection* 2013, *19* (2), E106-E112.
23. Lamprinaki, D.; Garcia-Vello, P.; Marchetti, R.; Hellmich, C.; McCord, K. A.; Bowles, K. M.; Macauley, M. S.; Silipo, A.; De Castro, C.; Crocker, P. R.; Juge, N., Siglec-7 Mediates Immunomodulation by Colorectal Cancer-Associated Fusobacterium nucleatum ssp. animalis. *Frontiers in Immunology* 2021, *12*.
24. De Castro, C.; Parrilli, M.; Holst, O.; Molinaro, A., Chapter Five - Microbe-Associated Molecular Patterns in Innate Immunity: Extraction and Chemical Analysis of Gram-Negative Bacterial

Lipopolysaccharides. In *Methods in enzymology*, Fukuda, M., Ed. Academic Press: 2010; Vol. 480, pp 89-115.

25. Vinogradov, E.; St. Michael, F.; Homma, K.; Sharma, A.; Cox, A. D., Structure of the LPS O-chain from Fusobacterium nucleatum strain 10953, containing sialic acid. *Carbohydr. Res.* 2017, *440-441*, 38-42.
26. Di Carluccio, C.; Forgione, R. E.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B. A.; Marchetti, R.; Silipo, A., Molecular recognition of sialoglycans by streptococcal Siglec-like adhesins: toward the shape of specific inhibitors. *RSC Chemical Biology* 2021, *2* (6), 1618-1630.
27. 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., Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22. *iScience* 2021, *24* (1), 101998.
28. Di Carluccio, C.; Forgione, M. C.; Martini, S.; Berti, F.; Molinaro, A.; Marchetti, R.; Silipo, A., Investigation of protein-ligand complexes by ligand-based NMR methods. *Carbohydr Res* 2021, *503*, 108313.
29. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden , T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; SalomonFerrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, M.; Kollman, P. A. *AMBER 2018*, University of California, San Francisco, 2018.
30. Attrill, H.; Imamura, A.; Sharma, R. S.; Kiso, M.; Crocker, P. R.; van Aalten, D. M. F., Siglec-7 Undergoes a Major Conformational Change When Complexed with the α(2,8)-Disialylganglioside GT1b*. *Journal of Biological Chemistry* 2006, *281* (43), 32774-32783.
31. Nieto-Fabregat, F.; Lenza, M. P.; Marseglia, A.; Di Carluccio, C.; Molinaro, A.; Silipo, A.; Marchetti, R., Computational toolbox for the analysis of protein–glycan interactions. *Beilstein Journal of Organic Chemistry* 2024, *20*, 2084-2107.
32. Engevik, M. A.; Danhof, H. A.; Ruan, W.; Engevik, A. C.; Chang-Graham, A. L.; Engevik, K. A.; Shi, Z.; Zhao, Y.; Brand, C. K.; Krystofiak, E. S.; Venable, S.; Liu, X.; Hirschi, K. D.; Hyser, J. M.; Spinler, J. K.; Britton, R. A.; Versalovic, J., Fusobacterium nucleatum Secretes Outer Membrane Vesicles and Promotes Intestinal Inflammation. *mBio* 2021, *12* (2).

33. Skinner, S. P.; Fogh, R. H.; Boucher, W.; Ragan, T. J.; Mureddu, L. G.; Vuister, G. W., CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *Journal of biomolecular NMR* 2016, *66* (2), 111-124.
34. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* 2013, *73*, 1-16.
35. Frisch, M. J., gaussian09. *http://www.gaussian.com/* 2009.
36. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. J. J. A. C. S., Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* 2001, *222* (1).
37. Humphrey, W.; Dalke, A.; Schulten, K., VMD - Visual Molecular Dynamics. *J. Molec. Graphics* 1996, *14*, 33-38.
38. *Schrödinger Release 2022-3*, Maestro, Schrödinger, LLC: New York, 2021.
39. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation* 2015, *11* (8), 3696-3713.
40. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* 2008, *29* (4), 622-55.
41. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 2013, *9* (7), 3084-3095.
42. *The PyMOL Molecular Graphics System*, Version 2.3.4 Schrödinger,LLC.
43. St Michael, F.; Fleming, P.; Cox, A. D.; Vinogradov, E., Structural analysis of the core oligosaccharides from Fusobacterium nucleatum lipopolysaccharides. *Carbohydrate Research* 2021, *499*, 108198.

Molecular insights into O-linked Sialoglycans Recognition by the Siglec-like SLBR-N ($SLBR_{UB10712}$) of *Streptococcus gordonii*

ACS central science

Open Access

http://pubs.acs.org/journal/acscii

Article

Molecular Insights into O-Linked Sialoglycans Recognition by the Siglec-Like SLBR-N ($SLBR_{UB10712}$) of *Streptococcus gordonii*

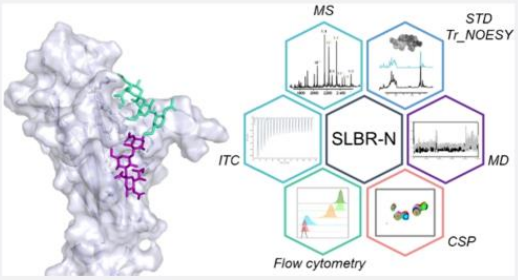
Cristina Di Carluccio, Linda Cerofolini, Miguel Moreira, Frédéric Rosu, Luis Padilla-Cortés, Giulia Roxana Gheorghita, Zhuojia Xu, Abhishek Santra, Hai Yu, Shinji Yokoyama, Taylor E. Gray, Chris D. St. Laurent, Yoshiyuki Manabe, Xi Chen, Koichi Fukase, Matthew S. Macauley, Antonio Molinaro, Tiehai Li, Barbara A. Bensing, Roberta Marchetti,* Valérie Gabelica, Marco Fragai, and Alba Silipo*

Cite This: https://doi.org/10.1021/acscentsci.3c01598

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: *Streptococcus gordonii* is a Gram-positive bacterial species that typically colonizes the human oral cavity, but can also cause local or systemic diseases. Serine-rich repeat (SRR) glycoproteins exposed on the *S. gordonii* bacterial surface bind to sialylated glycans on human salivary, plasma, and platelet glycoproteins, which may contribute to oral colonization as well as endocardial infections. Despite a conserved overall domain organization of SRR adhesins, the Siglec-like binding regions (SLBRs) are highly variable, affecting the recognition of a wide range of sialoglycans. SLBR-N from the SRR glycoprotein of *S. gordonii* UB10712 possesses the remarkable ability to recognize complex core 2 *O*-glycans. We here employed a multidisciplinary approach, including flow cytometry, native mass spectrometry, isothermal titration calorimetry, NMR spectroscopy from both protein and ligand perspectives, and computational methods, to investigate the ligand specificity and binding preferences of SLBR-N when interacting with mono- and disialylated core 2 *O*-glycans. We determined the means by which SLBR-N preferentially binds branched α2,3-disialylated core 2 *O*-glycans: a selected conformation of the 3′SLn branch is accommodated into the main binding site, driving the sTa branch to further interact with the protein. At the same time, SLBR-N assumes an open conformation of the CD loop of the glycan-binding pocket, allowing one to accommodate the entire complex core 2 *O*-glycan. These findings establish the basis for the generation of novel tools for the detection of specific complex *O*-glycan structures and pave the way for the design and development of potential therapeutics against streptococcal infections.



INTRODUCTION

The human oral cavity is densely populated with bacteria, some of which can be associated with a variety of systemic infections.[1] Bacteremia may result from lesions in the oral epithelium, occurring for instance after dental manipulation, and also following daily procedures, such as chewing, toothbrushing, and flossing. Once in the bloodstream, bacteria can reach and colonize other organs and consequently establish infection and inflammation. Streptococcal bloodstream infections are among the most frequent causes of infective endocarditis (IE), an inflammatory disease that impacts the endocardium and causes serious damage of the heart.[2] In particular, *Streptococcus sanguinis*, *Streptococcus gordonii*, and *Streptococcus mitis/oralis*, commensals of the oral microbiota, are common in patients affected by IE.[3] Despite the guidelines on prevention, diagnosis, and treatment,[4] IE incidence has remained unchanged over the past 30 years. Antibiotic administration, the treatment of choice, often fails due to antibiotic resistance, thus making urgent the development of novel measures and approaches for IE prevention and treatment.

S. sanguinis, *S. gordonii*, and *S. oralis* express surface sialic acid-binding adhesins, including serine-rich repeat glycoproteins (SRRPs)[5] and AsaA proteins[6] that mediate adherence to host sialylated glycoproteins and tissue colonization. Generally, the structural organization of SRRPs includes an N-terminal signal peptide, a ligand binding region (BR) positioned between two SRR regions (a short SRR1 and long SRR2), and a C-terminal

Received: December 21, 2023
Revised: January 23, 2024
Accepted: January 23, 2024

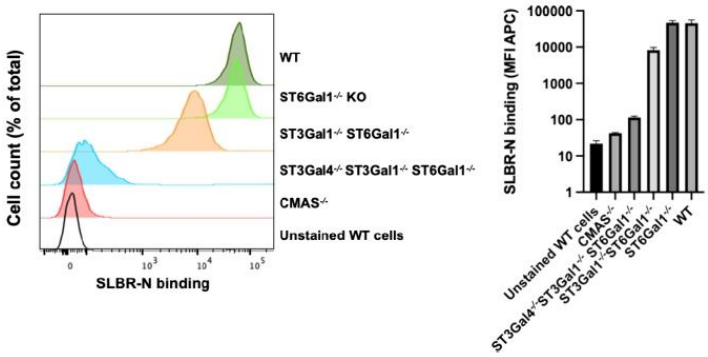


Figure 1. Flow cytometry analysis. SLBR-N binding to U937 WT (green), ST6Gal1 KO (lime green), ST3Gal1/ST6Gal1 KO (orange), ST3Gal1/4, ST6Gal1 KO (blue), CMAS KO (red), and unstained WT (black) cells. Left panel: Flow cytometry histograms. Right panel: Quantification of the mean fluorescence intensity (MFI) for three technical replicates.

LPXTG motif anchored to the cell wall (Scheme S1). The "Siglec-like" BRs (SLBRs) include two domains that are crucial for glycan recognition: the Siglec domain and the adjacent Unique domain, involved in the modulation of the Siglec domain conformation. The SLBRs mediate recognition of α2−3-linked sialic acid-galactose epitopes located at the termini of O-linked glycans displayed on mucins and mucin-like proteins. The O-glycosylated host proteins bound by the SLBRS include the salivary mucin MUC7[7,8] and platelet glycoprotein Ibα,[9,10] and these interactions may contribute to oral colonization and IE, respectively.

Streptococcus gordonii, an oral commensal sometimes associated with periodontitis, can act as an opportunistic pathogen and cause a variety of systemic diseases, including empyema in the lung, spondylodiscitis in the spine, perihepatic abscess in the liver, and endocarditis.[11]

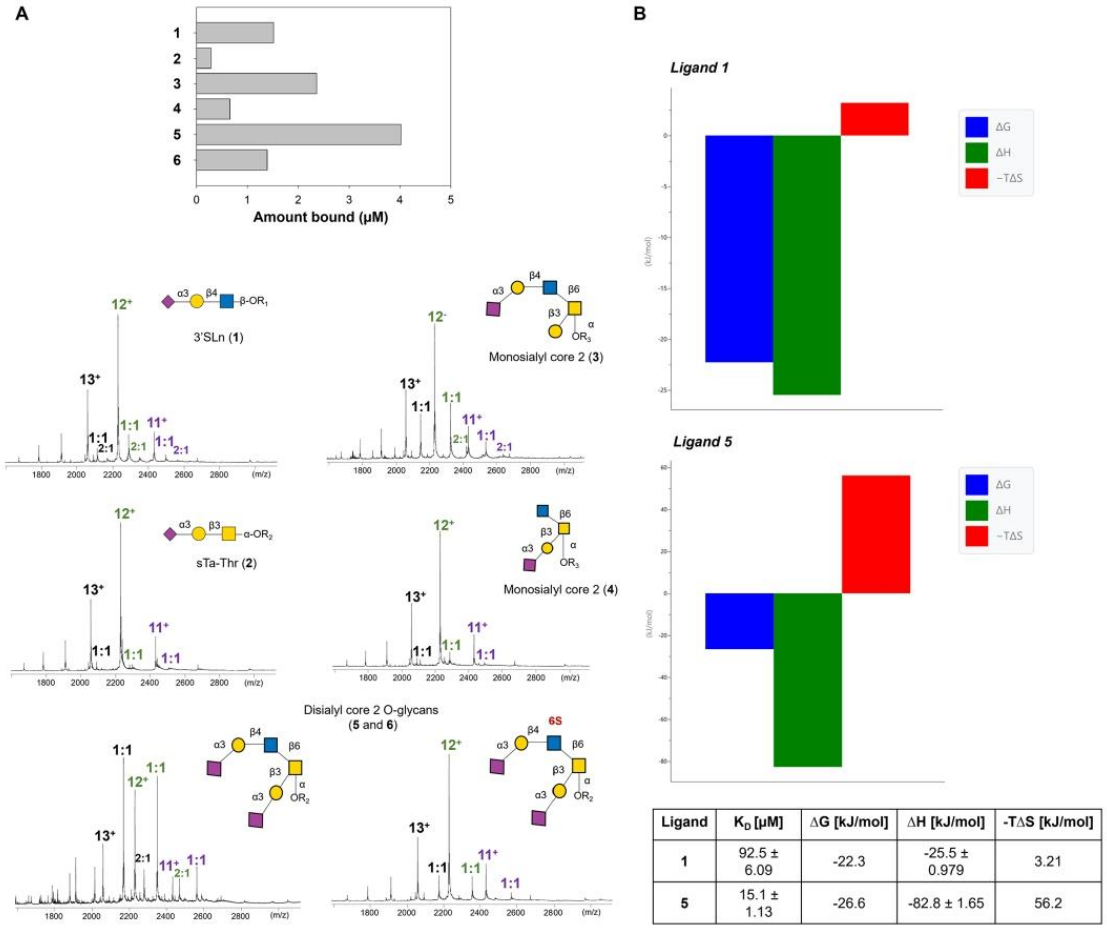
Despite the conserved structural fold of the Siglec domains and the crucial role of a ΦTRX motif in sialoglycan binding,[11,12] the primary structures of SLBRs are highly divergent and show different selectivity for α2−3 sialoglycans. Recent chimeragenesis experiments pointed out how the specificity of Siglec-like adhesins for different ligands may be determined by hypervariability in the CD, EF, and FG loops of the V-set Ig fold in the Siglec domain, influencing the interaction with host cells, and then pathogenicity and commensalism.[13] The diversity of ligands recognized is thought to parallel the range of *O*-glycan structures displayed on MUC7. Importantly, the SLBR ligand repertoire impacts the recognition of plasma and platelet glycoproteins, in addition to virulence and pathogenesis.[13,14] Indeed, loop alterations in SLBRs modify the recognition of glycoforms of MUC7, revealing the importance of SLBR selectivity in bacterial adhesion to host receptors and then their potential implications for bacterial infections. The SLBRs of SRR adhesins from *Streptococcus gordonii* M99 and Challis strains, SLBR-B and SLBR-H, respectively, have been extensively characterized,[12−17] and their contribution to pathogenicity was demonstrated in animal models of endocarditis. SLBR-B is selective for a core 1 trisaccharide known as sialyl T-antigen (sTa), whereas SLBR-H shows broader specificity and can bind sTa as well as the closely related structures 3′-sialyl-*N*-acetyllactosamine (3′SLn) and sialyl Lewis C (sLe^C). In contrast, SLBR-N of *S. gordonii* strain UB$_{10712}$ binds more avidly to different α2,3 sialoglycans, in particular to 3′-sialyl-*N*-acetyllactosamine (3′SLn) more avidly than to sialyl-T-antigen (sTa or) or sialyl Lewis C (sLe^C). Despite that SLBR-H and SLBR-N share 80% sequence identity, only the latter recognizes fucosylated and sulfated derivatives of 3′SLn, including sialyl Lewis X (sLe^X) and 6S-sLe^X.[6] Interestingly, SLBR-N also displayed the ability to bind larger, branched disialylated core 2 *O*-glycans, as demonstrated by probing glyco-engineered HEK293 cells and glycan arrays.[14,18] However, the means by which SLBR-N may preferentially bind disialylated core 2 glycans is unclear. Furthermore, the ability of SLBR-N to detect O-linked sialoglycans on CD44, a complex transmembrane glycoprotein common marker of cancer stem cells (CSCs) in breast cancer, makes SLBR-N particularly attractive not only for its biological role, but also as a tool to detect core 2 disialylated *O*-glycans on tumor cells.[20] In addition, SLBR-N has been recently shown to enhance the infectivity and transmissibility of HIV-1 by binding the *O*-glycans exposed on the surface of the envelope glycoprotein gp120 and increasing the affinity of HIV-1 for the CD4 receptor.[21]

To deeply understand how SLBRs can discriminate between closely related sialoglycan structures, and with the long-term goal of developing potential mimetics to hinder streptococcal infections, we undertook a comprehensive study of the SLBR-N recognition and binding to chemoenzymatically synthesized monosialylated and complex core 2 branched *O*-glycans.[19] We employed a combination of multidisciplinary and complementary methods, consisting of flow cytometry, high-resolution ligand- and protein-based NMR experiments, native mass spectrometry, isothermal titration calorimetry, and computational approaches, including docking studies, molecular dynamic simulations, and CORCEMA-ST (complete relaxation and conformational exchange matrix analysis of saturation transfer), to achieve information about binding preferences, affinities, and 3D molecular features of such protein−ligand complexes. Our results showed a preference of SLBR-N in interacting with glycans containing a sialylactosamine branch; moreover, the presence of two sialic acids, as in disialylated core 2 *O*-glycan structures, results in a wider protein binding site, thus strengthening the interaction.

RESULTS

SLBR-N Recognition of Host Glycans. To dissect the interaction of SLBR-N with host *O*-glycans, we analyzed the protein binding to host ligands. These include 3′-sialylactosamine (3′SLn, ligand **1**), sialyl-T-antigen (sTa-Thr, ligand **2**),

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX



Ligand	K_D [μM]	ΔG [kJ/mol]	ΔH [kJ/mol]	-TΔS [kJ/mol]
1	92.5 ± 6.09	-22.3	-25.5 ± 0.979	3.21
5	15.1 ± 1.13	-26.6	-82.8 ± 1.65	56.2

Figure 2. (A) Native mass spectrometry analysis. Electrospray mass spectra of protein SLBRN with ligands **1**−**6**. The 1:1 ligand:protein peaks were considered; lower intensity peaks corresponding to a 2:1 ligands:protein complex could be also observed for **1**, **4**, and **6**, likely due to nonspecific interactions. R1, ethanolamine; R2, threonine; and R3, methoxybenzene. The histogram derived from the MS-determined equilibrium association constants (see the Supporting Information) reflects the different affinity of SLBR-N for the different sialoglycans, revealing the following ligand preference: **5** ≫ **3** > **1**/**6** ≫ **4** > **2**. (B) Isothermal titration calorimetry analysis. Thermodynamic profiles for the interaction of **1** and **5** with SLBR-N, as measured by ITC experiments. Affinity constants and thermodynamic parameters of the interactions were shown in the table.

and monosialylated core 2 *O*-glycans (ligands **3** and **4**); we then focused on the more complex disialylated core 2 *O*-glycans (ligands **5** and **6**, Scheme S2), containing both 3′SLn and sTa-Thr branches.

Flow Cytometry Analysis. The ability of SLBR-N to preferentially recognize α2−3-linked sialic acid containing glycans has been confirmed by flow cytometry experiments as follows. We knocked out several key sialyltransferases from U937 cells and tested the binding of fluorescently labeled SLBR-N to cells by flow cytometry (Figure 1). Knocking out ST6Gal1, which installs α2−6-linked sialic acid onto *N*-glycans, had no impact on SLBR-N binding. Further deletion of ST3Gal1, an enzyme that installs α2−3-linked sialic acid onto the core 1 branch of *O*-glycans (Scheme 2), led to the 4-fold decrease in binding of SLBR-N to cells. Further deletion of ST3Gal4, which likely installs α2−3-linked sialic acid onto the core 2 branch of *O*-glycans, caused a much greater impact on SLBR-N binding, decreasing binding by another 100-fold. It has previously been shown that both ST3Gal1 and ST3Gal4 are implicated in making the ligand for SLBR-N.[5,18,20] This is supported by our findings that, however, point to ST3Gal4 as being the more critical enzyme in creating ligands recognized by SLBR-N on cells.

Native Mass Spectrometry and Binding Affinities. Native electrospray ionization mass spectrometry (ESI−MS) was used to obtain the binding affinities of the investigated glycans for SLBR-N. First, MS analysis on SLBR-N was performed using ion mobility MS.[21] The charge states 12^+ and 13^+ were compact (Figure S1A); hence, presumably the protein was well folded and produced via a charged residue process and used for the quantification. Lower charge states (11^+, 10^+...) were not sufficiently declustered for quantification, while higher charge states could be partially unfolded, presumably produced partially by chain ejection, and, for this reason, were excluded from the

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

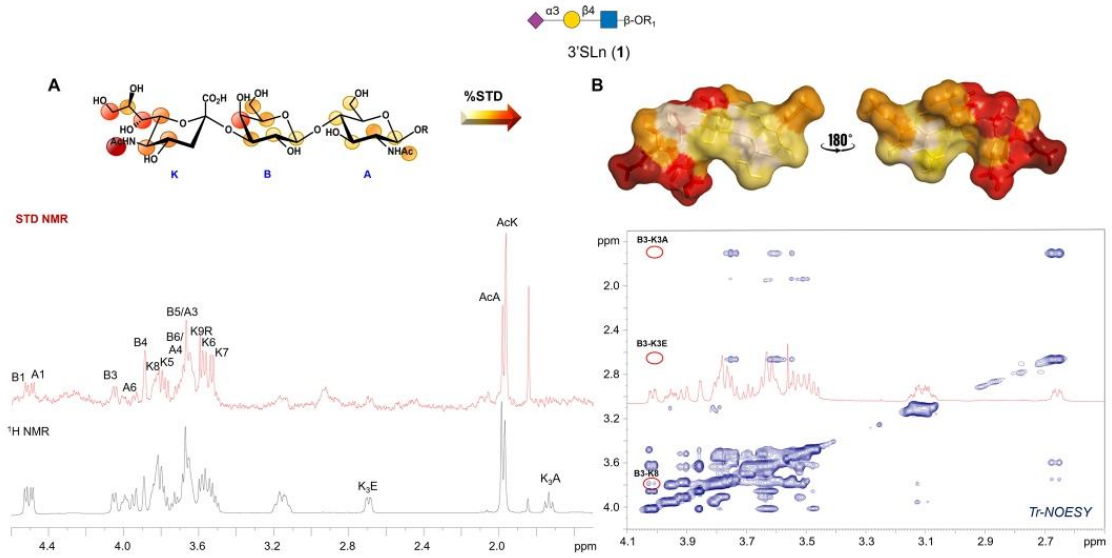


Figure 3. NMR results from the 3′SLn ligand perspective in interaction with SLBR-N. (A) STD NMR analysis of SLBR-N and ligand **1** with an indication of ligand epitope mapping calculated by $(I_0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ was the signal intensity in the STD-NMR spectrum (red) and I_0 was the peak intensity of the off-resonance spectrum (black). The highest STD signal referring to the Ac group of Neu5Ac was set to 100%, and the other protons were normalized to this value. (B) Bioactive conformation of ligand **1** as obtained by the tr-NOESY spectrum; the ligand surface was colored according to the STD effects.

quantification. Electrospray mass spectra then were acquired on the protein SLBR-N in complex with ligands **1**−**6** (Figure 2A). Stoichiometry and quantification of the complexes were then obtained from the position and the relative intensities of the peaks, respectively.[22−24] Thus, ESI−MS allowed one to determine the global equilibrium binding constants (sum of 1:1 complexes and of 2:1 complexes) of the different systems (Figures 2A and S1A). Ligand **5** had the highest affinity for SLBR-N, with the stronger intensity of the peak corresponding to the 1:1 complex, followed by ligands **3** and **1** that also showed good binding, with the latter comparable to **6**. Regarding ligands **2** and **4**, the intensity was very low even at a 1:1 ligand:protein ratio, diagnostic of a weak affinity of SLBR-N for these ligands ($K_D > 100\,\mu M$). The MS-derived affinity constants for these two ligands also serve to prove that the 1:1 and 2:1 complexes detected for the others are specific in the sense that they exist in solution and are not MS artifacts. However, as no NMR signal was found for a second binding site (see below), we conclude that the 2:1 complex detected by MS is not site-specific, but rather the sum of weakly interacting complexes over the protein surface.

Isothermal Titration Calorimetry. ITC was also performed to gain information regarding affinity constants (K_D) and thermodynamic parameters of the binding for the two main ligands (**1** and **5**) (Figures 2B and S1B). In both cases, the formation of the complex was spontaneous ($\Delta G < 0$), and the reaction was enthalpically driven, with ΔH more negative for ligand **5** with respect to **1** (Figure 2B and Supporting Information). The ITC results fitted well into a single-site binding model, and the derived K_D value agreed with those obtained by native MS (see below for further discussion on ITC results). Therefore, ESI−MS, ITC, and flow cytometry analyses provided information on the SLBR-N preferential recognition of 3-linked host sialoglycans and the binding affinities of protein−ligand complexes, showing a net preference of SLBR-N for disialylated core 2 *O*-GalNAc structures (Figures 2 and S1).

Interaction of SLBR-N with 3′SLn: A Ligand- and Protein-Based Study. The molecular recognition features of 3′SLn (ligand **1**) by SLBR-N were unveiled by a combination of NMR and MD simulations. First, saturation transfer difference (STD) NMR confirmed the protein−ligand binding and mapped the recognized epitope, identifying the ligand protons in closest proximity to SLBR-N (Figure 3A). The acetyl group of sialic acid (AcK) received the highest magnetization from the receptor and was set to 100%; the relative STD intensities of the other protons were calculated accordingly. STD NMR experiments identified an extended binding epitope, involving all of the sugar moieties, with protons belonging to sialic acid and galactose units giving the strongest STD response, diagnostic of their closer proximity to the receptor (Figure 3A). The conformational behavior of **1** in both free and bound states was explored by NOESY and tr-NOESY experiments. The conformational flexibility for α2−3 sialoglycans depends on the behavior of the torsion angles around the Neu5Ac-α-(2−3)-Gal glycosidic linkage, namely, φ (C1−C2−O−C3′) and ψ (C2−O−C3′−H3′). Therefore, an almost stable ψ angle (around −11°), and a φ torsion oscillating around −60°, 180°, and 60°, were diagnostic of an equilibrium between the −g, t, and g conformers, respectively;[25,26] this equilibrium between different conformational states was indeed confirmed by NOESY analysis on **1** in the free state (Figure S2 and Table S1). The bioactive conformation adopted by ligand **1** upon binding was then explored by tr-NOESY experiments. A detailed analysis of key trNOE contacts indicated a propensity of **1** for adopting the −g conformation upon binding to SLBR-N (Figure 3B and Table S1). This was supported by the key NOE between H3 Gal and H8 Neu5Ac (B3−K8) and the absence of NOEs between H3 Gal and the diastereotopic H3 protons of Sia, indicative of the

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

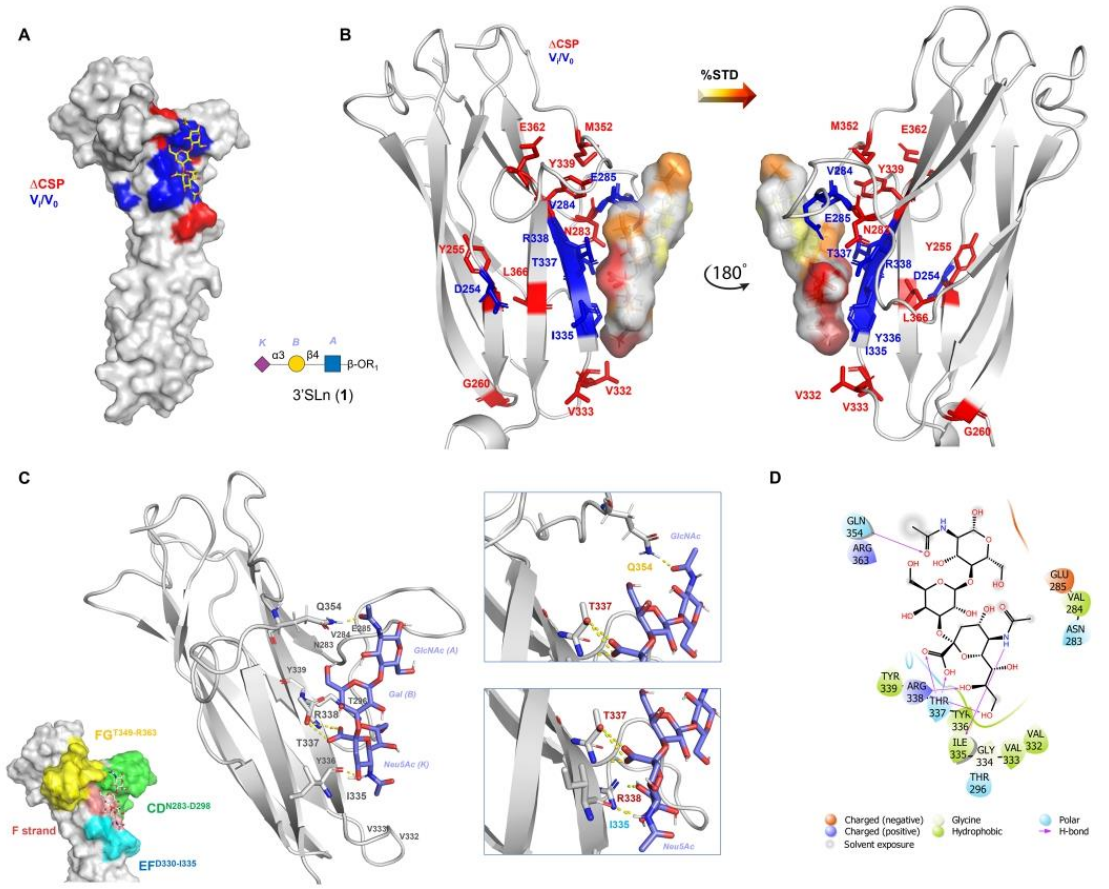


Figure 4. 3D model of the SLBR-N−**1** complex. (A) 3D complex with protein surface colored according to chemical shifts perturbation (in red) and intensity decreases (in blue) detected by protein-based NMR titration. (B) 3D views of the SLBR-N−**1** complex: the amino acids involved in the interaction as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code. (C) Representation of the SLBR-N−**1** complex with the CD, EF, and FG loops and F strand highlighted in green, cyan, yellow, and pink, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops and F strand legend). (D) 2D plot of interactions occurring at the SLBR-N−**1** interface.

3′SLn propensity to adopt in the bound state a φ torsion at the Neu5Ac-Gal glycosidic linkage around −60°. Therefore, NMR binding experiments confirmed that **1** displayed an extended recognized epitope and underwent conformer selection upon binding to SLBR-N.

Protein-based NMR experiments were used to evaluate the interaction from the protein viewpoint. Triple-labeled $^{2}H^{13}C^{15}N$ SLBR-N was expressed in *E. coli*, and then the protein resonances were assigned by 3D NMR experiments (Figure S3 and Supporting Information). Despite the low molecular weight (~23 kDa), deuteration and the acquisition of spectra with the TROSY-scheme were required to extend the T_2 relaxation times, due to the observed aggregation propensity of the protein at the concentration used for the NMR experiments. The evaluation of the binding of ligand **1** to SLBR-N was performed by titrating the ligand to the solution of the [U-15N] protein (Figure S4A) and following the variation in chemical shift and/or signal intensity of the cross-peaks in the 2D $^{1}H-^{15}N$ TROSY-HSQC spectrum of the protein.

Interestingly, the presence of **1** induced both a decrease in cross-peaks intensity (Figure S4B) and a chemical shift perturbation (CSP, Figure S4C) of some signals of SLBR-N at substoichiometric concentrations of the ligand. These variations allowed one to describe the binding region of SLBR-N in the presence of **1** (Figure S4D). In particular, the signals affected by the largest decreases in intensity were assigned to amino acids located on the CD loop (V284 and E285) and mainly to residues of the F strand (I335−R338) (see Scheme S1 and Figure S4A,B). Of note, signals corresponding to residues V284, E285, Y336, and T337 almost disappeared in the presence of low concentrations of the ligand (50 μM).

This result agreed with the expected binding mode of ligand **1**, because Y336 and T337 belonged to the φTRY consensus motif of SLBR-N and were bound to the sialic acid, while V284−E285 are on the adjacent CD loop. At the same time, the CSP analysis (Figure S4C) showed that residues of the F strand, and, in particular, the arginine and tyrosine in the φTRY consensus motif (R338 and Y339), as well as the adjacent residues of the

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

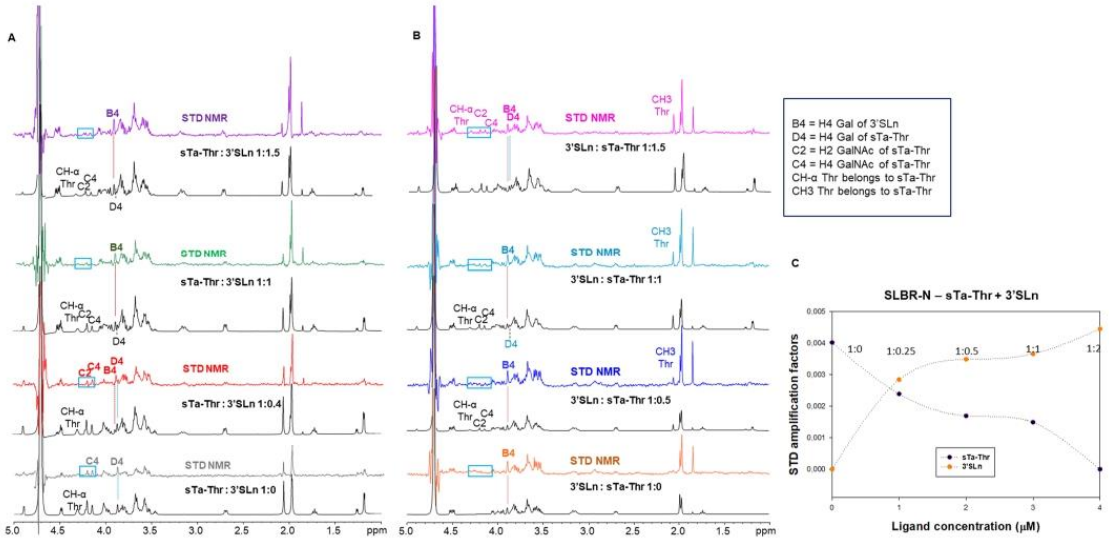


Figure 5. STD-derived competition experiments to compare the molecular binding of SLBR-N with ligands **1** and **2**. Two competition experiments were considered: the addition of ligand **1** to a mixture of SLBR-N + ligand **2**, and the addition of ligand **2** to a mixture of SLBR-N + ligand **1**. (A) STD NMR and off resonance spectra of the mixture containing SLBR-N (20 μM) and ligand **2** (2 mM) titrated with increasing concentrations of ligand **1** (from 800 μM to 3 mM). Key protons of the ligands were indicated in the STD NMR spectrum. At equal molar ligands ratio (1:1), the STD signals of ligand **2** disappeared, favoring the protein binding to ligand **1**. (B) STD NMR and off resonance spectra of the mixture containing SLBR-N (20 μM) and ligand **1** (2 mM) that was titrated with increasing concentrations of ligand **2** (from 500 μM to 3 mM). Key protons of the ligands were indicated in the STD NMR spectrum. STD NMR signals of ligand **2** were absent until an equal molar ligand ratio; low signals from threonine were detected only at a high molar excess of ligand **2**, likely deriving from nonspecific interactions. (C) Graphic view of the titration of SLBR-N−ligand **2** with ligand **1**: the STD amplification factors of H4 Gal belonging to ligand **2** (purple) and to ligand **1** (orange) were calculated at different ligands ratios (x axis), clearly confirming the preference of SLBR-N for **1** with respect to **2**.

EF loop (V332 and V333), experienced a large CSP. Changes in the chemical shifts of residues Y255, G260, L366, and N283, the latter belonging to the CD loop as well as M352 and E362 of the FG loop, were also observed. Of note, I335 and R338 were influenced by both CSP and variation of cross peak intensity. Furthermore, D254, Y255, and L366 of the β-strands close to the F-strand and M352 and E362 of the FG loop were indirectly affected by the interaction occurring between the neighboring amino acids and **1**, thus varying their chemical environment.

Overall, the spectral changes experienced by SLBR-N upon the addition of ligand **1** were in agreement with a protein−ligand binding affinity in the low micromolar range (see also native MS and ITC analyses in Figures 2B and S1). The combination of ligand- and protein-based NMR experiments allowed one to model **1** into the binding site of SLBR-N, and MD simulation analysis of the complex was performed to monitor the complex stability and dynamics of the binding in solution (Figures 4 and S5). Representative MD poses were then subjected to CORCEMA-ST (complete relaxation and conformational exchange matrix analysis of saturation transfer) analysis,[27] a program that calculates the predicted STD-NMR intensities for proposed molecular models of a ligand−receptor complex; the calculated and measured STD values can be then compared and used to validate a given complex. Notably, the calculated STDs, predicted by CORCEMA-ST analysis, matched well with those experimentally observed, therefore confirming the reliability of the 3D model (Figures 4 and S5D).

Therefore, the SLBR-N−3′SLn 3D model showed a full accommodation of the ligand into the receptor binding pocket defined by the Siglec domain, with the sialic acid residue exhibiting the most stable and significant interactions.

Interaction of SLBR-N with sTa-Thr. ESI−MS (Figure 2) and NMR experiments (Figure S6) revealed that sialyl-T-antigen (sTa-Thr, ligand **2**) was poorly recognized by SLBR-N. From a ligand perspective, the epitope mapping showed ligand **2** recognition by SLBR-N with the sialic acid residue mostly involved in the interaction (Figure S6A) with respect to the other residues, as was further confirmed by the contacts observed along the MD simulation (Figure S7). The CSP and the reduction of signal intensity were detected during the NMR titration of [U−^{15}N] SLBR-N with ligand **2** (Figure S6B−D); most of the intensity decreases and chemical shift variations were found in the Siglec domain region, and some residues of the Unique domain experienced a slight CSP. In particular, the residues on the F strand (I335, Y336, T337, and R338), containing the φTRY consensus motif, experienced a decrease in signal intensity, and all, but T337, which disappeared completely, experienced chemical shift changes. However, these effects were observed at only a high molar excess of the ligand, meaning a weak interaction between SLBR-N and ligand **2**.

These results were in accordance not only with the data derived by the MS analysis (Figure 2) but also were corroborated by computational studies. The weak affinity of ligand **2** for SLBR-N detected by ESI−MS experiments was indeed proven by the root-mean-square deviation (RMSD) observed along the trajectory of the MD simulation (Figure S7). Although ligand **2** remained in the SLBR-N binding pocket, the RMSD plot indicated a variation of the ligand coordinates along

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

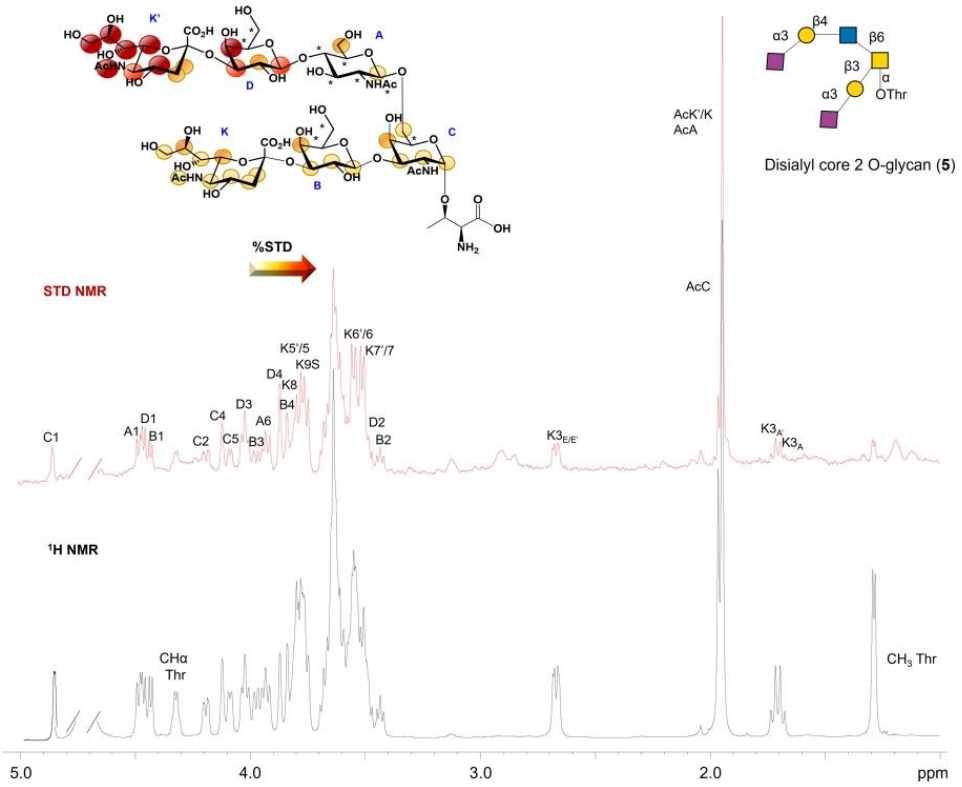


Figure 6. STD NMR analysis of SLBR-N and ligand **5**. The epitope mapping was calculated by $(I_0 - I_{sat})/I_0$, where $(I_0 - I_{sat})$ was the signal intensity in the STD-NMR spectrum (red) and I_0 was the peak intensity of the off-resonance spectrum (black). Sialic acid residues were not distinguishable by NMR, and their epitope mapping was derived according to protein NMR experiment theoretical STD effects calculated by CORCEMA-ST. The overlapped protons were indicated with asterisks on the chemical structure.

the MD trajectory as compared to the starting frame (Figure S7B); indeed, the absence of significant interactions between SLBR-N and **2** allowed a dynamic behavior of the ligand within the Siglec domain, differently from ligand **1** (Figures 4 and S5). As an example, the carboxylate group of Neu5Ac made H-bonds with T337 and could also establish a salt bridge with R338. Overall, the F strand of SLBR-N containing a φTRY consensus motif, together with I335, was the main portion of the protein in contact with ligand **2**, particularly with the sialic acid, while the rest of the ligand established fewer contacts with the receptor. However, the global interactions at the SLBR-N–ligand **2** interface were not significatively reproduced along the trajectory, suggesting the formation of a weak affinity complex (Figure S7).

Preference for 3′SLn versus sTa-Thr: Competition NMR Experiments and Monosialylated Core 2 *O*-Glycans. The higher affinity of SLBR-N for 3′SLn versus sTa-Thr (ligands **1** and **2**) measured by native ESI–MS and ITC experiments was also evaluated and confirmed by STD NMR competition experiments (Figure 5). The mixture of SLBR-N and ligand **2** was titrated with ligand **1** (Figure 5A). The addition of the competing high-affinity ligand was determined by an increase of STD signals of ligand **1** and a progressive reduction, until complete disappearance, of STD signals of ligand **2** (Figures 5A and S8). On the other hand, when the mixture of SLBR-N and ligand **1** was titrated with ligand **2** (Figures 5B and S9), almost no STD NMR response from sTa protons was observed. These data clearly indicated that ligand **2** was weakly recognized by SLBR-N, which in turn exhibited high affinity for **1**. Competition experiments were also performed by protein-based NMR experiments, confirming the above ligand-based NMR results. Indeed, upon addition of ligand **1** to the SLBR-N–ligand **2** mixture (Figure S10), CSP and a large decrease in cross peak intensity were detected, and the signals affected were almost the same observed during SLBR-N titration with ligand **1** (see above and Figure 4), further proof of a higher affinity complex. Finally, by comparing the root-mean-square fluctuation (RMSF) of the CD, EF, and FG loops of free and bound SLBR-N, a major stabilization was observed for the complex with ligand **1** (Figure S11).

Binding studies with monosialylated core 2 *O*-glycans (Scheme S1 and Figure 2, ligands **3** and **4**), alternatively terminating with 3′SLn and sTa-branch, further confirmed the preference of the protein for the sialyl-lactosamine epitope. These data revealed that both monosialyl core 2 *O*-glycans could be recognized by SLBR-N, with Sia's accommodating in the binding pocket according to the corresponding trisaccharides ligands **1** and **2** (see details in the Supporting Information and in Figure S12).

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

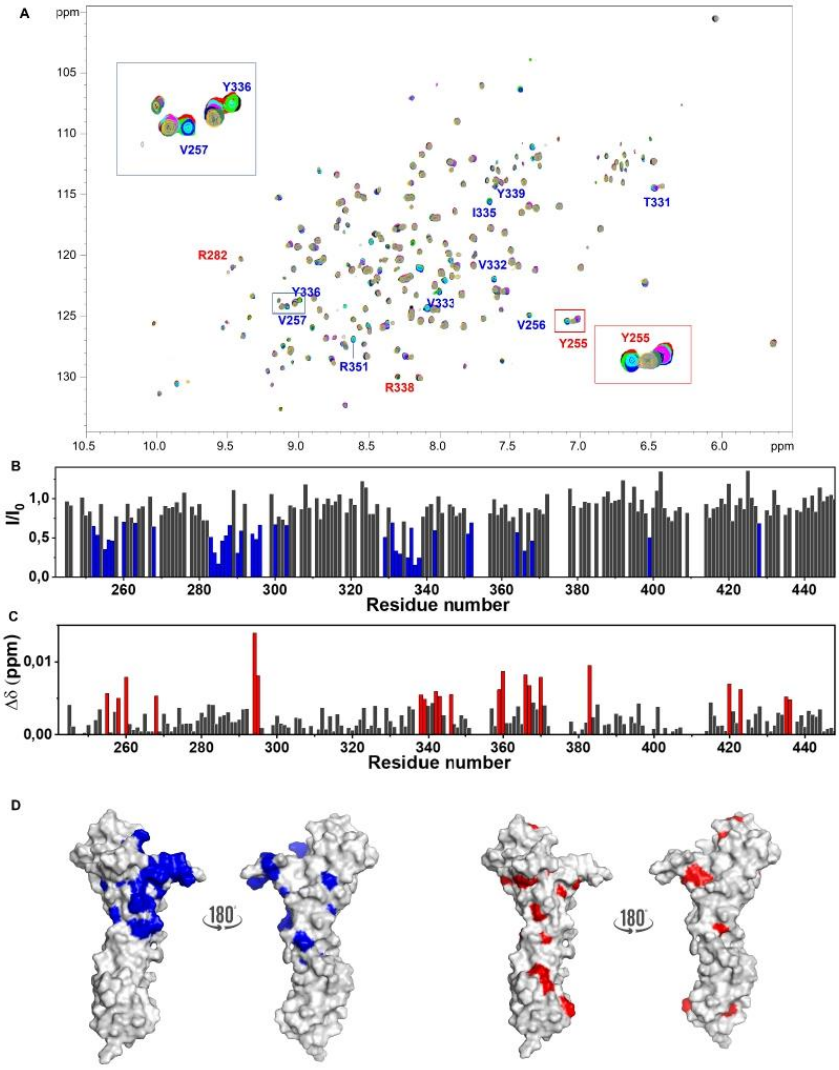


Figure 7. Protein-based NMR analysis of the SLBR-N–ligand **5** mixture. (A) 2D ^{1}H ^{15}N TROSY-HSQC NMR spectra of the SLBR-N free (black) and in the presence of 12.5 μM (red), 25 μM (green), 50 μM (blue), 100 μM (cyan), 200 μM (magenta), 400 μM (dark blue), 800 μM (dark green), 1600 μM (yellow), and 2080 μM (gray) of ligand **5**. The addition of the ligand leads to CSPs (red labels) and the decrease of intensity (blue labels) of some amino acids. Solutions were prepared in phosphate buffer pH 7.4 acquired on a spectrometer operating at 1.2 GHz at 298 K. (B) Plot representing the decreases in signal intensity per residue of SLBR-N in the presence of 200 μM of ligand **5** with respect to the protein (200 μM). The residues experiencing the largest signal intensity decrease (Y255, V256, V257, E268, N283, V284, E285, L286, D287, K288, T290, N291, Y294, L295, T296, L300, S303, L329, V332, V333, I335, Y336, T337, R338, A342, R351, M352, F364, L366, V368, V399, and Y428) have been highlighted in blue. (C) Plot of the CSP of the protein in the presence of 200 μM disialylated core 2-O glycan with respect to protein experiments that explored the binding and mapped the interacting epitope (Y258, E268, Y294, L295, R338, Y339, A342, T343, A346, D359, G360, L366, T367, S370, T383, Q420, T423, T435, and I436), which have been highlighted in red. (D) Surface representations of a model of the protein with highlighted residues experiencing the largest decreases in signal intensity (in blue) and the largest CSP (in red) in the presence of 100 μM ligand **5**.

Interaction of SLBR-N with Disialylated Core 2 *O*-Glycans. ESI–MS, ITC, NMR, and computational studies provided important information on how disialylated core 2 *O*-glycans (ligand **5**) could be recognized by SLBR-N. STD NMR experiments explored the binding and mapped the interacting epitope, showing that ligand **5** was extensively involved in the binding to SLBR-N (Figure 6).

Furthermore, titration of [U–^{15}N] SLBR-N with ligand **5** (Figure 7) showed that the area of the protein affected upon binding was more extensive if compared to ligands **1** and **2** (Figures S4 and S6). Indeed, ITC results showed that the binding was enthalpically driven, mainly due to the numerous establishments of hydrogen bonds between SLBR-N and ligand **5** (Figure S1B). The enthalpy measured by SLBR-N–ligand **5**

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX–XXX

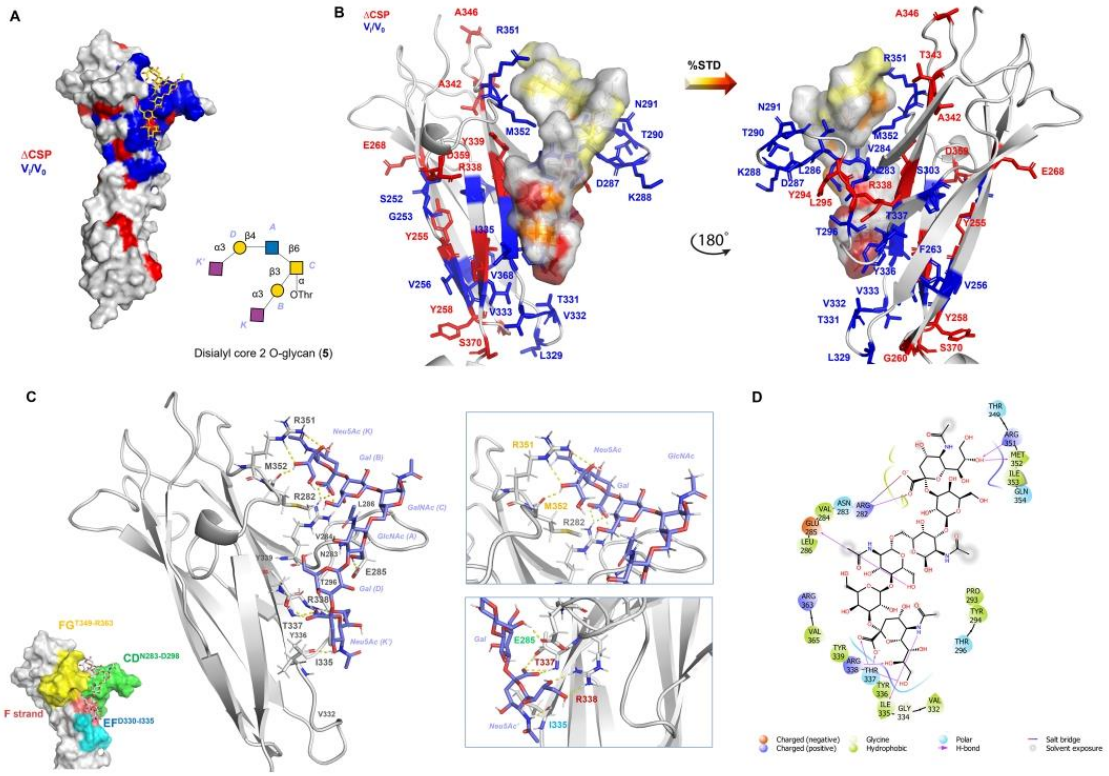


Figure 8. 3D model of the SLBR-N−**5** complex. (A) 3D complex with protein surface colored according to chemical shifts perturbation (in red) and intensity decreases (in blue) detected by protein-based NMR titration. (B) 3D views of the SLBR-N−**5** complex: the amino acids involved in the interaction as revealed by protein-based NMR experiments were represented as sticks; the ligand surface was colored according to the STD edit code. (C) Representation of the SLBR-N−**5** complex with CD, EF, and FG loops and the F strand highlighted in green, cyan, yellow, and pink, respectively. Different views highlighting the H-bonds monitored by MD simulation were shown (in the zoom, the amino acids were colored according to the loops and F strand legend). (D) 2D plot of interactions occurring at the SLBR-N−**5** interface.

binding was more favorable than that obtained by SLBR-N−ligand **1** interaction (−82.8 ± 1.65 vs −25.5 ± 0.979 kJ/mol); this result was reasonable given the higher number of contacts established at the interface with the protein, driven by the longer glycan chain of **5**, with two Sia residues.

Amino acids defining the binding site experienced a decrease in intensity as well as CSPs and mainly belonged to the Siglec domain (Figure 7).

In particular, residues Y336 and R338 of the φTRY consensus motif, as well the proximal residues, were affected by both decreases in signal intensity and CSP. On the basis of NMR data, MD simulation, and CORCEMA-ST (Figures S13 and S14), a 3D model of the complex was obtained. The entire ligand **5** was accommodated into the protein, with the 3′SLn branch bound to the F-strand of the Siglec domain, the region containing the φTRY sequence of SLBR-N (Figure 8).

In detail, Y336, T337, and R338 established H-bond interactions with Neu5Ac of the 3′SLn branch, including carboxylate and acetyl groups as well as the OH at positions 8 and 9 of the glycerol chain. On the other hand, Neu5Ac of the sTa branch established H-bonds with R351 and M352, both affected by an intensity decrease during protein titration, and a salt bridge with R282, affected by CSP. The above results indicated that the ligand was entirely splayed on the protein surface, with the 3′SLn arm sitting in the (main) protein binding site and the sTa branch establishing further polar interactions with amino acids of CD and FG loops. A recurrent H-bond between E285 and OH at position 6 of GlcNAc was further detected by MD simulation. Interestingly, this amino acid residue was affected by a decrease of signal intensity by NMR titration and significantly stabilized the protein−ligand interaction.

Indeed, as for ligand **6**, sulfation of position 6 of GlcNAc drastically reduced the affinity of the ligand to SLBR-N, as reported by ESI−MS data (Figure 2A). NMR binding data indicated that, although both arms of the ligand were recognized, the 3′SLn branch was the main engaged by the protein upon binding (Figure S15). By comparing the NMR and MD data of sulfated (ligand **6**) and nonsulfated (ligand **5**) *O*-glycans (Figure S16), it was possible to observe a similar accommodation into the binding site of SLBR-N, although more populated poses from MD cluster analysis and more stable contacts were found in the complex with ligand **5**, making it the preferred ligand (see also Discussion).

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

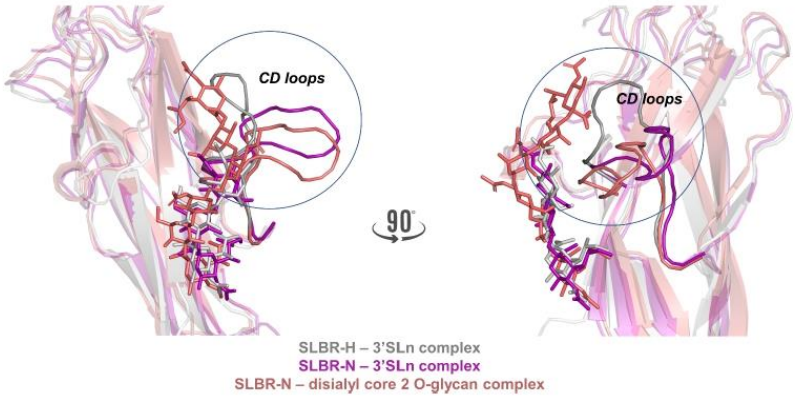


Figure 9. Comparison of the SLBR-H and SLBR-N binding modes with ligands **1** and **5**. The more open CD loop conformation of SLBR-N accommodates the complex disialyl core 2 *O*-glycan, which would cause steric hindrance with SLBR-H.

■ DISCUSSION

We here focused on the molecular details of host *O*-glycan recognition by the streptococcal Siglec-like adhesin SLBR-N of *S. gordonii* strain UB10712. Such protein-glycan systems are dynamic, intricate, and challenging complexes driven by a combination of forces including hydrogen bonds, van der Waals, electrostatic, and hydrophobic interactions, which were here investigated by combining complementary approaches and techniques. The combination of flow cytometry, native ESI−MS, ITC, NMR spectroscopy, and MD simulations allowed one to detect ligand binding affinities and to identify branched disialylated core 2 *O*-glycan as the preferred substrate of SLBR-N. Native mass spectrometry has emerged as an important tool in studying the structure and properties of the complexes. It is a versatile method used to evaluate noncovalently driven assemblies in a native-like folded state, providing information on the relative binding affinities and stoichiometry. Here, we could quantify weak interactions by using minimal amounts of protein and ligands, revealing the entire distribution of ligand-bound states and providing the K_D between SLBR-N and ligands **1**−**6**. The investigation, without the need for labeling or cross-linking, was performed on picomoles of material, while at the same time offering high resolution and a speed of analysis in the time scale of milliseconds. The order of magnitude of the association constants provided by the native mass spectrometry analysis is consistent with the NMR and ITC data. The addition of submillimolar/millimolar concentrations of the ligands to the investigated protein resulted in a sizable decrease in the intensity of the signals from the residues forming the specific binding site. Also, in the presence of a large excess of the ligands, new signals corresponding to the protein−ligand complexes were not observed.

Moreover, small chemical shift variations are observed for some signals. This behavior is consistent with an affinity in the micromolar range. Affinity constants calculated by ESI−MS were also in accordance with ITC results, confirming the preference of SLBR-N for ligand **1** with respect to **2** and showing ligand **5** as the preferred epitope. The NH backbone of the protein was assigned under per-deuteration conditions of expression and by performing 3D triple-resonance NMR experiments; the assignment was essential for the analysis of ligand interactions. In all complexes here studied, the main binding site was always comprised of the region containing $Y^{336}T^{337}R^{338}Y^{339}$ amino acids, a consensus motif common to all Siglec-like adhesins. We further revealed that SLBR-N was the only adhesin of its family that contained an isoleucine (I335) in the EF loop that formed a recurrent key contact with the acetyl group of sialic acid, confirming the importance of the amino acids on the F strand in the binding with sialoglycans (see Scheme S1). The main binding site of SLBR-N exhibited a clear preference for accommodating sialyl-lactosamine portion, as unveiled by native MS and ITC, showing the strongest affinities for the ligands containing the 3′SLn branch with respect to the sTa branch, and also confirmed by competition NMR experiments (Figures 2, 5, and S8−S10).

Interestingly, in complex core 2 *O*-glycans, both 3′SLn and sTa arms contributed to the interaction, as evidenced (i) by a wider binding region of SLBR-N when titrated with this longer ligand, (ii) by the STD NMR effects, variously and extensively involving most, if not all ligand protons of core 2 glycan (Figure 6), and (iii) by the numerous contacts monitored by MD, including polar, electrostatic, and hydrophobic interactions, which involved both sialic acids at the termini of the ligand arms (Figure 8). This is consistent with the thermodynamic parameters obtained by ITC, and in particular with the high negative enthalpy (ΔH −82.8 kcal/mol, Figures 2 and S1B) that confirmed the proposed 3D model (see also CORCEMA data, Figure S14) that involves additional contacts with respect to the shorter ligand **1**. Despite the large variation of enthalpies, the ΔG values were similar, because this latter term was compensated by a negative entropy. In carbohydrate recognition events, an entropic penalty is typically observed, due to glycan loss of flexibility from the free state to the interaction with a protein. Overall, for **1** and **5**, the reduction of entropy was compensated by a favorable enthalpic contribution. This was observed for ligand **1** that experienced a conformational selection upon binding around the glycosidic linkage (Figure 3B and Table S1). Furthermore, ligand **5** undergoes a stronger loss of conformational entropy upon complex formation, and this reduction of flexibility (Figures S1B and S14) supports the unfavorable entropy.[28] Thus, in complex core 2 *O*-glycans, accommodation of the 3′SLn branch into the main binding site represents the driving event that allows the interaction of the sTa branch to SLBR-N. Interestingly, while the carboxylate group of Sia belonging to the 3′SLn branch formed an H-bond with T337 of the main binding site, the carboxylate group of the Sia

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

belonging to the sTa branch established a salt bridge with R282, a crucial interaction observed in sialoglycans recognition by mammalian Siglec's.[29,30] Moreover, the presence of the sulfate group on the 3′SLn moiety of the core glycan did not influence the general orientation of the ligand into the SLBR-N binding site; rather, it precluded the binding of the GlcNAc residue to E285, due to electrostatic repulsions (Figure S16). Indeed, previous studies demonstrated that E295R substitution generated preference for the sulfated ligand with respect to the OH at position 6 of GlcNAc.[19] In the 3D models provided in this study, the negatively charged sulfate was instead found in contact with K288 of the CD loop of SLBR-N, but this interaction was weakly repeated along the MD simulation and, for this reason, considered negligible.

Despite the high sequence identity (80%) between SLBR-N and SLBR-H, Siglec-like adhesins from *S. gordonii* UB10712 and Challis, respectively, the ligand binding preferences are different, and their selectivity depends on the loop composition and flexibility.[14,20] Indeed, differently from SLBR-H,[15] the CD loop of SLBR-N assumed a more open conformation as shown by our 3D models of protein−ligand complexes (Figure 9) that defined a wider binding pocket with both 3′SLn and sTa branches and, consequently, accommodated the entire complex core 2 *O*-glycans. These data likely explain why the binding epitope of SLBR-H corresponds to a monosialylated core 1 *O*-glycan, supporting the results shown by Narimatsu et al.[18] The strongest preference of SLBR-N for the disialylated ligand was instead ascribable to the extended interactions with the amino acids on the F strand (YTRY) and those on CD, EF, and FG loops detected by NMR and MD. Among these, the hydrogen bonds between $I335^{EF}$ and the acetyl group of Sia of the 3′SLn branch, $E285^{CD}$ with OH6 of GlcNAc, $R351^{FG}$ and $M352^{FG}$ with the glycerol chain of Sia belonging to the sTa branch, as well as the hydrophobic interactions that involved $V284^{CD}$, $L285^{CD}$, $V332^{EF}$, and $M352^{FG}$ all contribute to an increased affinity of SLBR-N for the disialyl core 2 *O*-glycan.

Based on the above results, we can assess that the structural features of the binding between SLBR-N and host *O*-glycans are now clear at the atomic level. Our findings allowed us to determine the ligands epitope maps and their bioactive conformations in the receptor binding site as well as to define the amino acids of the protein binding pocket. Collectively, these outcomes described the architecture and the nature of the molecular interactions, providing the 3D models of different sialoglycans in complex with SLBR-N (Figure S17). Structural biology studies are always a necessary preliminary step toward the drug design and the chemical synthesis of entities that can tune a biochemical/immunological process. The versatility of the combined techniques used in this work, which include native mass spectrometry, calorimetry, biophysical, NMR, and computational approaches, represents a fundamental step to develop a site-specific interaction-based design strategy. By our approach encompassing an array of biophysical techniques, we have given new insights for developing potential high affinity mimetics that can maintain oral commensalism while hindering systemic streptococcal infections.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.3c01598.

Additional figures and experimental procedures employed for the production and assignment of SLBR-N, interaction with sialoglycans, MS, ITC, NMR ligand conformation experiments, and CORCEMA-ST (PDF)

Figure S17 (PDF)

Figures S1−S16 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Roberta Marchetti − *Department of Chemical Sciences, University of Naples Federico II, Naples 80138, Italy;* orcid.org/0000-0002-7173-7099; Email: roberta.marchetti@unina.it

Alba Silipo − *Department of Chemical Sciences, University of Naples Federico II, Naples 80138, Italy;* orcid.org/0000-0002-5394-6532; Email: silipo@unina.it

Authors

Cristina Di Carluccio − *Department of Chemical Sciences, University of Naples Federico II, Naples 80138, Italy*

Linda Cerofolini − *Magnetic Resonance Centre (CERM), CIRMMP, and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy*

Miguel Moreira − *Department of Chemical Sciences, University of Naples Federico II, Naples 80138, Italy;* orcid.org/0000-0003-1204-0416

Frédéric Rosu − *IECB Institut Européen de Chimie et Biologie, Pessac 33600, France;* Present Address: Spectrométrie de Masse Bio-Analytique Section des Sciences Pharmaceutiques Université de Genève, 1211 Geneva 4 Switzerland; orcid.org/0000-0003-3674-7539

Luis Padilla-Cortés − *Magnetic Resonance Centre (CERM), CIRMMP, and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy*

Giulia Roxana Gheorghita − *Magnetic Resonance Centre (CERM), CIRMMP, and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy; Giotto Biotech s.r.l., Sesto Fiorentino 50019, Italy*

Zhuojia Xu − *Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*

Abhishek Santra − *Department of Chemistry, University of California, Davis, Davis, California 95616, United States;* Present Address: Department of Organic Synthesis and Process Chemistry, CSIR-Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500007, India

Hai Yu − *Department of Chemistry, University of California, Davis, Davis, California 95616, United States;* orcid.org/0000-0002-4378-0532

Shinji Yokoyama − *Graduate School of Science, Osaka University, Osaka 565-0871, Japan*

Taylor E. Gray − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2R3, Canada*

Chris D. St. Laurent − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2R3, Canada*

Yoshiyuki Manabe − *Graduate School of Science, Osaka University, Osaka 565-0871, Japan;* orcid.org/0000-0002-5515-3923

Xi Chen − *Department of Chemistry, University of California, Davis, Davis, California 95616, United States;* orcid.org/0000-0002-3160-614X

Koichi Fukase − *Graduate School of Science, Osaka University, Osaka 565-0871, Japan;* orcid.org/0000-0001-8844-0710

Matthew S. Macauley − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2R3, Canada; Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta T6G 2R3, Canada*

Antonio Molinaro − *Department of Chemical Sciences, University of Naples Federico II, Naples 80138, Italy; Graduate School of Science, Osaka University, Osaka 565-0871, Japan;* orcid.org/0000-0002-3456-7369

Tiehai Li − *Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China;* orcid.org/0000-0002-3600-1828

Barbara A. Bensing − *San Francisco Veterans Affairs Medical Center and University of California, San Francisco, California 94121, United States*

Valérie Gabelica − *IECB Institut Européen de Chimie et Biologie, Pessac 33600, France;* Present Address: Spectrométrie de Masse Bio-Analytique Section des Sciences Pharmaceutiques Université de Genève, 1211 Geneva 4 Switzerland; orcid.org/0000-0001-9496-0165

Marco Fragai − *Magnetic Resonance Centre (CERM), CIRMMP, and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy;* orcid.org/0000-0002-8440-1690

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.3c01598

Author Contributions

A.S. and R.M. conceived and designed the study. C.D.C., L.C., L.P.-C., A.M., R.M., M.F., and A.S. carried out the NMR experiments and analyzed the results. C.D.C. ran and analyzed the ITC experiments and computational studies. C.D.C., G.R.G., M.M., and M.F. produced the protein. F.R. and V.G. carried out and analyzed the MS spectra. B.A.B. provided the plasmid. Z.X., A.S., H.Y., Y.M., X.C., K.F., and T.L provided the glycans. T.E.G., C.D.S.L., and M.S.M. ran the flow cytometry experiments. All authors wrote, revised, and reviewed the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy centre. M.F., A.S., L.P.-C., and G.R.G. acknowledge "Glytunes" H2020-MSCA-ITN-2020 (contract no. 956758). A.S. acknowledges the Ministry of Education, Universities and Research, PRIN2017 (2017XZ2ZBK, 2019−2023). A.S. and M.F. acknowledge PRIN MUR 2022 (2022ZEZS45). The United States National Institutes of Health (NIH) grant R01AI130684 for funding supported the synthetic efforts (X.C.). B.A.B. acknowledges the NIH and R03 DE029516. T.L. acknowledges the National Natural Science Foundation of China (22077130). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 851356 to R.M. A.S., A.M., and R.M. acknowledge PNRR, Missione 4 - Componente 2 - NextGenerationEU - Partenariato Esteso INF-ACT - One Health Basic and Translational Research Actions Addressing Unmet Needs on Emerging Infectious Diseases MUR: PE00000007.

■ ABBREVIATIONS

SRRP, serine-rich repeat glycoproteins; BR, binding region; IE, infective endocarditis; CSC, cancer stem cell; HIV, human immunodeficiency virus; ESI−MS, electrospray mass spectrometry; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; MD, molecular dynamic; CORCEMA-ST, complete relaxation and conformational exchange matrix analysis of saturation transfer; HSQC, heteronuclear single quantum coherence; TROSY, transverse relaxation optimized spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; ROESY, rotating frame Overhauser enhancement spectroscopy; CSP, chemical shift perturbation; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; 3′SLn, 3′-sialyl-*N*-acetyllactosamine; sTa, sialyl-T-antigen; sLeC, sialyl Lewis C; sLeX, sialyl Lewis X

■ REFERENCES

(1) Chamat-Hedemand, S.; Dahl, A.; Østergaard, L.; Arpi, M.; Fosbøl, E.; Boel, J.; Oestergaard, L. B.; Lauridsen, T. K.; Gislason, G.; Torp-Pedersen, C. Prevalence of infective endocarditis in streptococcal bloodstream infections is dependent on streptococcal species. *Circulation* **2020**, *142* (8), 720−730.

(2) Del Giudice, C.; Vaia, E.; Liccardo, D.; Marzano, F.; Valletta, A.; Spagnuolo, G.; Ferrara, N.; Rengo, C.; Cannavo, A.; Rengo, G. Infective endocarditis: A focus on oral microbiota. *Microorganisms* **2021**, *9* (6), 1218.

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

(4) Habib, G.; Hoen, B.; Tornos, P.; et al. Guidelines on the prevention, diagnosis, and treatment of infective endocarditis (new version 2009): the Task Force on the Prevention, Diagnosis, and Treatment of Infective Endocarditis of the European Society of Cardiology (ESC). Endorsed by the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) and the International Society of Chemotherapy (ISC) for Infection and Cancer. *Eur. Heart J.* **2009**, *30* (19), 2369−2413.

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

(6) Gaytán, M. O.; Singh, A. K.; Woodiga, S. A.; Patel, S. A.; An, S.-S.; Vera-Ponce de León, A.; McGrath, S.; Miller, A. R.; Bush, J. M.; van der Linden, M. A novel sialic acid-binding adhesin present in multiple species contributes to the pathogenesis of Infective endocarditis. *PLoS pathogens* **2021**, *17* (1), No. e1009222.

(7) Bensing, B. A.; Khedri, Z.; Deng, L.; Yu, H.; Prakobphol, A.; Fisher, S. J.; Chen, X.; Iverson, T. M.; Varki, A.; Sullam, P. M. Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. *Glycobiology* **2016**, *26*, 1222−1234.

(8) Takamatsu, D.; Bensing, B. A.; Prakobphol, A.; Fisher, S. J.; Sullam, P. M. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infection and immunity* **2006**, *74* (3), 1933−1940.

(9) Takamatsu, D.; Bensing, B. A.; Cheng, H.; Jarvis, G. A.; Siboo, I. R.; López, J. A.; Griffiss, J. M.; Sullam, P. M. Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibα. *Molecular microbiology* **2005**, *58* (2), 380−392.

(10) Plummer, C.; Wu, H.; Kerrigan, S. W.; Meade, G.; Cox, D.; Ian Douglas, C. A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. *British journal of haematology* **2005**, *129* (1), 101−109.

(11) Pyburn, T. M.; Bensing, B. A.; Xiong, Y. Q.; Melancon, B. J.; Tomasiak, T. M.; Ward, N. J.; Yankovskaya, V.; Oliver, K. M.; Cecchini, G.; Sulikowski, G. A. A structural model for binding of the serine-rich

repeat adhesin GspB to host carbohydrate receptors. *PLoS pathogens* **2011**, *7*, No. e1002112.
(12) Bensing, B. A.; Loukachevitch, L. V.; McCulloch, K. M.; Yu, H.; Vann, K. R.; Wawrzak, Z.; Anderson, S.; Chen, X.; Sullam, P. M.; Iverson, T. Structural Basis for Sialoglycan Binding by the Streptococcus sanguinis SrpA Adhesin*. *J. Biol. Chem.* **2016**, *291* (14), 7230−7240.
(13) Bensing, B. A.; Stubbs, H. E.; Agarwal, R.; Yamakawa, I.; Luong, K.; Solakyildirim, K.; Yu, H.; Hadadianpour, A.; Castro, M. A.; Fialkowski, K. P. Origins of glycan selectivity in streptococcal Siglec-like adhesins suggest mechanisms of receptor adaptation. *Nat. Commun.* **2022**, *13*, 2753.
(14) Bensing, B. A.; Li, L.; Yakovenko, O.; Wong, M.; Barnard, K. N.; Iverson, T.; Lebrilla, C. B.; Parrish, C. R.; Thomas, W. E.; Xiong, Y. Recognition of specific sialoglycan structures by oral streptococci impacts the severity of endocardial infection. *PLoS pathogens* **2019**, *15* (6), No. e1007896.
(15) Di Carluccio, C.; Forgione, R. E.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B. A. Molecular recognition of sialoglycans by streptococcal Siglec-like adhesins: toward the shape of specific inhibitors. *RSC Chemical Biology* **2021**, *2* (6), 1618−1630.
(16) Agarwal, R.; Bensing, B. A.; Mi, D.; Vinson, P. N.; Baudry, J.; Iverson, T. M.; Smith, J. C. Structure based virtual screening identifies novel competitive inhibitors for the sialoglycan binding protein Hsa. *Biochem. J.* **2020**, *477*, 3695−3707.
(17) Xiong, Y. Q.; Bensing, B. A.; Bayer, A. S.; Chambers, H. F.; Sullam, P. M. Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. *Microbial pathogenesis* **2008**, *45* (4), 297−301.
(18) Narimatsu, Y.; Joshi, H. J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y.-H.; Schjoldager, K. T.; Steentoft, C. An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Molecular Cell* **2019**, *75* (2), 394−407.
(19) Xu, Z.; Deng, Y.; Zhang, Z.; Ma, W.; Li, W.; Wen, L.; Li, T. Diversity-Oriented Chemoenzymatic Synthesis of Sulfated and Non-sulfated Core 2 O-GalNAc Glycans. *Journal of Organic Chemistry* **2021**, *86* (15), 10819−10828.
(20) Walker, M. R.; Goel, H. L.; Mukhopadhyay, D.; Chhoy, P.; Karner, E. R.; Clark, J. L.; Liu, H.; Li, R.; Zhu, J. L.; Chen, S. O-linked α2,3 sialylation defines stem cell populations in breast cancer. *Sci. Adv.* **2022**, *8*, No. eabj9513.
(21) Gabelica, V.; Livet, S.; Rosu, F. Optimizing native ion mobility Q-TOF in helium and nitrogen for very fragile noncovalent structures. *Journal of The American Society for Mass Spectrometry* **2018**, *29* (11), 2189−2198.
(22) Khristenko, N.; Rosu, F.; Largy, E.; Haustant, J.; Mesmin, C.; Gabelica, V. Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism. *J. Am. Chem. Soc.* **2023**, *145* (1), 498−506.
(23) van Schaick, G.; Haselberg, R.; Somsen, G. W.; Wuhrer, M.; Domínguez-Vega, E. Studying protein structure and function by native separation-mass spectrometry. *Nature Reviews Chemistry* **2022**, *6* (3), 215−231.
(24) Gabelica, V. Native mass spectrometry and nucleic acid G-quadruplex biophysics: advancing hand in hand. *Acc. Chem. Res.* **2021**, *54* (19), 3691−3699.
(25) Soares, C. O.; Grosso, A. S.; Ereño-Orbea, J.; Coelho, H.; Marcelo, F. Molecular recognition insights of sialic acid glycans by distinct receptors unveiled by NMR and molecular modeling. *Frontiers in Molecular Biosciences* **2021**, *8*, 727847.
(26) Forgione, R. E.; Di Carluccio, C.; Guzmán-Caldentey, J.; Gaglione, R.; Battista, F.; Chiodo, F.; Manabe, Y.; Arciello, A.; Del Vecchio, P.; Fukase, K. Unveiling molecular recognition of sialoglycans by human siglec-10. *Iscience* **2020**, *23* (6), 101231.
(27) Jayalakshmi, V.; Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand-Receptor Complexes. *J. Magn. Reson.* **2002**, *155* (1), 106−18.
(28) Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M. A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F. J. Minimizing the entropy penalty for ligand binding: lessons from the molecular recognition of the histo blood-group antigens by human galectin-3. *Angew. Chem., Int. Ed.* **2019**, *58* (22), 7268−7272.
(29) Crocker, P. R.; Paulson, J. C.; Varki, A. Siglecs and their roles in the immune system. *Nature Reviews Immunology* **2007**, *7* (4), 255−266.
(30) Di Carluccio, C.; Forgione, R. E.; Molinaro, A.; Crocker, P. R.; Marchetti, R.; Silipo, A. Exploring the fascinating world of sialoglycans in the interplay with Siglecs. *Carbohydr. Chem.* **2020**, *44* (44), 31−55.

https://doi.org/10.1021/acscentsci.3c01598
ACS Cent. Sci. XXXX, XXX, XXX−XXX

Solid-state NMR – A complementary technique for protein framework characterization

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2023, 59, 776

Received 21st October 2022,
Accepted 15th December 2022

DOI: 10.1039/d2cc05725e

rsc.li/chemcomm

Solid-state NMR – a complementary technique for protein framework characterization†

Linda Cerofolini, [ab] Kiefer O. Ramberg, [c] Luis C. Padilla,[ad] Paweł Antonik,[c] Enrico Ravera, [abd] Claudio Luchinat, [abd] Marco Fragai*[abd] and Peter B. Crowley *[c]

Protein frameworks are an emerging class of biomaterial with medical and technological applications. Frameworks are studied mainly by X-ray diffraction or scattering techniques. Complementary strategies are required. Here, we report solid-state NMR analyses of a microcrystalline protein–macrocycle framework and the rehydrated freeze-dried protein. This methodology may aid the characterization of low-crystallinity frameworks.

Protein crystals, which for decades have enabled advances in biomedical research, are currently in development as reaction vessels and templating devices with potential therapeutic applications.[1,2] Such materials are attractive as they are sustainable, biocompatible, programmable (from primary structure to crystalline assembly), and possess highly selective recognition and catalytic activities. Considerable effort is being invested in porous protein crystals, which possess large well-defined pores that permit the uptake and release of substrate/product cargo, enabling, for example, controlled drug delivery. Such "frameworks", with solvent contents >50%, can be achieved by using naturally porous cage proteins such as ferritin[3,4] and viral capsids,[5,6] or by engineered protein assembly with inducers that direct the formation of porous structures.[7–12] Among the assembly inducing strategies, metal ions/complexes[3,9] and organic ligands[7,8,10,11] are being used to noncovalently crosslink proteins. For example, Chen and co-workers reported an example of ligand-induced assembly based on rhodamine-sugar conjugates that enabled frameworks of concanavalin A.[7] Commercially-available macrocycles such as cucurbit[7]uril and sulfonato-calix[8]arene (**sclx$_8$**) have been shown to yield frameworks of different proteins including *Ralstonia solanacearum* lectin (RSL).[10–12]

To date, the structural characterization of protein frameworks has relied mainly on X-ray diffraction or scattering techniques.[4,6,10,11] However, alternative strategies are required to cater for frameworks that have reduced crystallinity or that lose crystallinity upon guest uptake.[4,10,12] Spectroscopic methods could be useful for framework characterization, providing additional information on the residues involved in ligand complexation. One such method is solid-state NMR (ssNMR), which thanks to on-going experimental advances yields spectra of microcrystalline, sedimented or freeze-dried proteins comparable in quality to solution-state spectra used for structural studies.[13–28] The sample conditions, in particular the protein hydration state (*e.g.* rehydration of freeze-dried samples), are crucial for achieving good quality spectra.[13,14,19]

Recently, we showed that solid-state spectral quality is sufficient to ensure resonance assignment, epitope mapping, and the calculation of structural models.[22,27] Here, we present a ssNMR characterization of microcrystalline RSL – **sclx$_8$** precipitates,[11] which form spontaneously at low pH and low salt. Complete solution-state NMR assignments of RSL in the sugar-free (BMRB 25952), D-mannose-(BMRB 25950) or L-fucose-bound (BMRB 25951) forms made this protein an ideal candidate.[29] For comparison, a freeze-dried sample of RSL devoid of **sclx$_8$** was characterized also. Such comparative analysis of the same protein in two distinct conditions provides information on the protein–macrocycle framework[11] that are additive with respect to the crystallographic data. Thus, we demonstrate the potential of ssNMR for the characterization of ligand–induced protein frameworks. We also put forward calixarene-mediated protein precipitation as a means of rapidly generating ordered protein solids suitable for ssNMR analysis.

RSL is a thermostable trimer with a 6-bladed β-propeller fold (C_3-symmetry) and micromolar affinity for L-fucose and related sugars.[29,30] The ~40% sequence identity between the N- and C-terminal blades of the RSL monomer results in pseudo C_6

[a] *Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, Sesto Fiorentino 50019, Italy. E-mail: fragai@cerm.unifi.it*
[b] *Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine (CIRMMP), Via L. Sacconi 6, Sesto Fiorentino 50019, Italy*
[c] *School of Biological and Chemical Sciences, University of Galway, University Road, Galway H91 TK33, Ireland. E-mail: peter.crowley@nuigalway.ie*
[d] *Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3, Sesto Fiorentino 50019, Italy*

† Electronic supplementary information (ESI) available: Sample preparation, NMR data acquisition, spectra and tables of assignments. See DOI: https://doi.org/10.1039/d2cc05725e

Table 1 RSL – **sclx$_8$** co-crystal forms[11]

Form	Space group	PDB id[a]	S.C. (%)[b]	$[(NH_4)_2SO_4]$[c] (M)	pH
I	$P2_13$	6Z60	36	>1.4	4.8–9.5
II	$I23$	6Z5G	66	0.8–1.2	≤4.0
III	$P3$	6Z5Q	59	0	≤4.2

[a] Representative PDB entries. [b] Solvent content estimated from total mass (protein plus **sclx$_8$**). [c] Approximate precipitant concentration at *circa* 1 mM RSL.

symmetry.[10] Frameworks of RSL and **sclx$_8$** have been characterized by X-ray crystallography and evidence for protein–calixarene binding in solution was obtained by NMR spectroscopy.[11] RSL co-crystallizes with **sclx$_8$** in at least three space groups depending on the precipitant and pH conditions (Table 1). Crystal forms II and III are frameworks (solvent content >50%) in which the crystal packing is dictated by the calixarene and there are no protein–protein contacts. In contrast, the freeze-dried RSL (devoid of **sclx$_8$**) requires protein–protein contacts. Crystal form III is particularly interesting as it occurs in the absence of a precipitant such as ammonium sulfate or polyethylene glycol. Millimetre-scale crystals, suitable for X-ray diffraction, grow within hours at pH 4 and 4 °C. Microcrystalline RSL – **sclx$_8$** precipitates are obtained within seconds at pH 3.4 and room temperature.[11]

Samples of microcrystalline RSL – **sclx$_8$** (Fig. S1, ESI†) or freeze-dried RSL were prepared with uniformly $^{13}C/^{15}N$-labelled protein in 20 mM potassium phosphate, 50 mM NaCl, 5 mM D-fructose, plus 0 or 10 mM **sclx$_8$** and pH adjusted to 3.4 (by adding HCl). Microcrystalline (9.0 mg) or freeze-dried (13.4 mg) samples were packed into 3.2 mm zirconia rotors (Bruker). All spectra were acquired on a Bruker Avance III spectrometer operating at 800 MHz (19 T, 201.2 MHz ^{13}C Larmor frequency) equipped with a Bruker 3.2 mm Efree NCH probe-head. Spectra were recorded at 14 or 20 kHz magic-angle spinning (MAS) frequency and the sample temperature was maintained at ~280 K. Complete details on sample preparation and data collection (Table S1, ESI†) are given in the ESI.†

The ssNMR spectra of microcrystalline RSL – **sclx$_8$** exhibit sharp, well-resolved signals and the 2D ^{15}N–^{13}C NCA spectrum was assigned by comparison with the solution state assignments for L-fucose-bound RSL (BMRB 25951).[29] The assignment was confirmed by analysing the 3D ssNMR spectra (Fig. S2, ESI†) and is reported in Fig. 1 and Table S2 (ESI†). As noted previously for the solution state assignments, homologous residues in the N- and C-terminal blades of the RSL monomer had similar chemical shifts in the solid-state spectrum (Fig. 1, *e.g.* D32/D77; G33/G78; G35/G80; W36/W81; G39/G84). All of the 2D ^{15}N–^{13}C NCA resonances, except for the N-terminus (S1-Q4), K34 and the C-terminus (T89-N90) were assigned. The N- and C-termini of RSL are mobile, as evidenced by high temperature factors in X-ray structures, which may explain the lack of cross-peaks.[11,30] K34 is a key residue in the **sclx$_8$** binding site and the occurrence of multiple conformations may explain the absence of a well-resolved cross-peak. The K34 side chain is well-defined in RSL – **sclx$_8$** co-crystal structures at pH ~ 4 but tends to be disordered at high pH.[11] While the microcrystalline sample used for ssNMR was obtained under conditions

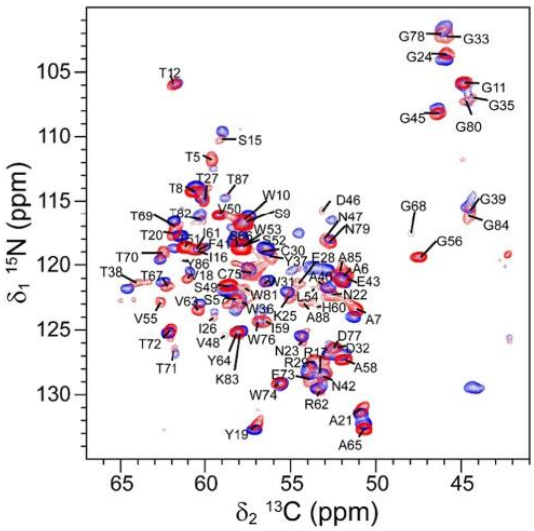


Fig. 1 The assigned 2D ^{15}N–^{13}C NCA solid-state spectrum of microcrystalline RSL – **sclx$_8$** (red) overlaid on the spectrum of rehydrated freeze-dried RSL (blue).

similar to the *P*3 crystal form (Table 1), there are differences in the samples that may give rise to multiple K34 conformations and resonance broadening.

The RSL side chains were assigned by analysing the 3D ^{15}N–^{13}C NCACX (Fig. S2, ESI†) and 2D ^{13}C–^{13}C dipolar assisted rotational resonance (DARR) solid-state spectra (Table S2, ESI†). A visual inspection of the DARR spectrum (Fig. 2) revealed that ~40% of the signals were lower intensity than the rest. The lower intensity signals belong to residues V13-R17, I26-E28, C30-A40, D46-V48, H60-R62, G80-N90 spread throughout the protein but involved or close to the sugar and **sclx$_8$** binding sites. Structural heterogeneity[29] may be responsible for this signal broadening. The occurrence of cross-peaks in the 2D ^{13}C–^{13}C DARR spectra was consistent with preservation of the trimeric structure of RSL, as expected for this highly stable protein. Cross-peaks between T5 and N47, A6 and S49, and S9 and T51,

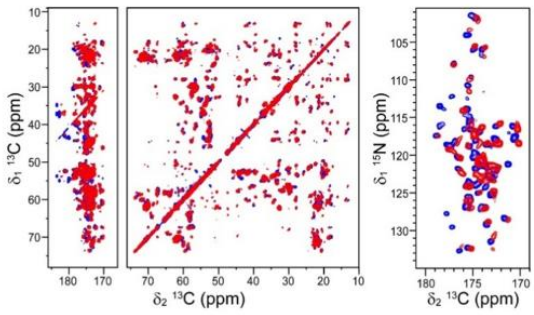


Fig. 2 Overlaid 2D ^{13}C–^{13}C DARR and ^{15}N–^{13}C NCO spectra of microcrystalline RSL – **sclx$_8$** (red) and rehydrated freeze-dried RSL (blue).

were assigned in the ^{13}C–^{13}C DARR spectrum with a 100 ms mixing time. These residues are located on adjacent β-strands at the inter-monomer interface and line the central channel of the β-propeller fold. Signal overlap prevented the identification of other inter-monomer cross-peaks.

For comparison, freeze-dried RSL was characterized by ssNMR. The freeze-dried sample was rehydrated until the maximum resolution was achieved in the 1D $\{^1H\}^{13}C$ cross-polarization (CP) spectrum (ωH = 78 kHz; ωC = 50 kHz, Fig. S3, ESI†).[13,14,19] This optimisation process required ~24 hours, in stark contrast to the microcrystalline sample that required no optimisation. A comparison of the 1D $\{^1H\}^{13}C$ CP spectra of microcrystalline RSL – **sclx$_8$** and rehydrated freeze-dried RSL revealed slightly broader signals in the latter (Fig. S4, ESI†). 1D projections along the ^{13}C dimension of representative signals showed 5–10% broader peaks in rehydrated freeze-dried RSL compared to microcrystalline RSL – **sclx$_8$** (Fig. S5, ESI†). The signal-to-noise ratio was ~1.1-fold better for the rehydrated freeze-dried sample (230) compared to microcrystalline RSL – **sclx$_8$** (203), which can be attributed to the higher amount of protein in the former and to the porous nature (60% solvent) of the latter.[19] The 2D ^{15}N–^{13}C NCA spectrum of rehydrated freeze-dried RSL (Fig. 1) was assigned by comparison with the assignments for microcrystalline RSL – **sclx$_8$**, and confirmed by 3D ssNMR spectral analysis (Table S3, ESI†). Similar to the microcrystalline sample, all of the 2D ^{15}N–^{13}C NCA resonances were assigned, except for the N-terminus (S1-Q4), K34 and the C-terminus (T89-N90). It is intriguing that the K34 resonance was also unassigned in this sample. The other two lysines, K25 and K83, were assigned (Fig. 1). Interestingly, some signals belonging to residues in loops or flexible regions (G33, L54, V55, G56, G68, T69, A88) had a lower intensity.

The assigned 2D ^{15}N–^{13}C spectra for both rehydrated freeze-dried RSL and microcrystalline RSL – **sclx$_8$** enabled an analysis of the effects of the calixarene and different packing (protein–macrocycle *versus* protein–protein) in the two materials. Fig. 3 illustrates the chemical shift perturbations (CSP, $\Delta\delta$) between the ^{15}N and $^{13}C^{\alpha}$ resonances of rehydrated freeze-dried RSL and microcrystalline RSL – **sclx$_8$**, evaluated according to $\Delta\delta = \frac{1}{2}\sqrt{(\Delta\delta_{C\alpha}/2)^2+(\Delta\delta_N/5)^2}$. The resonances with perturbations above the cut-off (mean + std. dev. ≥0.16 ppm) were Arg17, Glu28, Trp36, Tyr37, Thr38, Ala40, Phe41, Asp46, Trp81, Gly84, Ala85, Tyr86, and Ala88. Of these 13 residues, all (except Arg17) occur near the protein surface with either or both the N and C^{α} groups in proximity with solvent. Thus, these groups may be sensing different protein–macrocycle, protein–protein and/or protein–solvent packing that occur in the two materials. Only three of the residues (Arg17, Tyr37 and Asp46) make contact with **sclx$_8$** in the X-ray crystal structure (PDB 6Z5Q) that was obtained under similar conditions to the microcrystalline sample used for ssNMR analysis. Interestingly, the largest CSP occurred for Asp46, which undergoes a substantial side chain conformation change to complex **sclx$_8$**.[11] Other important **sclx$_8$**-binding residues such as Val13, Asn23 and Thr67 did not exhibit significant $\Delta\delta$ effects.

In addition to CSP effects, there were also significant differences in line widths. Signal intensity variations in the 2D ^{15}N–^{13}C NCA

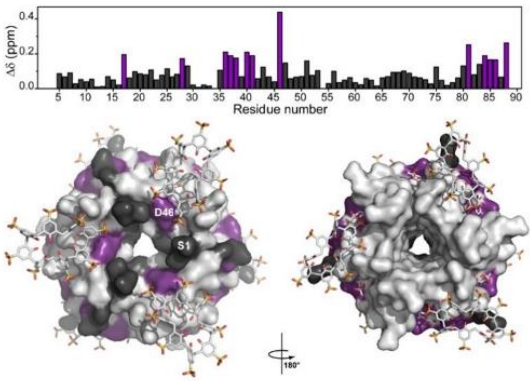


Fig. 3 Plot of CSP between the ^{15}N and $^{13}C^{\alpha}$ resonances of rehydrated freeze-dried RSL and microcrystalline RSL – **sclx$_8$**. Residues with $\Delta\delta \geq$ 0.16 ppm are highlighted purple and mapped to the RSL – **sclx$_8$** crystal structure (PDB 6Z5Q). The RSL trimer is rendered as a light grey surface. Residues for which data were unavailable (1–4, K34 and 89–90) are dark grey. **sclx$_8$** and D-fructose are shown as sticks.

spectra of microcrystalline RSL – **sclx$_8$** and rehydrated freeze-dried RSL were analysed as follows: the peak intensities in each spectrum were normalized *versus* the average signal intensity; the normalized peak intensities were subtracted to give a difference plot (Fig. 4) yielding 13 and 18 higher intensity resonances (based on a cut-off of 0.36) in the microcrystalline and rehydrated freeze-dried spectra, respectively. This analysis indicates structural heterogeneity across the protein including surface patches and features involved in calixarene binding. Intensity differences were evident also in the 2D ^{13}C–^{13}C DARR spectrum, in particular, for the carboxylate groups. RSL contains six acidic residues, some of which had sharper signals in the rehydrated freeze-dried RSL compared with microcrystalline

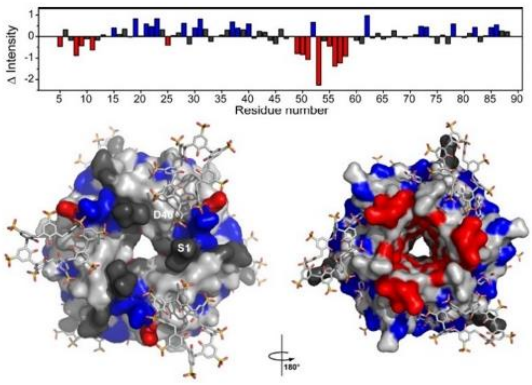


Fig. 4 Plot of the normalised intensity changes between the ^{15}N–$^{13}C^{\alpha}$ resonances of rehydrated freeze-dried RSL and microcrystalline RSL – **sclx$_8$**. Higher intensity signals are indicated by positive bars for rehydrated freeze-dried RSL and by negative bars for microcrystalline RSL – **sclx$_8$**. Residues with significant difference in signal intensity (highlighted blue and red, respectively) were mapped on to the RSL – **sclx$_8$** crystal structure (PDB 6Z5Q).

RSL – $\mathbf{sclx_8}$ (Fig. S6, ESI†). In particular, Asp32 and Asp46 that bind $\mathbf{sclx_8}$ were significantly broadened in the microcrystalline sample. CSP and line broadening were evident also for the carboxylate resonances of homologous Glu28 and Glu73, which form hydrogen bonds to D-fructose in the sugar binding sites. Glu28 (near Lys34) might sense calixarene binding *via* the sugar as this site makes contact with $\mathbf{sclx_8}$, while the site containing Glu73 does not bind the calixarene. Previous comparative ssNMR analyses of protein dynamics in crystalline and rehydrated freeze-dried samples indicated no significant differences in backbone or side chain mobility.[16]

In summary, high-quality ssNMR spectra were acquired for both microcrystalline RSL – $\mathbf{sclx_8}$ and rehydrated freeze-dried RSL. The microcrystalline sample exhibits slightly better resolved spectra, as expected for a crystalline sample, but fewer cross-peaks. This seemingly contradictory finding is explained by a larger structural heterogeneity arising from residue masking by $\mathbf{sclx_8}$ that resulted in signal broadening, in some cases beyond detection. The quality of the multidimensional spectra was sufficient to allow near complete resonance assignment and the identification of restraints useful for structural calculation. Therefore, calixarene-mediated protein precipitation may be a straightforward and rapid route to high-quality ssNMR samples. Despite the differences in intensity and the CSP affecting some signals, the spectra of microcrystalline RSL – $\mathbf{sclx_8}$ and rehydrated freeze-dried RSL are largely superimposable (Fig. 1 and 2). A good match occurred also between the cross-peaks in the solution-state[29] and solid-state spectra. Furthermore, the data show the versatility of ssNMR and its virtual independence from a highly ordered distribution of the proteins in the material as required for X-ray crystallography. This feature is particularly relevant for the characterization of new protein-based materials where the presence of additional guest molecules can lead to decreased crystallinity, thus preventing X-ray characterization.[4] This advance is relevant to ferritin, which has been characterized by ssNMR[17] and is prevalent in protein framework studies.[1,3]

This work was supported by Regione Toscana (CERM-TT and BioEnable), the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN 2017A2KEPL), the Recombinant Proteins JOYNLAB laboratory, the project FISR2021_SYLCOV, University of Galway, NUI Travelling Studentship, Teagasc Walsh fellowship and Science Foundation Ireland (grant 13/CDA/2168). We acknowledge also Instruct-ERIC, specifically CERM/CIRMMP, H2020 INFRAIA iNEXT-Discovery (contract 871037), H2020 PANACEA (grant 101008500), EOSC-Life (contract 824087) and H2020-MSCA-ITN-2020 (contract 956758).

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 M. Kojima, S. Abe and T. Ueno, *Biomater. Sci.*, 2022, **10**, 354–367.
2 R. Fernández-Penas, C. Verdugo-Escamilla, S. Martínez-Rodríguez and J. A. Gavira, *Cryst. Growth Des.*, 2021, **21**, 1698–1707.
3 J. B. Bailey, L. Zhang, J. A. Chiong, S. Ahn and F. A. Tezcan, *J. Am. Chem. Soc.*, 2017, **139**, 8160–8166.
4 K. Han, Y. Na, L. Zhang and F. A. Tezcan, *J. Am. Chem. Soc.*, 2022, **144**, 10139–10144.
5 M. Uchida, K. McCoy, M. Fukuto, L. Yang, H. Yoshimura, H. M. Miettinen, B. LaFrance, D. P. Patterson, B. Schwarz, J. A. Karty, P. E. Prevelige, B. Lee and T. Douglas, *ACS Nano*, 2018, **12**, 942–953.
6 M. Uchida, N. E. Brunk, N. D. Hewagama, B. Lee, P. E. Prevelige, V. Jadhao and T. Douglas, *ACS Nano*, 2022, **16**, 7662–7673.
7 F. Sakai, G. Yang, M. S. Weiss, Y. Liu, G. Chen and M. Jiang, *Nat. Commun.*, 2014, **5**, 4634.
8 B. E. Partridge, P. H. Winegar, Z. Han and C. A. Mirkin, *J. Am. Chem. Soc.*, 2021, **143**, 8925–8934.
9 L. Vandebroek, H. Noguchi, K. Kamata, J. R. H. Tame, L. van Meervelt, T. N. Parac-Vogt and A. R. D. Voet, *Chem. Commun.*, 2020, **56**, 11601–11604.
10 F. Guagnini, S. Engilberge, K. O. Ramberg, J. Pérez and P. B. Crowley, *Chem. Commun.*, 2020, **56**, 360–363.
11 K. O. Ramberg, S. Engilberge, T. Skorek and P. B. Crowley, *J. Am. Chem. Soc.*, 2021, **143**, 1896–1907.
12 K. O. Ramberg, F. Guagnini, S. Engilberge, M. A. Wrońska, M. L. Rennie, J. Pérez and P. B. Crowley, *Chem. – Eur. J.*, 2021, **27**, 14619–14627.
13 S. D. Kennedy and R. G. Bryant, *Biopolymers*, 1990, **29**, 1801–1806.
14 J. Pauli, B. van Rossum, H. Förster, H. J. de Groot and H. Oschkinat, *J. Magn. Reson.*, 2000, **143**, 411–416.
15 F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein and H. Oschkinat, *Nature*, 2002, **420**, 98–102.
16 A. Krushelnitsky, Y. Gogolev, R. Golbik, F. Dahlquist and D. Reichert, *Biochim. Biophys. Acta.*, 2006, **1764**, 1639–1645.
17 I. Bertini, C. Luchinat, G. Parigi, E. Ravera, B. Reif and P. Turano, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10396–10399.
18 M. Chan-Huot, L. Duma, J.-B. Charbonnier, J.-E. Herbert-Pucheta, L. Assairi, Y. Blouquit, D. Abergel and G. Bodenhausen, *Cryst. Growth Des.*, 2012, **12**, 6199–6207.
19 M. Fragai, C. Luchinat, G. Parigi and E. Ravera, *J. Biomol. NMR*, 2013, **57**, 155–166.
20 M. Fragai, C. Luchinat, T. Martelli, E. Ravera, I. Sagi, I. Solomonov and Y. Udi, *Chem. Commun.*, 2014, **50**, 421–423.
21 E. Ravera, S. Ciambellotti, L. Cerofolini, T. Martelli, T. Kozyreva, C. Bernacchioni, S. Giuntini, M. Fragai, P. Turano and C. Luchinat, *Angew. Chem., Int. Ed.*, 2016, **55**, 2446–2449.
22 L. Cerofolini, S. Giuntini, A. Carlon, E. Ravera, V. Calderone, M. Fragai, G. Parigi and C. Luchinat, *Chem. – Eur. J.*, 2019, **25**, 1984–1991.
23 J. Mao, V. Aladin, X. Jin, A. J. Leeder, L. J. Brown, R. C. D. Brown, X. He, B. Corzilius and C. Glaubitz, *J. Am. Chem. Soc.*, 2019, **141**, 19888–19901.
24 B. R. Sahoo, S. J. Cox and A. Ramamoorthy, *Chem. Commun.*, 2020, **56**, 4627–4639.
25 K. Jaudzems, A. Kirsteina, T. Schubeis, G. Casano, O. Ouari, J. Bogans, A. Kazaks, K. Tars, A. Lesage and G. Pintacuda, *Angew. Chem., Int. Ed.*, 2021, **60**, 12847–12851.
26 S. Sarkar, B. Runge, R. W. Russell, K. T. Movellan, D. Calero, S. Zeinalilathori, C. M. Quinn, M. Lu, G. Calero, A. M. Gronenborn and T. Polenova, *J. Am. Chem. Soc.*, 2022, **144**, 10543–10555.
27 D. Rizzo, L. Cerofolini, S. Giuntini, L. Iozzino, C. Pergola, F. Sacco, A. Palmese, E. Ravera, C. Luchinat, F. Baroni and M. Fragai, *J. Am. Chem. Soc.*, 2022, **144**, 10006–10016.
28 S. Ahlawat, K. R. Mote, N. A. Lakomek and V. Agarwal, *Chem. Rev.*, 2022, **122**, 9643–9737.
29 P. M. Antonik, A. N. Volkov, U. N. Broder, D. Lo Re, N. A. J. van Nuland and P. B. Crowley, *Biochemistry*, 2016, **55**, 1195–1203.
30 N. Kostlánová, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerová and A. Imberty, *J. Biol. Chem.*, 2005, **280**, 27839–27849.

5.2 NMR for the Characterization of Drug Delivery Systems

Characterization of the preservation of the Higher Order Structure of erythrocyte-encapsulated biologics

Characterization of the preservation of the High Order Structure of erythrocyte-encapsulated biologics

Luis Padilla-Cortés[a#], Giulia Roxana Gheorghita[a,b#], Bianca Susini[a], Francesco Currò[a], Pasquale Russomano[a], Linda Cerofolini[a], Marco Fragai[a,b]

# Equal Contribution

[a] Magnetic Resonance Centre (CERM), CIRMMP and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

[b] Giotto Biotech s.r.l., Sesto Fiorentino, Italy

Introduction

Therapeutic enzymes are currently used in the treatment of several diseases. Enzymatic replacement in genetically deficient patients or the clearance of essential precursors for cancer cells are two of the most representative examples[1,2]. Often, the benefits are limited due to poor in vivo stability, immunogenicity, and the related inactivating antibodies. These problems can be partially resolved by coating the therapeutic proteins with biocompatible polymers like PEG, which extend their half-life by increasing the size and masking the enzyme from proteases and the reticuloendothelial system[3,4]. However, pegylation is also plagued by several side effects that are emerging because of its more extensive use. In particular, hypersensitivity and immune reactions to PEG up to anaphylaxis, and cytoplasmic vacuolation have been documented for pegylated protein delivery[5,6]. The idea of using of erythrocytes to deliver enzymes is even older than pegylation and it was firstly reported for beta-glucosidase and beta-galactosidase and proposed Gaucher's disease[7]. Erythrocytes have been experimentally tested in treating patients due to their biocompatibility, biodegradability, long life in the bloodstream, and ability to decrease side effects. They protect the drug from the immune system and plasma proteases, allowing for increased therapeutic effects and safer blood concentrations[8].

Although red cells are now being considered also for the delivery of small molecules, several erythrocyte-encapsulated enzymes are currently under development, with one of them having undergone clinical evaluation. L-Asparaginase was one of the first approved biologics used to treat acute lymphoblastic leukemia (ALL) and lymphoblastic lymphoma[9,10]. This enzyme derived from various sources is still used in its unmodified and pegylated forms, and its encapsulated erythrocyte formulation (GRASPA) has already been evaluated in clinical trials against ALL and pancreatic cancer[11–13]. Although the company is no longer pursuing GRASPA’s approval for clinical use against ALL, this therapeutic candidate remains a model for developing analytical tools to characterize this innovative class of biologics. To date the characterization of the erythrocyte-encapsulated enzymes has indeed largely relied on assessing their enzymatic activity, either within the

encapsulated formulation itself or after extracting the enzyme following cell lysis. However, the assessment of the preservation of the high order structure (HOS) is important for a comprehensive understanding of these formulations and for the optimization of the encapsulation protocols, handling of the sample and storage. In this regard, solution NMR is extensively used to investigate the structural features of proteins and peptides also embedded in cells, hydrogels and in general, in confined environments. Here we show how the NMR characterization of three erythrocyte-encapsulated proteins provide information on their HOS preservation together with a semiquantitative evaluation of the encapsulation yield.

Results

Biologics, unlike small molecules, are produced in living systems, and are characterized by a primary, secondary, tertiary and sometimes quaternary structure. Assessment of the Higher Order Structure (HOS) of therapeutic proteins encapsulated within red blood cells (RBCs) is crucial for ensuring their functionality and stability[14]. In this study, three proteins –Carbonic Anhydrase II (CAII), Transthyretin (TTR), and Asparaginase II (ANSII)– were selected as model biomolecules to evaluate their structural integrity upon encapsulation based on expression levels, structural complexity and therapeutic relevance. Using solution NMR, we investigated the preservation of the HOS of these proteins within the erythrocyte environment. The results highlight the suitability of NMR for assessing protein structure in complex matrices and offer insights into the encapsulation efficiency and structural stability of these therapeutics.

Selection of protein drug targets

CAII is a 30 kDa enzyme that catalyzes the interconversion between CO_2 and H_2O to the respective carbonic acid ions, such as carbonate and bicarbonate. The enzyme was chosen as a model for the optimization of the encapsulation protocol given its high over-expression levels in a bacterial vector and its stability in concentrated solution, up to 3 mM.

TTR, is a 55 kDa homotetrameric hormone carrier with a complex quaternary structure[15,16]. The protein is a potential drug carrier[17], as well as a suitable model to investigate the preservation of the HOS of a tetrameric biomolecule upon erythrocyte encapsulation[18].

ANSII, on the other hand, consists of a 138 kDa dimer of dimer and is an important therapeutic drug in the treatment of several types of cancer, in particular for acute lymphoblastic leukemia (ALL). The enzyme has already been encapsulated in red blood cells (GRASPA® [Globule Rouge ASPAraginase], Erytech)[13] and the formulation has been already evaluated in clinical trials against ALL and pancreatic cancer, hence, it comes as an ideal model for the structural characterization of biologics within the erythrocyte environment[19].

NMR Spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful tool for assessing the HOS of proteins, particularly in complex matrices or in pharmaceutical formulation, due to its atomic resolution, short acquisition times and little sample preparation. [U-^{15}N] CAII, [U-^{15}N] TTR and [U-^{2}H, ^{13}C, ^{15}N] ANSII, were encapsulated in red blood cells and used as models for the characterization of the preservation of the HOS of the proteins within the erythrocyte environment. Encapsulation and structural characterization were monitored using ^{1}H-^{15}N sofast-HMQC NMR experiments instead of ^{1}H-^{15}N-HSQC (for CAII) or ^{1}H-^{15}N-TROSY HSQC (for TTR and ANSII) given the low lifetime of the NMR samples. The ^{1}H-^{15}N sofast-HMQC spectra acquired on the encapsulated CAII and TTR exhibited good quality. The backbone resonances on the spectra are superimposable with those of the "free", non-encapsulated proteins as shown in Figure 1, demonstrating the preservation of the HOS.

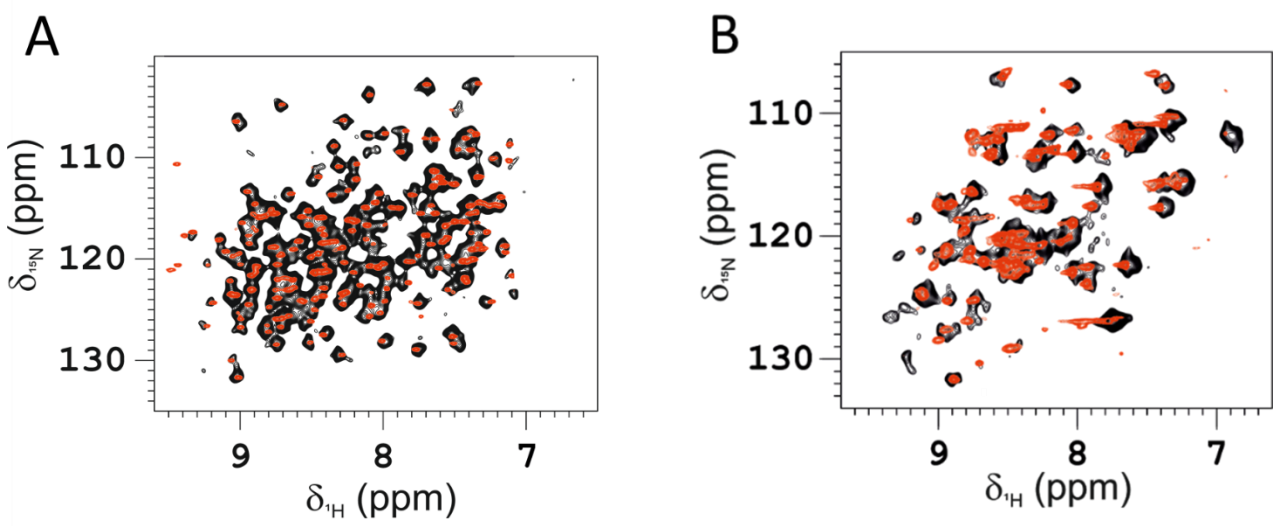


Figure 1. *^{1}H-^{15}N sofast-HMQC NMR spectra of A) CAII encapsulated in RBC (black) and free CAII (red), and B) TTR encapsulated in RBC (black) and free TTR (red). Experiments were acquired at a spectrometer operating at 1200 MHz (^{1}H Larmour Frequency), at 310 K.*

Furthermore, a partial assignment of the most-resolved peaks was easily obtained by comparison with the assigned ^{1}H-^{15}N HSQC spectrum of the free CAII with the ^{1}H-^{15}N sofast-HMQC spectrum acquired on the encapsulated protein (Figure 2A). Most of the residues whose resonances were re-assigned, are located in all of the secondary structural motifs of the protein (Figure 2B), serving as a further proof of the HOS preservation of the encapsulated biomolecule. In the case of TTR, the signals on the ^{1}H-^{15}N sofast-HMQC are affected by line broadening even for the free protein given its size (MW 55 kDa), and the quality of the spectrum inside the cells is close to that one of the free proteins.

Overall, much broader cross-peaks in the spectra for the encapsulated CAII and TTR are observed due to the intrinsic line-broadening effect obtained for a protein encapsulated inside a cell. These are a) a crowded and viscous environment that causes molecules to diffuse slowly, leading to increased transverse relaxation rates (R_2), which in turn causes line-broadening of the signal; b) magnetic field inhomogeneities resulting from the internal heterogenous structure of the RBCs; and c) the presence of paramagnetic species that can enhance relaxation rates, such as iron in the hemoglobin [20,21].

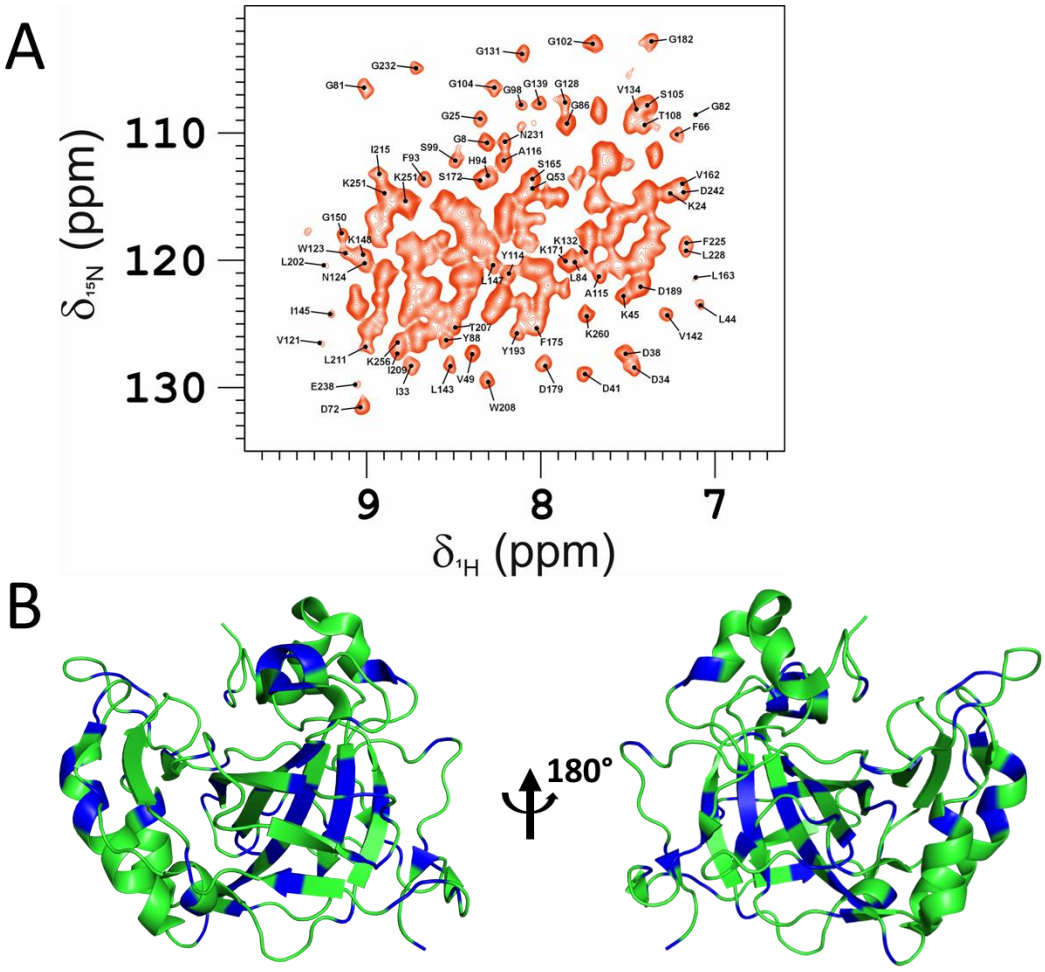


Figure 2. *A) NMR backbone resonance re-assignment of CAII encapsulated in Red Blood Cells obtained by comparison with the assigned ^{15}N HSQC spectra of free CAII with the sofast-HMQC spectra acquired on the CAII protein encapsulated in Red Blood Cells. Experiments were acquired at a spectrometer operating at 1200 MHz (^{1}H Larmour Frequency), at 310 K. B) X-ray structure of CAII (PDB code: 3KS3), with residues for which the backbone resonances have been reassigned inside the Red Blood Cells by superposition highlighted in blue*.

In order to confirm the encapsulation of the CAII inside the RBC, two different control spectra were acquired. Control 1, for a sample of RBC that underwent the dialysis protocol for encapsulation described in the methodology, but in the absence of the [U-^{15}N] CAII, and Control 2, for a sample of RBC in the presence of the [U-^{15}N] CAII which was not subjected to hypotonic dialysis. The ^{1}H-^{15}N sofast-HMQC spectra of the control confirm the absence

of the protein and the NMR signal from the 1st FID of the ^{1}H-^{15}N sofast-HMQC confirms the encapsulation of the protein with the protocol used (Figure 3). Background signal from natural abundance of ^{15}N-hemoglobin is observed given the high concentration of this protein (~5 mM in intact erythrocytes).

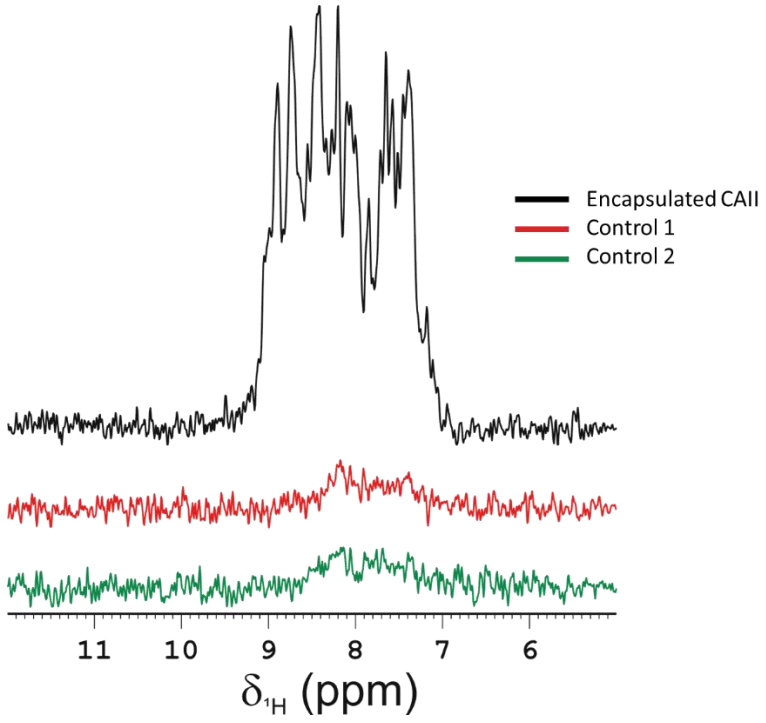


Figure 3. *NMR signal from the 1st FID of the so-fast HMQC of encapsulated ^{15}N CAII (black), control 1 (red) and control 2 (green). The spectra were acquired at 310 K at a spectrometer operating at 1200 MHz (^{1}H Larmour frequency).*

The ^{1}H-^{15}N sofast HMQC spectrum acquired on the erythrocyte-encapsulated [U-^{2}H,^{13}C,^{15}N] ANSII exhibits particularly clear, well-resolved signals, with a line-width comparable to that obtained on the spectrum of the free protein (Figure 4). Upon deuteration of the protein, the signal-to-noise in the NMR spectra is significantly improved by reducing spin diffusion and decreasing relaxation rates. Hence, the quality of spectra was significantly improved compared to the non-deuterated CAII and TTR. The remarkably resolved spectra of the encapsulated ANSII, not only confirms that the HOS of the protein is preserved within the erythrocyte environment, but also enable a semiquantitative estimation of the

encapsulated enzyme. Comparison of 1D ^{1}H, ^{15}N-filtered NMR experiments of free and encapsulated protein, allowed an estimation of the ANSII encapsulated within the red blood cells in ~25 µM, which corresponds to an encapsulation efficiency of ~10 %.

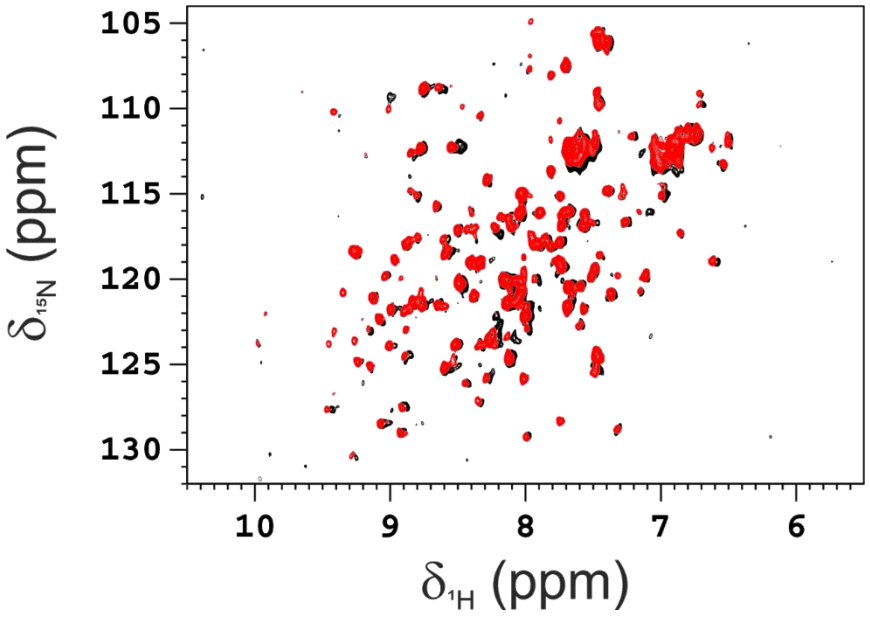


Figure 4. *^{1}H-^{15}N sofast-HMQC NMR spectra of ANSII encapsulated in RBC (black) and free ANSII (red). Experiments were acquired at a spectrometer operating at 900 MHz (^{1}H Larmour Frequency), at 310 K.*

Discussion

The preservation of the high order structure (HOS) of therapeutic proteins encapsulated in red blood cells (RBCs) is critical to ensure their functionality and stability within the pharmaceutical formulation. In this study, we applied solution NMR spectroscopy to investigate the structural integrity of three model proteins (Carbonic Anhydrase II (CAII), Transthyretin (TTR) and Asparaginase II (ANSII)), encapsulated in RBCs. The findings indicate that the proteins preserve their HOS within the erythrocyte environment without significant alterations.

CAII served as an optimal model for the initial optimization of the encapsulation protocol due to its stability and high expression levels. The protein maintained its native conformation upon encapsulation, as evidenced by the superposition of the resonances on the ^{1}H-^{15}N sofast HMQC NMR spectra of the free and encapsulated proteins. The most resolved resonances on the sofast-HMQC spectra of the encapsulated protein that were easily reassigned are located on all the different secondary structural motifs of the protein, which serves as a further confirmation of the preservation of the HOS. TTR, a larger protein with a complex quaternary structure, also retained its structural integrity upon encapsulation, as shown by the ^{1}H-^{15}N sofast-HMQC spectra. Given the size of the protein, the line broadening and quality of the spectra suggest the preservation of its quaternary structure. This is of uttermost importance, because maintaining the quaternary structure of TTR is crucial for its function as a carrier. These results demonstrate the preservation of HOS upon encapsulation in RBCs for structurally more complex proteins.

Finally, we demonstrated the preservation of the HOS of ANSII upon erythrocyte encapsulation, which is critical to ensure its enzymatic function and therapeutic potential. The ^{1}H-^{15}N sofast-HMQC spectra of the free and encapsulated ANSII are superimposable, with highly resolved resonances even for the protein in the complex matrix of erythrocytes. This aspect is attributed to the deuteration of the protein sample, that improves the resolution and sensitivity of the NMR spectra by minimizing spin diffusion and decreasing relaxation rates. This study represents the first structural characterization of a therapeutic biomolecule encapsulated within red blood cells. This is of significant importance, given the rapidly growing interest in the application of red blood cells as systems for drug delivery.

Conclusion

This study demonstrates the preservation of the HOS of different proteins upon erythrocyte encapsulation. By using three model proteins, we have demonstrated that erythrocyte encapsulation of such biomolecules does not perturb their higher order structure. This is critical for maintaining the therapeutic efficacy of biologics. This

methodology provides valuable insights for assessing the HOS of biologics in complex environments such as RBCs. The results obtained in this study support the potentiality of red blood cells as effective drug delivery systems for therapeutic proteins, while maintaining their structural and functional integrity, offering a reliable approach for the development and quality control of encapsulated biologics in pharmaceutical formulations.

Methods

Expression and purification of uniformly isotopically enriched Carbonic Anhydrase II [U-^{15}N]

The gene encoding α-CAII into the pCAM vector was transformed into *E.coli* BL21(DE3) cells, which were subsequently pre-cultured overnight in LB medium containing ampicillin (0.1 mg/mL) and 1 % glucose at 37 ℃ and at 160 rpm. The main culture is then incubated at 37 ℃, 160 rpm until reaches the OD_{600}=0.6 and harvested for 15 minutes at 4000 rpm, in order to proceed with the isotopic enrichment according to Marley method. The cell pellet is resuspended in 1L M9 medium supplemented with 1mM $MgSO_4$, 0.3 mM $CaCl_2$, 3 g glucose, 1.2 g ^{15}N-Amonium sulfate, 0.5 mM $ZnSO_4$ and 0.1 mg / mL ampicillin. The new culture is further incubated for 30 minutes at 37 ℃ and at 160 rpm, before inducing with 1 mM IPTG and incubating overnight at 25 ℃ and 160 rpm. The cell culture was harvested at 7500 rpm for 20 minutes. The cell pellet is resuspended in 70 mL of 20 mM Tris-SO_4, 500 µM $ZnSO_4$ pH 8 and sonicated for 30 seconds with a resting period of 3 minutes on ice, 60 % power, 10 cycles. The lysate is ultracentrifuged at 40 000 rpm for 40 minutes and the supernatant recovered is filtered with a 0.45 µm filter. The supernatant, containing crude α-CAII protein, was purified by Ni^{2+} affinity chromatography using a linear 0-0.5 M imidazole gradient on a HisTrap 5 mL (GE Healthcare). Finally, the α-CAII was subjected to size-exclusion chromatography on a HiLoad 26 / 60 Superdex 75 pg column (Amersham Biosciences) in 50 mM sodium phosphate buffer at pH 7.

Expression and purification of uniformly isotopically enriched TTR [U-^{15}N].

E. coli BL21(DE3) RIPL cells were transformed with pET28a(+) plasmid coding for TTR gene. The cells were cultured in enriched M9 minimal medium and further grown at 37 ℃, until OD_{600} reached 0.6-0.8 for the induction with 1 mM IPTG. The cells were incubated overnight at 37 ℃ and then harvested by centrifugation at 7500 rpm (JA-10 Beckman Coulter) for 15 min at 4 ℃. The pellet was suspended in 20 mM Tris-HCl, pH 8.6, buffer supplied with 5 mM DTT and sonicated at 4 ℃ for 10 cycles of 30 seconds ON and 3 minutes OFF, at 60 %

power. The suspension was centrifuged at 40 000 rpm (F15-6x100y Thermo Scientific) for 40 minutes and the pellet discarded. The protein was purified by anionic-exchange chromatography using a HiPrep Q FF 16 / 10 column (GE Healthcare Life Science), previously equilibrated with 20 mM Tris-HCl, pH 8.6. The protein was eluted in 20 mM Tris-HCl, pH 8.6 with a 0-1 M NaCl linear gradient. Fractions containing pure TTR were joined and purified by size exclusion chromatography using HiLoad 26 / 60 Superdex 75 pg in 50 mM phosphate buffer, pH 7.5.

Expression and purification of uniformly isotopically enriched L-Asparaginase II [U-^{2}H,^{13}C,^{15}N].

E. coli (DE3) C41 cells were transformed with pET21a(+) plasmid with ANSII insert. The cells were cultured in ^{2}D-^{13}C-^{15}N-enriched medium (*E. coli*-OD2 rich growth media) supplemented with ampicillin, and grown at 37 ℃, until OD_{600} reached 0.6, then further induced with 1 mM IPTG. All reagents were previously dissolved in D_2O. The culture was incubated at 37 ℃ overnight and then harvested by centrifugation at 6 500 rpm (JA-10 Beckman Coulter) for 15 minutes at 4 ℃. The pellet was suspended in 10 mM Tris-HCl, pH 8.0, 15 mM EDTA, 20 % sucrose buffer and incubated at 4 ℃ for 20 minutes under magnetic stirring. The suspension was centrifuged at 10 000 rpm (F15-6x100y Thermo Scientific) for 30 minutes and the supernatant discarded. The recovered pellet was re-suspended in milli-Q H_2O and newly incubated at 4 ℃ for 20 minutes under magnetic stirring. The suspension was centrifuged at 10 000 rpm (F15-6x100y Thermo Scientific) for 30 minutes and the pellet was discarded, whereas the supernatant was treated with ammonium sulfate. Still under magnetic stirring, solid ammonium sulfate was added in aliquots up to 50 % saturation. The precipitate was removed by centrifugation, then more ammonium sulfate was added up to 90 % saturation to trigger the precipitation of ANSII. The precipitated ANSII was re-dissolved in a minimal amount of 20 mM Tris-HCl, pH 8.6 and dialyzed extensively against the same buffer. ANSII was purified by anionic-exchange chromatography using a HiPrep 16 / 10 Q FF column (GE Healthcare Life Science). The protein was eluted in 20 mM Tris-HCl, pH 8.6 with

a 0-1 M NaCl linear gradient. Fractions containing pure ANSII were identified by Coomassie staining SDS-PAGE gels, then joined and dialyzed extensively against final buffer.

Washing of packed Red Blood Cells.

Commercially available Bovine Red Blood Cells Packed 100% from Innovative Research were used for encapsulation. RBCs were washed by diluting 5 mL of 100% packed RBC in 5 mL of cold PBS. The sample was centrifuged at 1200 g for 1 min at 4 °C and the supernatant removed. The process was repeated two more times.

Encapsulation of ^{15}N-labeled CAII in Red Blood Cells.

CAII, TTR and ANSII were encapsulated using a previously reported hypotonic dialysis method of encapsulation[22]. 7 parts (700 µL) of washed packed RBC were mixed with 3 parts (300 µL) of a concentrated solution of the protein (3 mM solution of CAII, 1 mM solution of TTR and 0.9 mM solution of ANSII). 1 mL of cell suspension was placed on a dialysis tube (12 kDa MWCO). The sample was dialyzed against 150 mL of the hypotonic buffer (5 mM KH_2PO_4, 5 mM K_2HPO_4, pH 7.4) at 4 °C with a rotation of ~20 rev/min for 180 minutes. The erythrocytes were then resealed by transferring the dialysis tube into a container holding 150 mL of the isotonic buffer (PBS 1X, 5 mM glucose, 5 mM $MgCl_2$) at 37 °C under continuous stirring for 60 minutes. In order to wash away non-encapsulated ^{15}N isotopically enriched protein or released components from the erythrocytes, the sample was transferred into a dialysis tube of 1000 kDa MWCO and dialyzed against 1L of isotonic buffer at 4 °C for 12 hours. This dialysis was repeated 2 times. The dialyzed erythrocytes were then washed with an equal volume of cold PBS three times. A 1D ^{15}N sofast-HMQC spectra of a sample from the last wash was acquired in order to confirm the absence of any ^{15}N-labeled protein outside of the Red Blood Cells.

Control experiments

Two different control experiments were prepared to rule out the possibility of having the protein absorbed on the surface of the Red Blood Cells and to confirm the efficacy of the washing procedure. Control 1, in which the protein was added but not encapsulated, was

prepared by mixing 7 parts of washed packed RBC with 3 parts of a 3 mM solution of the ^{15}N-labeled CAII. 1 mL of cell suspension was placed on a dialysis tube (12 kDa MWCO). The sample was dialyzed against 150 mL of isotonic buffer at 4 °C with a rotation of ~20 rev/min for 180 minutes. The sample was then dialyzed against the same isotonic buffer at 37 °C and continue stirring for 60 minutes. Control 2, in which the erythrocytes were subjected to hypotonic dialysis in the absence of the protein, was prepared by mixing 7 parts of washed packed RBC with 3 parts of cold PBS. 1 mL of cell suspension was placed on a dialysis tube (12 kDa MWCO). The sample was dialyzed against 150 mL of hypotonic buffer at 4 °C with a rotation of around ~20 rev/min for 180 minutes. The dialysis tube was then transferred to a container holding 150 mL of isotonic buffer at 37 °C and continue stirring for 60 minutes. Both controls were washed as described in the section of 'Encapsulation of ^{15}N-labeled CAII in Red Blood Cells'. Both control experiments were performed in parallel with the experiment of encapsulation of ^{15}N-labeled CAII.

Sample preparation and NMR measurements.

NMR experiments were recorded on a Bruker Avance III spectrometer operating at 900 MHz ^{1}H Larmor frequency (28.2 T) at 310 K.

The NMR samples were prepared with 540 µL of RBC suspension and 60 µL of PBS 1X in D_2O for field frequency lock purposes. sofast-HMQC experiments were acquired with an H_N offset of 8.1 ppm and an excitation H_N bandwidth of 4 ppm. NS = 256. All spectra were processed on a Bruker TopSpin 3.4 software package.

References

(1) Schiffmann, R.; Kopp, J. B.; Austin Iii, H. A.; Sabnis, S.; Moore, D. F.; Weibel, T.; Balow, J. E.; Brady, R. O. Enzyme Replacement Therapy in Fabry Disease: A Randomized Controlled Trial. *JAMA* **2001**, *285* (21), 2743. https://doi.org/10.1001/jama.285.21.2743.

(2) Kakkis, E. D.; Muenzer, J.; Tiller, G. E.; Waber, L.; Belmont, J.; Passage, M.; Izykowski, B.; Phillips, J.; Doroshow, R.; Walot, I.; Hoft, R.; Yu, K. T.; Okazaki, S.; Lewis, D.; Lachman, R.; Thompson, J. N.; Neufeld, E. F. Enzyme-Replacement Therapy in Mucopolysaccharidosis I. *N. Engl. J. Med.* **2001**, *344* (3), 182–188. https://doi.org/10.1056/NEJM200101183440304.

(3) Veronese, F. M.; Mero, A. The Impact of PEGylation on Biological Therapies: *BioDrugs* **2008**, *22* (5), 315–329. https://doi.org/10.2165/00063030-200822050-00004.

(4) Cerofolini, L.; Giuntini, S.; Carlon, A.; Ravera, E.; Calderone, V.; Fragai, M.; Parigi, G.; Luchinat, C. Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. *Chem. – Eur. J.* **2019**, *25* (8), 1984–1991. https://doi.org/10.1002/chem.201804488.

(5) Sellaturay, P.; Nasser, S.; Islam, S.; Gurugama, P.; Ewan, P. W. Polyethylene Glycol (PEG) Is a Cause of Anaphylaxis to the Pfizer/BioNTech mRNA COVID-19 Vaccine. *Clin. Exp. Allergy* **2021**, *51* (6), 861–863. https://doi.org/10.1111/cea.13874.

(6) Lee, S. H.; Cha, J. M.; Lee, J. I.; Joo, K. R.; Shin, H. P.; Baek, I. H.; Jeon, J. W.; Lim, J. U.; Lee, J. L.; Lee, H. M.; Cho, Y.-H. Anaphylactic Shock Caused by Ingestion of Polyethylene Glycol. *Intest. Res.* **2015**, *13* (1), 90. https://doi.org/10.5217/ir.2015.13.1.90.

(7) Ihler, G. M.; Glew, R. H.; Schnure, F. W. Enzyme Loading of Erythrocytes. *Proc. Natl. Acad. Sci.* **1973**, *70* (9), 2663–2666. https://doi.org/10.1073/pnas.70.9.2663.

(8) Koleva, L.; Bovt, E.; Ataullakhanov, F.; Sinauridze, E. Erythrocytes as Carriers: From Drug Delivery to Biosensors. *Pharmaceutics* **2020**, *12* (3), 276. https://doi.org/10.3390/pharmaceutics12030276.

(9) Dolowy, W. C.; Henson, D.; Cornet, J.; Sellin, H. Toxic and Antineoplastic Effects of L-Asparaginase: Study of Mice with Lymphoma and Normal Monkeys and Report on a Child with Leukemia. *Cancer* **1966**, *19* (12), 1813–1819. https://doi.org/10.1002/1097-0142(196612)19:12<1813::AID-CNCR2820191208>3.0.CO;2-E.

(10) Hill, J. M. L-Asparaginase Therapy for Leukemia and Other Malignant Neoplasms: Remission in Human Leukemia. *JAMA* **1967**, *202* (9), 882. https://doi.org/10.1001/jama.1967.03130220070012.

(11) Bertrand, Y.; Thomas, X.; Baruchel, A.; Mazingue, F.; Auvrignon, A.; Corm, S.; Dombret, H.; Chevallier, P.; Galambrun, C.; Huguet, F.; Legrand, F.; Mechinaud, F.; Vey, N.; Chabaud, S.; Baron, F. N.; Philip, I.; Liens, D.; Godfrin, Y. GRASPALL 2005.01 Clinical Study: L-Asparaginase Loaded into Red Blood Cells Is Effective at Depleting Serum Asparagine in Children and Adults with Relapsed Acute Lymphoblastic Leukaemia (ALL). *Blood* **2008**, *112* (11), 306–306. https://doi.org/10.1182/blood.V112.11.306.306.

(12) Domenech, C.; Thomas, X.; Chabaud, S.; Baruchel, A.; Gueyffier, F.; Mazingue, F.; Auvrignon, A.; Corm, S.; Dombret, H.; Chevallier, P.; Galambrun, C.; Huguet, F.; Legrand, F.; Mechinaud, F.; Vey, N.; Philip, I.; Liens, D.; Godfrin, Y.; Rigal, D.; Bertrand, Y. L -asparaginase Loaded Red Blood Cells in Refractory or Relapsing Acute Lymphoblastic Leukaemia in

Children and Adults: Results of the GRASPALL 2005-01 Randomized Trial. *Br. J. Haematol.* **2011**, *153* (1), 58–65. https://doi.org/10.1111/j.1365-2141.2011.08588.x.
(13) Thomas, X.; Le Jeune, C. Erythrocyte Encapsulated L -Asparaginase (Graspa) in Acute Leukemia. *Int. J. Hematol. Oncol.* **2016**, *5* (1), 11–25. https://doi.org/10.2217/ijh-2016-0002.
(14) Zarrin, A.; Foroozesh, M.; Hamidi, M. Carrier Erythrocytes: Recent Advances, Present Status, Current Trends and Future Horizons. *Expert Opin. Drug Deliv.* **2014**, *11* (3), 433–447. https://doi.org/10.1517/17425247.2014.880422.
(15) Herbert, J.; Wilcox, J. N.; Pham, K.-T. C.; Fremeau, R. T.; Zeviani, M.; Dwork, A.; Soprano, D. R.; Makover, A.; Goodman, D. S.; Zimmerman, E. A.; Roberts, J. L.; Schon, E. A. Transthyretin: A Choroid Plexus-specific Transport Protein in Human Brain: The 1986 S. Weir Mitchell Award. *Neurology* **1986**, *36* (7), 900–900. https://doi.org/10.1212/WNL.36.7.900.
(16) Hamilton, J. A.; Benson, M. D. Transthyretin: A Review from a Structural Perspective: *Cell. Mol. Life Sci.* **2001**, *58* (10), 1491–1521. https://doi.org/10.1007/PL00000791.
(17) Walker, K. W.; Foltz, I. N.; Wang, T.; Salimi-Moosavi, H.; Bailis, J. M.; Lee, F.; An, P.; Smith, S.; Bruno, R.; Wang, Z. The Serum Protein Transthyretin as a Platform for Dimerization and Tetramerization of Antibodies and Fab Fragments to Enable Target Clustering. *J. Biol. Chem.* **2020**, *295* (30), 10446–10455. https://doi.org/10.1074/jbc.RA120.013135.
(18) Cerofolini, L.; Vasa, K.; Bianconi, E.; Salobehaj, M.; Cappelli, G.; Bonciani, A.; Licciardi, G.; Pérez-Ràfols, A.; Padilla-Cortés, L.; Antonacci, S.; Rizzo, D.; Ravera, E.; Viglianisi, C.; Calderone, V.; Parigi, G.; Luchinat, C.; Macchiarulo, A.; Menichetti, S.; Fragai, M. Combining Solid-State NMR with Structural and Biophysical Techniques to Design Challenging Protein-Drug Conjugates. *Angew. Chem. Int. Ed.* **2023**, *62* (31), e202303202. https://doi.org/10.1002/anie.202303202.
(19) Wang, H.; Wang, L.; Li, C.; Wuxiao, Z.; Chen, G.; Luo, W.; Lu, Y. Pegaspargase Combined with Concurrent Radiotherapy for Early-Stage Extranodal Natural Killer/T-Cell Lymphoma, Nasal Type: A Two-Center Phase II Study. *The Oncologist* **2020**, *25* (11), e1725–e1731. https://doi.org/10.1634/theoncologist.2020-0144.
(20) Luchinat, E.; Banci, L. A Unique Tool for Cellular Structural Biology: In-Cell NMR. *J. Biol. Chem.* **2016**, *291* (8), 3776–3784. https://doi.org/10.1074/jbc.R115.643247.
(21) Luchinat, E.; Banci, L. In-Cell NMR: A Topical Review. *IUCrJ* **2017**, *4* (2), 108–118. https://doi.org/10.1107/S2052252516020622) Bax, B. E.; Bain, M. D.; Talbot, P. J.; John Parker-Williams, E.; Chalmers, R. A. Survival of Human Carrier Erythrocytes in Vivo. *Clin. Sci.* **1999**, *96* (2), 171–178. https://doi.org/10.1042/cs0960171.

Combining Solid-State NMR with Structural and Biophysical Techniques to Design Challenging Protein-Drug Conjugates

How to cite: *Angew. Chem. Int. Ed.* **2023**, *62*, e202303202
doi.org/10.1002/anie.202303202

NMR Spectroscopy

Combining Solid-State NMR with Structural and Biophysical Techniques to Design Challenging Protein-Drug Conjugates

Linda Cerofolini[+], Kristian Vasa[+], Elisa Bianconi[+], Maria Salobehaj, Giulia Cappelli, Alice Bonciani, Giulia Licciardi, Anna Pérez-Ràfols, Luis Padilla-Cortés, Sabrina Antonacci, Domenico Rizzo, Enrico Ravera, Caterina Viglianisi, Vito Calderone, Giacomo Parigi, Claudio Luchinat, Antonio Macchiarulo, Stefano Menichetti, and Marco Fragai**

Abstract: Several protein-drug conjugates are currently being used in cancer therapy. These conjugates rely on cytotoxic organic compounds that are covalently attached to the carrier proteins or that interact with them via non-covalent interactions. Human transthyretin (TTR), a physiological protein, has already been identified as a possible carrier protein for the delivery of cytotoxic drugs. Here we show the structure-guided development of a new stable cytotoxic molecule based on a known strong binder of TTR and a well-established anticancer drug. This example is used to demonstrate the importance of the integration of multiple biophysical and structural techniques, encompassing microscale thermophoresis, X-ray crystallography and NMR. In particular, we show that solid-state NMR has the ability to reveal effects caused by ligand binding which are more easily relatable to structural and dynamical alterations that impact the stability of macromolecular complexes.

Introduction

The development of a suitable drug delivery system is a crucial step in drug design to ensure extended half-lives and efficient targeting, thus achieving high therapeutic efficacy.[1,2] The conjugation of small organic drugs to protein-based biomaterials or synthetic polymers is often used to decrease renal excretion and increase their half-lives.[3] When a protein-based carrier is used, specific recognition of receptors and thus efficient targeting can be also achieved. Human serum albumin is an archetypical example of a protein used as a carrier of drugs and contrast agents.[4] It is also used as a component of nanoparticles to deliver cytotoxic molecules to cancer cells,[5,6] or fused with therapeutic peptides to prevent their fast clearance, proteolytic degradation, and to improve solubility.[7–9]

Another protein that has been considered as a drug carrier is human transthyretin (TTR hereafter), which is present in blood plasma and cerebrospinal fluid where it carries the holo-retinol binding protein and the thyroxine T_4 hormone. Functional TTR is an assembly of four identical subunits, a dimer of dimers (D_2 symmetry, total molecular mass 55 kDa). Mutations in the gene coding for TTR decrease the stability of the assembly, leading to the dissociation of the tetramer into monomers. Monomers can partially unfold and form amyloid fibrils.[10] Extracellular accumulation of TTR amyloid fibrils in different tissues and organs leads to severe disorders, and ultimately to fatal multiorgan failure. A recent therapeutic approach to treat

[*] Dr. L. Cerofolini,[+] M. Salobehaj, G. Licciardi, Dr. A. Pérez-Ràfols, L. Padilla-Cortés, S. Antonacci, Dr. D. Rizzo, Prof. Dr. E. Ravera, Prof. Dr. V. Calderone, Prof. Dr. G. Parigi, Prof. C. Luchinat, Prof. Dr. M. Fragai
Magnetic Resonance Centre (CERM), University of Florence
Via L. Sacconi 6, 50019 Sesto Fiorentino (Italy)
E-mail: fragai@cerm.unifi.it

Dr. L. Cerofolini,[+] M. Salobehaj, G. Licciardi, L. Padilla-Cortés, Dr. D. Rizzo, Prof. Dr. E. Ravera, Prof. Dr. V. Calderone, Prof. Dr. G. Parigi, Prof. C. Luchinat, Prof. Dr. M. Fragai
Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine (CIRMMP)
Via L. Sacconi 6, 50019 Sesto Fiorentino (Italy)

Dr. L. Cerofolini,[+] K. Vasa,[+] M. Salobehaj, G. Cappelli, A. Bonciani, G. Licciardi, L. Padilla-Cortés, S. Antonacci, Dr. D. Rizzo, Prof. Dr. E. Ravera, Prof. Dr. C. Viglianisi, Prof. Dr. V. Calderone, Prof. Dr. G. Parigi, Prof. C. Luchinat, Prof. S. Menichetti, Prof. Dr. M. Fragai
Department of Chemistry "Ugo Schiff", University of Florence
Via della Lastruccia 3-13, 50019 Sesto Fiorentino (Italy)
E-mail: stefano.menichetti@unifi.it

E. Bianconi,[+] Prof. Dr. A. Macchiarulo
Department of Pharmaceutical Sciences, University of Perugia
Via Fabretti n.48, 06123 Perugia (Italy)

Dr. A. Pérez-Ràfols, Prof. C. Luchinat
Giotto Biotech s.r.l, Sesto Fiorentino
Via della Madonna del Piano 6, 50019 Florence (Italy)

[+] These authors contributed equally to this work.

the familial amyloid polyneuropathy relies on small organic molecules that target thyroxine T_4 hormone binding sites.[11,12] One of these molecules, Tafamidis (2-(3,5-dichlorophenyl)benzo[*d*]oxazole-6-carboxylic acid), binds TTR at these two sites in a negative cooperative manner, but with nanomolar affinity (K_ds of ≈2 nM and ≈200 nM respectively), and prevents dissociation into monomers.[13] The high affinity of Tafamidis makes it an ideal anchor in the design of conjugates harboring hydrophobic cytotoxic drugs that would otherwise result poorly soluble. This strategy, but with a different high-affinity TTR ligand, has been already used to generate drug-protein conjugates with an improved selectivity against cancer cells.[14–16]

The rational design of protein-drug conjugates to maximize effectiveness, pharmacokinetics, and stability in vivo while minimizing their structural complexity is receiving more and more interest, and could benefit from highly accurate structural information from X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy.[17–26] However, proteins that are covalently bound to, or that interact strongly with, relatively large drugs through long linkers can be difficult to crystallize. Furthermore, these systems are too big for NMR spectroscopy in solution, but still neither big nor rigid enough to allow for the use of cryo-electron microscopy.[27] Solid-state NMR may overcome these limitations, and it is already used to investigate non-crystalline protein samples, biologics and biomaterials.[28–39] Significant enhancements in sensitivity have been obtained by the recent achievements in the NMR probe technology and in biomolecular Dynamic Nuclear Polarization (DNP).[40–44]

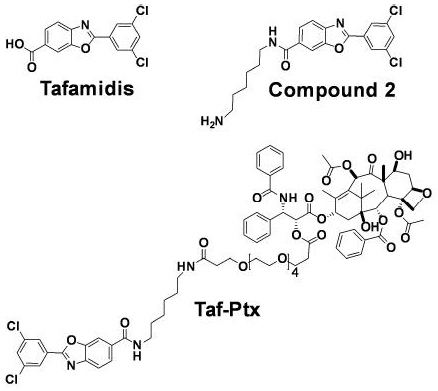
We here report the design and synthesis of a new molecule that results from the conjugation of the cytotoxic Paclitaxel with Tafamidis (Scheme 1) to form a stable non-covalent protein-drug conjugate (PDC) with TTR. The design of this molecule starts with the combination of structural data from X-ray crystallography, and solution and solid-state NMR. Paclitaxel and Tafamidis are linked through a long PEGylated linker containing an easily hydrolysable ester bond for the release of the cytotoxic agent to its pharmacological target (Taf-Ptx hereafter, Scheme 1).

The approach used for the design of this molecule, which is aimed at maintaining a high affinity for TTR, demonstrates the strength of an integrated structural biology strategy (relying upon X-ray, solution and solid-state NMR) in the structure-based design of novel PDCs.

Results and Discussion

Tafamidis was chosen to anchor Paclitaxel to TTR due to its high affinity and because it is an approved drug with minimal side effects.[45] A relaxometry analysis was carried out to investigate the stability of the tetrameric protein and the overall steadiness of protein dynamics upon ligand binding.[30,46–48] ^{1}H NMRD profiles of 1.2 mM wild-type TTR in water solutions were acquired with and without Tafamidis (shown in Figure S1). Multiple correlation times should be considered to account for many motional processes of different water protons interacting with protein (see "NMR relaxometry measurements" in the Supporting Information).[47] The analysis of the profiles indicates the major contributions from a correlation time in agreement with the overall reorientation time expected for tetrameric TTR, as calculated with HydroNMR,[49] and from a faster correlation time of few nanoseconds. In summary, the profiles indicate that i) the reorientation time is in line with that expected for a tetrameric protein assembly, ii) there are extensive internal motions, and iii) no sizable changes in the overall dynamics of the protein occur upon ligand binding.

The X-ray structure of the Tafamidis-TTR complex (PDB code: 3TCT)[13] shows the binding mode of Tafamidis (see Figure 1 and Figure S2, S3), suggesting that functionalization of the molecule on the carboxylic acid in position 6 should not affect its interaction with TTR. Therefore, a derivative of Tafamidis bearing a six carbon atoms linker in position 6 (Compound **2**) was synthesized (see Schemes 1 and S1) as a precursor of Taf-Ptx and used to soak crystals



Scheme 1. Structures of Tafamidis, Compound **2**, and Taf-Ptx.

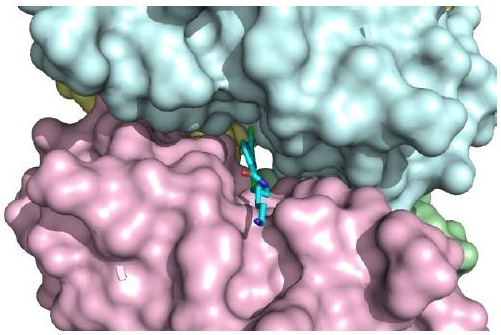


Figure 1. Detail of the surface representation of TTR interacting with compound **2** displayed as stick presented in this work (8AWW).

of TTR. The resulting X-ray structure shows that Compound **2** and Tafamidis bind TTR very similarly, and that the structure of the protein part is maintained. However, while the electron density is very well defined for all the aromatic rings in the structure of Tafamidis-bound TTR (3TCT), the Tafamidis ring that was functionalized with the hydrocarbon chain shows a weaker and less defined electron density in the structure of the adduct with Compound **2**, and this suggests a slightly higher ligand-binding heterogeneity, which in turn could be linked to a slightly lower affinity. Furthermore, the long hydrocarbon chain shows only very faint and scattered patches. This is certainly due to the mobility of this chain that is completely leaning towards the solvent. The interactions between the protein and Compound **2** are shown in Figure S4. Summarizing, the X-ray structure proves that Compound **2** preserves the binding pose of Tafamidis with the linker sticking-out from the central channel toward the solvent (see Figure 1), supporting the correct design strategy for Taf-Ptx. The X-ray structure has been deposited in the Protein Data Bank under the accession code 8AWW.

Then, the Paclitaxel molecule was conjugated to Compound **2** through an additional nineteen-atoms PEG spacer to minimize any possible steric clash with the protein and to increase the solubility. Paclitaxel is connected to the PEG spacer through an ester linkage labile in vivo,[50] which ensures its release into the cell allowing for the inhibition of tubulin polymerization.[51] As expected Taf-Ptx (Schemes 1 and S1), fails to enter the TTR crystals by soaking. Likewise, co-crystallization with TTR fails. Apparently, Taf-Ptx is too bulky, and X-ray characterization of the complex cannot be carried out. As anticipated, in cases like this, NMR could provide precious information on the structural features of the PDC, and thus validate the whole design strategy.

NMR Analysis and Assignment of Free TTR in Solution

The assignment of free tetrameric TTR in solution was obtained comparing the assignments available in the literature for the monomeric and tetrameric states of the protein,[52–55] and analyzing triple resonance spectra recorded on the perdeuterated sample of TTR. All residues (but the N-terminus, Gly-1 and Asn-98) were assigned in the spectra (Figure S5). The present assignment is the most complete and has been deposited in the bmrb under the accession code 51818.

The number of cross-peaks present in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum, and the absence of signal splitting, are both consistent with the preservation of the D_2 symmetry of TTR in the tetrameric assembly. Interestingly, the signals within the same 2D spectra are characterized by different line broadening. Specifically, sharp and intense signals were observed for the residues forming loops and on the external surface of the tetramer, while broad signals were observed for the residues at the interfaces between the monomers (*i.e.*, Cys10-Lys15; His91-Phe95; Tyr105-Val121). In the 2D ^{1}H-^{15}N TROSY-HSQC spectrum recorded on the deuterated sample, these signals are also broad, suggesting the occurrence of a conformational heterogeneity/exchange for the protein (Figure S6). These features also affect the quality of the 3D ^{1}H-^{15}N NOESY spectrum, where only few NOE correlations are visible (Figure S7). As expected for a folded protein of 55 kDa, in the 2D ^{13}C-^{15}N CON spectrum only signals of the flexible regions, which are not well-defined in most X-ray structures, can be observed (Thr3-Cy10; Ala37-Thr40; Glu51-Ser52; Asn124-Glu127; see Figure S8).

NMR Analysis and Assignment of Free TTR in the Solid-State

The solid-state NMR spectra of re-hydrated freeze-dried tetrameric TTR are of good quality and characterized by a good signal resolution (Figure S9). Nevertheless, around 20% of the expected resonances are missing and some signals are characterized by large line-broadening. Assignment of the free tetrameric protein was also obtained in the solid-state (Figure S10). The available assignment of the free protein in solution was used as starting point and complemented by the analysis of carbon-detected spectra acquired in the solid-state. The residues whose signals are missing in the spectra are mainly located at the N-terminus (up to Lys15) and in flexible regions: Asp38-Thr40, Gly57-Leu58, Phe64, Ser117-Thr119, Thr123-Asn124, Lys126. Nevertheless, as much as 80% of the spin systems of the protein could be reassigned in the solid-state NMR spectra.

NMR Analysis and Assignment of TTR-Tafamidis and TTR-Taf-Ptx in Solution

The binding-mode of Tafamidis to [U-^{13}C, ^{15}N] TTR was first analyzed by solution NMR. During the NMR titration, the cross-peaks of the free protein in the 2D ^{1}H-^{15}N TROSY-HSQC spectra decrease in intensity upon the addition of increasing concentrations of the ligand, while new cross-peaks, corresponding to the complex between TTR and Tafamidis, appear and increase in intensity. This behavior indicates that the ligand is in slow exchange regime on the NMR timescale, and confirms its expected high affinity towards the protein.[13,56] In the presence of Tafamidis at 100 μM concentration (ligand:tetramer ratio equal to 0.5:1) the cross-peaks corresponding to the free protein and to the protein bound to the ligand have similar intensities (Figure S11A), as it can be seen for the signal of Ser112 located at the interface between the dimers of the tetrameric assembly (PDB code: 3TCT).[13] During the titration the quality of spectra decreases up to a 1:1 ligand:tetramer ratio; some signals are broadened and some disappeared. The signal's line-width sharpens again at 2:1 ligand:tetramer ratio (Figure S12), in line with the lower affinity for the second binding event. The analysis of the chemical shift perturbation (CSP) was thus performed with a ligand in slight excess with respect to the 2:1 ratio. Figure S12 confirmed that the residues experiencing the largest changes (Lys15, Leu17, Ala19, Val20, Arg21, Gly22, Ser23, Ile26, Gly53, His88, Val94, Tyr105, Thr106, Ile107, Leu111,

Ser112, Ala120, Val122) are in the expected Tafamidis binding site (Figure S13). Some of the signals experiencing large perturbations (Ala19, Arg21, Gly22, Leu111) have been tentatively reassigned with some uncertainty (Figure S13A). Interestingly, the signals corresponding to residue Ser117 and Thr118, which are almost missing in the spectra of the free protein, appear with increased intensity in the spectrum of TTR in the presence of Tafamidis. This last observation is consistent with Tafamidis stabilizing the protein tetramer.

NMR titration in solution was also performed with the newly designed ligand, Taf-Ptx. The evolution of the spectra upon addition of increasing amounts of Taf-Ptx was superimposable to that previously observed for Tafamidis, with several protein resonances experiencing a slow exchange regime on the NMR timescale. After the addition of Taf-Ptx, in the presence of a ligand:tetramer ratio equal to 0.5:1, the signals of the free protein and those of the protein bound to the ligand have similar intensities (Figure S11B). As with Tafamidis, some protein signals broaden/disappear at 1:1 ligand:tetramer ratio and sharpen/reappear again in the 2:1 complex (Figure S12C and D). The analysis of CSP of the protein signals in the 2:1 complex (Figure S13B) confirmed that the residues experiencing the largest changes (Lys15, Val20, Arg21, Gly22, Ala25, Leu111, Ser112, Ala120, Val122) are in the same protein region affected by the presence of Tafamidis (Figure S13D). It is important to point out that the signals of some residues mostly affected by CSP in the spectrum of Tafamidis-TTR (Ala19, His88, Val94, Tyr105, T106) are broadened beyond detection in the spectrum of Taf-Ptx-TTR. This is probably related to the previously described slightly higher ligand-binding heterogeneity of Taf-Ptx-TTR with respect to that of Tafamidis-TTR.

NMR Analysis and Assignment of Tafamidis-TTR and Taf-Ptx-TTR in the Solid-State

Further interesting features of the complexes between TTR and Tafamidis or Taf-Ptx are revealed by solid-state NMR. The 2D solid-state NMR spectra of these complexes exhibit a higher number of cross-peaks with respect to those of the free protein. Specifically, in the 2D ^{13}C-^{13}C dipolar-assisted rotational resonance (DARR) spectrum of TTR in the presence of either Tafamidis or Taf-Ptx, several new signals appear or increase in intensity (Gly57, Arg103, Ile107, Ala108, Ser117, Thr118, Thr119, Ala120, Thr123; see Figure 2, panel A and B). This also occurs for a few signals in the 2D ^{15}N-^{13}C NCA spectra (Figure 3A and B). These signals belong to residues located at the tetramer interface, where Tafamidis binds (Figure 3, panels E and F). The increase in signal intensity can be explained by a higher rigidity of this region after the binding of Tafamidis, which is known to stabilize the tetrameric form of the protein. As shown by the high similarity of the spectra of TTR bound to one ligand or to the other, the high affinity of Tafamidis, and its stabilizing effect on the tetramer, are still present also when the Tafamidis unit is conjugated to the Paclitaxel unit. A lower intensity of the signals corresponding to only residues S117 and T118 is observed when TTR binds to Taf-Ptx compared to when it binds to Tafamidis.

The analysis of the CSP of TTR bound to Tafamidis or Taf-Ptx with respect to the free protein indicates that the signals influenced by the ligands are largely the same. Most of the signals experiencing the largest perturbations correspond to residues at the dimer/tetramer interface (Figure 3).

To rule out the possibility that the differences, observed in the analysis of CSPs performed for solution and solid-state NMR data, are due to the comparison of different nuclei (*i.e.* ^{1}H and ^{15}N in solution, and $^{13}C\alpha$ and ^{15}N in the solid-state), the ^{15}N chemical shifts were separately compared. The analysis confirms that the largest perturbations are observed in the same areas (Figure S14).

Microscale Thermophoresis for Ligand Binding Assay

The interaction between TTR and Taf-Ptx was analyzed using a ligand binding assay with microscale thermophoresis (MST). MST detects the migration of a macromolecule in a temperature gradient, which strongly depends on size, charge, conformation and hydration shell parameters of the macromolecule. Upon a ligand binding event to the macromolecule, at least one of these parameters change, resulting in a different thermophoretic behavior.[57] In the MST experiment, 16 capillary tubes are prepared containing a fluorescent labelled protein at constant concentration and a serial titration of unlabeled ligand. An infrared laser is used to generate a temperature gradient in each tube and induce migration of the ligand-bound complex, whose fluorescence is monitored in real time. Fluorescence variations are then used to generate a binding curve as a function of ligand concentration, which is instrumental to derive the dissociation constant (K_d).[58]

MST binding experiments were performed adding Taf-Ptx to fluorescently labelled TTR (RED-TTR). A biphasic binding curve was observed, suggesting the presence of more than one Taf-Ptx binding site on RED-TTR (Figure S15A). Specifically, at low concentrations of Taf-Ptx, a first binding event is observed with the occupancy of a high affinity site by Taf-Ptx in the target protein. Then, at higher ligand concentrations, a second binding event occurs with the occupancy of a low affinity binding site. These two binding events are well separated by approximately three orders of magnitude and thus can be independently analyzed with good approximation to obtain the relative dissociation constants ($K_d1 = 0.065 \pm 0.017$ μM, Figure S15C; $K_d2 = 9.21 \pm 1.14$ μM, Figure S15B). The second dissociation constant is found lower than the first for Taf-Ptx as for Tafamidis alone, and this is consistent with the NMR observations.

Overall, these results are in agreement with the NMR measurements and with literature data,[13] supporting a high affinity interaction between Taf-Ptx and TTR with a conjugated ligand/protein stoichiometry of 2:1.

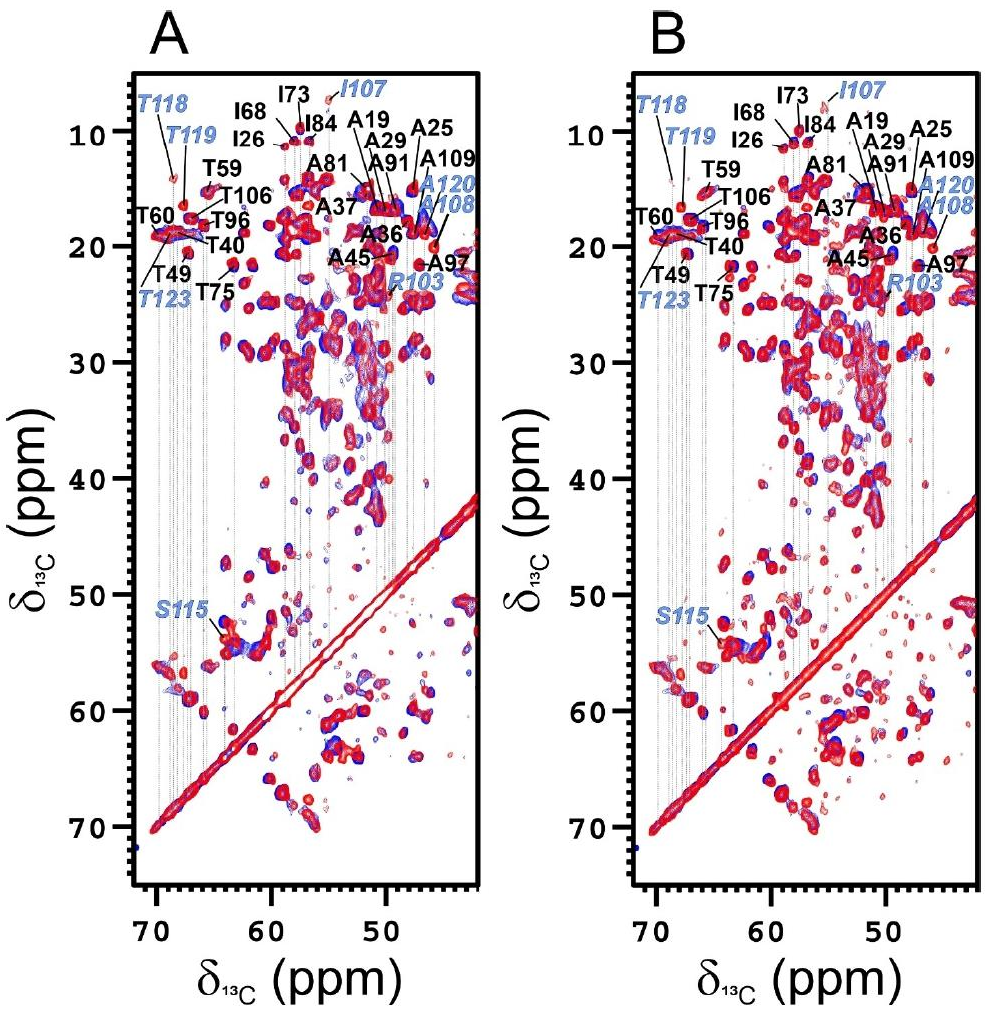


Figure 2. Comparison of a region of 2D ^{13}C-^{13}C DARR spectra (mixing time 50 ms, A, B) of free rehydrated freeze-dried TTR (blue) and TTR in the presence of the ligands (red), Tafamidis (A) or Taf-Ptx (B). The assignment of the cross-peaks related to correlations of Cβ/Cγ2 in threonine, Cα/Cδ1 in isoleucine and Cα/Cβ in alanine has been reported in the spectra. The signals with increased intensity in the spectra of TTR in complex with the ligands with respect to the free protein, have been labeled in italic font and colored in cyan. Assignment is reported also for the cross-peaks between Cβ/Cα of serine 115 and Cα/Cβ of arginine 103.

Chemical Denaturation Assay for Evaluating Taf-Ptx-Induced Stability of TTR

The impact of Taf-Ptx binding to TTR on the stability of the tetrameric protein was then evaluated using a chemical denaturation assay. Specifically, the intrinsic fluorescence of tryptophan residues (Trp41, Trp79) was monitored while denaturing the protein assembly with increasing concentrations of urea in the absence and presence of Taf-Ptx (see *Chemical denaturation assay* section in Supporting Information). In the presence of 50 μM Taf-Ptx, the tetrameric protein complex gains stability, with the relative denaturation curve shifting rightward and not reaching an unfolding plateau (Figure S16). A ΔG value $>47.3\pm2.5$ kJ mol^{-1} can be estimated in this case, suggesting a significant Taf-Ptx induced stability of TTR complex (Table S3).

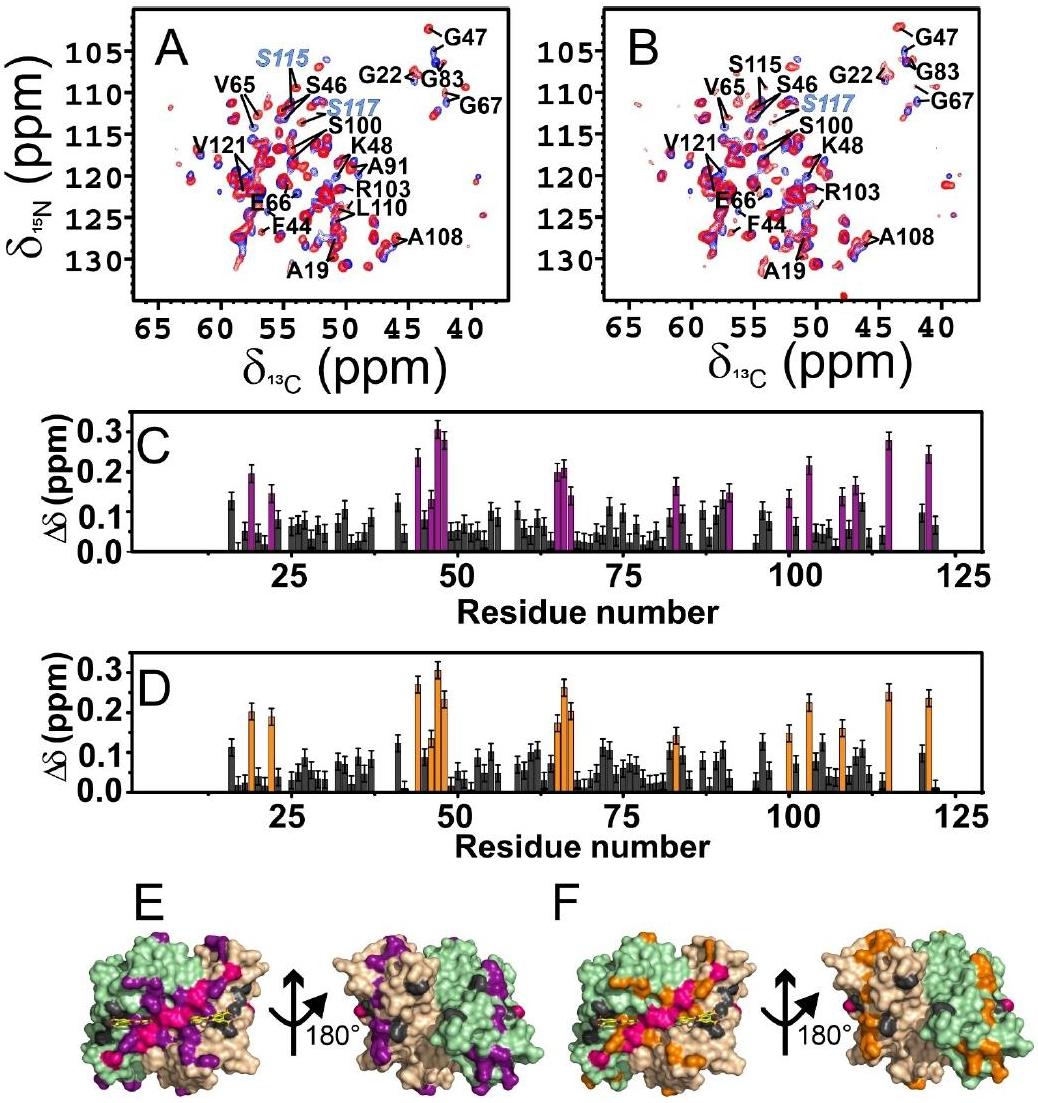


Figure 3. A), B) 2D ^{15}N ^{13}C NCA spectra of free rehydrated freeze-dried TTR (blue) and TTR in the presence of the ligands (red), Tafamidis (A) or Taf-Ptx (B). Assignment has been reported for the cross-peaks experiencing the largest perturbations ($\Delta\delta \geq$ mean + std. dev.). The signals with increased intensity in the spectra of TTR in complex with the ligands with respect to the free protein, have been labeled in italic font and colored in cyan. C), D) Chemical shift perturbation (CSP) of rehydrated freeze-dried TTR in the presence of Tafamidis (C) and Taf-Ptx (D) with respect to free rehydrated freeze-dried TTR, evaluated according to the formula $\Delta\delta = {}^1/_2\sqrt{(\Delta\delta_{C_\alpha}/2)^2 + (\Delta\delta_N/5)^2}$.[59] The residues experiencing the largest variations have been highlighted in violet (Ala19, Gly22, Phe44, Ser46, Gly47, Lys48, Val65, Glu66, Gly67, Gly83, Ala91, S100, Arg103, Ala108, Leu110, Ser115, Val121, for Tafamidis) and orange (Ala19, Gly22, Phe44, Ser46, Gly47, Lys48, Val65, Glu66, Gly67, Gly83, Ser100, Arg103, Ala108, Ser115, Val121, for Taf-Ptx), respectively. The error for CSP has been evaluated considering the standard deviation of the values of CSP below the mean, calculated considering all the CSP values. E), F) CSP mapping on the X-ray structure of TTR in complex with Tafamidis (PDB code: 3TCT)[13] with the residues experiencing the largest CSP in the presence of Tafamidis or Taf-Ptx colored in violet (E) and orange (F), respectively. The residues experiencing an increase in signal intensity after Tafamidis or Taf-Ptx binding have been colored in magenta. The residues missing in solid-state spectra are colored in grey. The monomers are in different colors (wheat, green) and Tafamidis is shown as yellow sticks.

Conclusion

CSP in solution are extensively used to analyze and compare the binding mode of ligands interacting with proteins or nucleic acids. For TTR the spectral changes resulting from the binding of Tafamidis or Taf-Ptx clearly point to a slow exchange regime on the NMR time scale for both ligands. Most of the residues experiencing the largest chemical shift variations on the NH resonances are placed at the interface of the two dimers and form the central channel hosting the thyroxine T4 hormone. Specifically, Lys15, Val20, Arg21, Gly22, Leu111, Ser112, Ala120, Val122 show chemical shift perturbations in the presence of both ligands, Tafamidis and Taf-Ptx. More importantly, the analysis of the X-ray structure (3TCT) of the 2:1 Tafamidis-TTR complex reveals that only Lys15, Leu17, and Thr106 appear to interact directly with the two ligand molecules among the residues experiencing the largest CSP. Therefore, the largest chemical shift perturbations appear to be mostly determined by the structural rearrangement induced by the two ligand molecules, making chemical shift mapping for the identification of the binding mode extremely challenging.

The stabilizing effect of Tafamidis was unambiguously proven in vitro and in vivo.[13,45] However, the chemical shift mapping obtained by NMR in solution provides only indirect evidence of this important structural effect, which can be inferred from the slow exchange regime on the NMR timescale of the signals corresponding to the residues where the binding occurs. In this respect, the observation of the slightly better quality of the spectra recorded on TTR in the presence of Tafamidis and Taf-Ptx is more informative.

It is important to point out that the use of solid-state NMR is critical for achieving a complete picture about the structural and dynamical features of this system. Indeed, the analysis of the data recorded on the solid-state samples of TTR in the presence and in the absence of the two ligands provides information that is out of reach for both X-ray crystallography and solution NMR. The comparative analyses on Tafamidis and Taf-Ptx were carried out using ^{15}N-^{13}C isotopically enriched samples of TTR and ^{13}C-detected experiments. The analysis of the NCA spectra shows that TTR in complex with Tafamidis and Taf-Ptx exhibits the largest chemical shift variations on the same residues, with very few differences. This constitutes further experimental evidence of the very similar binding mode of Taf-Ptx and Tafamidis. The chemical shift variations affect several residues placed at the interface of the two dimers around the central channel as previously observed by NMR in solution. Four of these residues (Ala108, Leu110, Ser115 and Val121) are near the ligands in the X-ray structure 3TCT. The non-overlap between solution and solid-state data about the residues experiencing the largest effects is probably due to a slightly higher conformational heterogeneity of the ligand-protein complex in solution, although in both cases they are localized at the interface between the two dimers forming the assembly. This can be clearly inferred from Figure 4, where the residues experiencing the largest CSPs ($\Delta\delta \geq$ mean + std. dev.) in solution and in the solid-state are shown together.

NMRD measurements indicate that the tetrameric assembly of TTR is maintained, and possibly reinforced, in the presence of Tafamidis. However, relaxometry is not sensitive to the presence of multiple conformational states with similar reorientation time. Important information on this respect was obtained from the qualitative analysis of the signal intensity on DARR and NCA spectra. Figure 4C shows that several residues placed at the interface between the two dimers forming the assembly, including some also experiencing large chemical shift variation, increase in intensity or appear in the spectra in the presence of Tafamidis and Taf-Ptx. The increase in signal intensity or the appearance of a signal previously undetectable in the solid-state spectra is conclusive evidence of an equilibrium of the residue shifting toward a unique conformation. This is associated with the structural stabilization of the tetrameric assembly, here proven by the chemical denaturation assay, resulting from the interaction with two high affinity ligand molecules with TTR. Therefore, the analysis provides a map of the residues experiencing a decrease of the conformational heterogeneity, thus providing a different and more informative parameter to monitor the binding mode and to evaluate the effects of the interaction with the ligands. This

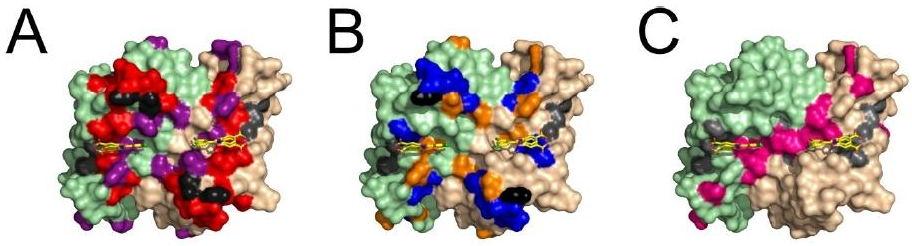


Figure 4. Comparative analysis of the NMR data collected in solution and in the solid-state of the effects of Tafamidis (A) and Taf-Ptx (B) on TTR. The residues experiencing in solution the largest CSP ($\Delta\delta \geq$ mean + std. dev.) in the presence of Tafamidis or Taf-Ptx are colored in red (A) or blue (B), respectively. The residues experiencing in the solid-state the largest CSP in the presence of Tafamidis or Taf-Ptx are colored in violet (A) or orange (B), respectively. Residues experiencing the largest CSP both in solution and in the solid state are reported in black. C) Residues experiencing in the solid-state an increase in signal intensity after the binding of Tafamidis or Taf-Ptx are colored in magenta. The residues missing in solid-state spectra are colored in grey. Monomers are in different colors (wheat, green) and Tafamidis is shown as yellow sticks.

information cannot be immediately inferred from the analysis of the CSP or from the observed improved quality of the NMR spectra recorded in solution.

When feasible, the integration of different structural biophysical methodologies is obviously the best option to develop PDCs. However, solid-state NMR can also be used on its own when the features of the investigated system prevent the use of other structural methodologies. The quality of the spectra obtained from the rehydrated freeze-dried samples, the high sensitivity of solid-state NMR to the effects of ligand binding and to small conformational heterogeneities make this technique extremely helpful to characterize the interaction of drug candidates with large monomeric/multimeric carrier proteins. In this regard, the information obtained from solid-state NMR data can be particularly important also for systems amenable to NMR characterization in solution, when the traditional chemical shift mapping based on the analysis of CSPs is not informative or resolutive.

Acknowledgements

This work has been supported by Regione Toscana (CERM-TT, BioEnable, and PANCREAS-AD bando salute 2018), the JOYNLAB laboratory, the Italian "Progetto Dipartimenti di Eccellenza 2023-2027 (DICUS2.0)". The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy centre. We acknowledge H2020 projects INFRAIA iNEXT-Discovery (contract n° 871037), FET-Open HIRES-MULTIDYN (contract n° 899683), PANACEA (contract n° 101008500), the Marie Sklodowska-Curie Action (MSCA) Innovative Training Networks (ITN) "RNAct" H2020-MSCA-ITN-2018 (contract n° 813239) and MSCA-ITN "Glytunes" (contract n° 956758), Fragment Screen (contract n° 101094131), and the project "Potentiating the Italian Capacity for Structural Biology Services in Instruct Eric (ITACA.SB)" (Project n° IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union NextGenerationEU. Authors AM and EB acknowledge research support by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041–VITALITY, and Università degli Studi di Perugia and MUR for support within the project Vitality.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Protein Data Bank https://www.rcsb.org/ and Biological Magnetic Resonance Bank https://bmrb.io/. The raw data of the SSNMR spectra are available at https://zenodo.org under the DOI: 10.5281/zenodo.8020132.

Keywords: Drug Delivery · Drug Design · NMR Spectroscopy · Protein-Drug Conjugates · Structural Biology

[1] P. Akkapeddi, S.-A. Azizi, A. M. Freedy, P. M. S. D. Cal, P. M. P. Gois, G. J. L. Bernardes, *Chem. Sci.* **2016**, *7*, 2954–2963.
[2] R. V. J. Chari, M. L. Miller, W. C. Widdison, *Angew. Chem. Int. Ed.* **2014**, *53*, 3796–3827.
[3] J. M. Harris, R. B. Chess, *Nat. Rev. Drug Discovery* **2003**, *2*, 214–221.
[4] P. Caravan, N. J. Cloutier, M. T. Greenfield, S. A. McDermid, S. U. Dunham, J. W. M. Bulte, J. C. Amedio, R. J. Looby, R. M. Supkowski, W. D. Horrocks, T. J. McMurry, R. B. Lauffer, *J. Am. Chem. Soc.* **2002**, *124*, 3152–3162.
[5] J. Huang, B. Wu, Z. Zhou, S. Hu, H. Xu, Y. Piao, H. Zheng, J. Tang, X. Liu, Y. Shen, *Nanomedicine: Nanotechnol. Biol. Med.* **2019**, *21*, 102058.
[6] O. A. Mandrup, S. C. Ong, S. Lykkemark, A. Dinesen, I. Rudnik-Jansen, N. F. Dagnæs-Hansen, J. T. Andersen, L. Alvarez-Vallina, K. A. Howard, *Commun. Biol.* **2021**, *4*, 310.
[7] A. Zorzi, S. Linciano, A. Angelini, *MedChemComm* **2019**, *10*, 1068–1081.
[8] A. Spada, J. Emami, J. A. Tuszynski, A. Lavasanifar, *Mol. Pharmaceutics* **2021**, *18*, 1862–1894.
[9] A. D. AlQahtani, D. O'Connor, A. Domling, S. K. Goda, *Biomed. Pharmacother.* **2019**, *113*, 108750.
[10] A. W. Yee, M. Aldeghi, M. P. Blakeley, A. Ostermann, P. J. Mas, M. Moulin, D. de Sanctis, M. W. Bowler, C. Mueller-Dieckmann, E. P. Mitchell, M. Haertlein, B. L. de Groot, E. Boeri Erba, V. T. Forsyth, *Nat. Commun.* **2019**, *10*, 925.
[11] M. M. Alhamadsheh, S. Connelly, A. Cho, N. Reixach, E. T. Powers, D. W. Pan, I. A. Wilson, J. W. Kelly, I. A. Graef, *Sci. Transl. Med.* **2011**, *3*, 97ra81.
[12] M. Miller, A. Pal, W. Albusairi, H. Joo, B. Pappas, M. T. Haque Tuhin, D. Liang, R. Jampala, F. Liu, J. Khan, M. Faaij, M. Park, W. Chan, I. Graef, R. Zamboni, N. Kumar, J. Fox, U. Sinha, M. Alhamadsheh, *J. Med. Chem.* **2018**, *61*, 7862–7876.
[13] C. E. Bulawa, S. Connelly, M. DeVit, L. Wang, C. Weigel, J. A. Fleming, J. Packman, E. T. Powers, R. L. Wiseman, T. R. Foss, I. A. Wilson, J. W. Kelly, R. Labaudinière, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9629–9634.
[14] A. Pal, W. Albusairi, F. Liu, M. T. H. Tuhin, M. Miller, D. Liang, H. Joo, T. U. Amin, E. A. Wilson, J. S. Faridi, M. Park, M. M. Alhamadsheh, *Mol. Pharmaceutics* **2019**, *16*, 3237–3252.
[15] S. C. Penchala, M. R. Miller, A. Pal, J. Dong, N. R. Madadi, J. Xie, H. Joo, J. Tsai, P. Batoon, V. Samoshin, A. Franz, T. Cox, J. Miles, W. K. Chan, M. S. Park, M. M. Alhamadsheh, *Nat. Chem. Biol.* **2015**, *11*, 793–798.
[16] F. Liu, T. Ul Amin, D. Liang, M. S. Park, M. M. Alhamadsheh, *J. Med. Chem.* **2021**, *64*, 14876–14886.
[17] J.-P. Renaud, A. Chari, C. Ciferri, W. Liu, H.-W. Rémigy, H. Stark, C. Wiesmann, *Nat. Rev. Drug Discovery* **2018**, *17*, 471–492.
[18] J. M. Lamley, D. Iuga, C. Öster, H.-J. Sass, M. Rogowski, A. Oss, J. Past, A. Reinhold, S. Grzesiek, A. Samoson, J. R. Lewandowski, *J. Am. Chem. Soc.* **2014**, *136*, 16800–16806.
[19] A. Mainz, T. L. Religa, R. Sprangers, R. Linser, L. E. Kay, B. Reif, *Angew. Chem. Int. Ed.* **2013**, *52*, 8746–8751.
[20] A. Mainz, S. Jehle, B. J. van Rossum, H. Oschkinat, B. Reif, *J. Am. Chem. Soc.* **2009**, *131*, 15968–15969.
[21] U. B. le Paige, S. Xiang, M. M. R. M. Hendrix, Y. Zhang, G. E. Folkers, M. Weingarth, A. M. J. J. Bonvin, T. G. Kutateladze,

I. K. Voets, M. Baldus, H. van Ingen, *Magn. Reson.* **2021**, *2*, 187–202.
[22] L. Eshun-Wilson, R. Zhang, D. Portran, M. V. Nachury, D. B. Toso, T. Löhr, M. Vendruscolo, M. Bonomi, J. S. Fraser, E. Nogales, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 10366–10371.
[23] H. Kato, H. van Ingen, B.-R. Zhou, H. Feng, M. Bustin, L. E. Kay, Y. Bai, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12283–12288.
[24] A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, G. M. Clore, *Science* **1991**, *253*, 657–661.
[25] E. Barbet-Massin, C.-T. Huang, V. Daebel, S.-T. D. Hsu, B. Reif, *Angew. Chem. Int. Ed.* **2015**, *54*, 4367–4369.
[26] D. Rizzo, L. Cerofolini, S. Giuntini, L. Iozzino, C. Pergola, F. Sacco, A. Palmese, E. Ravera, C. Luchinat, F. Baroni, M. Fragai, *J. Am. Chem. Soc.* **2022**, *144*, 10006–10016.
[27] Q. Wang, H. Yang, X. Liu, L. Dai, T. Ma, J. Qi, G. Wong, R. Peng, S. Liu, J. Li, S. Li, J. Song, J. Liu, J. He, H. Yuan, Y. Xiong, Y. Liao, J. Li, J. Yang, Z. Tong, B. D. Griffin, Y. Bi, M. Liang, X. Xu, C. Qin, G. Cheng, X. Zhang, P. Wang, X. Qiu, G. Kobinger, Y. Shi, J. Yan, G. F. Gao, *Sci. Transl. Med.* **2016**, *8*, 369ra179.
[28] S. Giuntini, L. Cerofolini, E. Ravera, M. Fragai, C. Luchinat, *Sci. Rep.* **2017**, *7*, 17934.
[29] L. Cerofolini, S. Giuntini, E. Ravera, C. Luchinat, F. Berti, M. Fragai, *npj Vaccines* **2019**, *4*, 20.
[30] L. Cerofolini, S. Giuntini, A. Carlon, E. Ravera, V. Calderone, M. Fragai, G. Parigi, C. Luchinat, *Chem. Eur. J.* **2019**, *25*, 1984–1991.
[31] E. Ravera, S. Ciambellotti, L. Cerofolini, T. Martelli, T. Kozyreva, C. Bernacchioni, S. Giuntini, M. Fragai, P. Turano, C. Luchinat, *Angew. Chem. Int. Ed.* **2016**, *55*, 2446–2449.
[32] L. Lecoq, M.-L. Fogeron, B. H. Meier, M. Nassal, A. Böckmann, *Viruses* **2020**, *12*, E1069.
[33] T. Wiegand, D. Lacabanne, A. Torosyan, J. Boudet, R. Cadalbert, F. H.-T. Allain, B. H. Meier, A. Böckmann, *Front. Mol. Biosci.* **2020**, *7*, 17.
[34] M. Lu, R. W. Russell, A. J. Bryer, C. M. Quinn, G. Hou, H. Zhang, C. D. Schwieters, J. R. Perilla, A. M. Gronenborn, T. Polenova, *Nat. Struct. Mol. Biol.* **2020**, *27*, 863–869.
[35] M. T. Eddy, T.-Y. Yu, G. Wagner, R. G. Griffin, *J. Biomol. NMR* **2019**, *73*, 451–460.
[36] M. Fragai, C. Luchinat, T. Martelli, E. Ravera, I. Sagi, I. Solomonov, Y. Udi, *Chem. Commun.* **2014**, *50*, 421–423.
[37] K. Jaudzems, A. Kirsteina, T. Schubeis, G. Casano, O. Ouari, J. Bogans, A. Kazaks, K. Tars, A. Lesage, G. Pintacuda, *Angew. Chem. Int. Ed.* **2021**, *60*, 12847–12851.
[38] D. Rizzo, L. Cerofolini, A. Pérez-Ràfols, S. Giuntini, F. Baroni, E. Ravera, C. Luchinat, M. Fragai, *Anal. Chem.* **2021**, *93*, 11208–11214.
[39] S. Gupta, R. Tycko, *J. Biomol. NMR* **2018**, *70*, 103–114.
[40] R. Gupta, H. Zhang, M. Lu, G. Hou, M. Caporini, M. Rosay, W. Maas, J. Struppe, J. Ahn, I.-J. L. Byeon, H. Oschkinat, K. Jaudzems, E. Barbet-Massin, L. Emsley, G. Pintacuda, A. Lesage, A. M. Gronenborn, T. Polenova, *J. Phys. Chem. B* **2019**, *123*, 5048–5058.
[41] T. Azaïs, S. Von Euw, W. Ajili, S. Auzoux-Bordenave, P. Bertani, D. Gajan, L. Emsley, N. Nassif, A. Lesage, *Solid State Nucl. Magn. Reson.* **2019**, *102*, 2–11.
[42] J. Viger-Gravel, F. M. Paruzzo, C. Cazaux, R. Jabbour, A. Leleu, F. Canini, P. Florian, F. Ronzon, D. Gajan, A. Lesage, *Chem. Eur. J.* **2020**, *26*, 8976–8982.
[43] A. Hassan, C. M. Quinn, J. Struppe, I. V. Sergeyev, C. Zhang, C. Guo, B. Runge, T. Theint, H. H. Dao, C. P. Jaroniec, M. Berbon, A. Lends, B. Habenstein, A. Loquet, R. Kuemmerle, B. Perrone, A. M. Gronenborn, T. Polenova, *J. Magn. Reson.* **2020**, *311*, 106680.
[44] R. Zhang, Y. Hong, T. Ravula, Y. Nishiyama, A. Ramamoorthy, *J. Magn. Reson.* **2020**, *313*, 106717.
[45] M. S. Maurer, J. H. Schwartz, B. Gundapaneni, P. M. Elliott, G. Merlini, M. Waddington-Cruz, A. V. Kristen, M. Grogan, R. Witteles, T. Damy, B. M. Drachman, S. J. Shah, M. Hanna, D. P. Judge, A. I. Barsdorf, P. Huber, T. A. Patterson, S. Riley, J. Schumacher, M. Stewart, M. B. Sultan, C. Rapezzi, *N. Engl. J. Med.* **2018**, *379*, 1007–1016.
[46] I. Bertini, M. Fragai, C. Luchinat, G. Parigi, *Magn. Reson. Chem.* **2000**, *38*, 543–550.
[47] G. Parigi, E. Ravera, M. Fragai, C. Luchinat, *Prog. Nucl. Magn. Reson. Spectrosc.* **2021**, *124–125*, 85–98.
[48] E. Ravera, G. Parigi, A. Mainz, T. L. Religa, B. Reif, C. Luchinat, *J. Phys. Chem. B* **2013**, *117*, 3548–3553.
[49] J. García de la Torre, M. L. Huertas, B. Carrasco, *J. Magn. Reson.* **2000**, *147*, 138–146.
[50] J. Brunetti, S. Piantini, M. Fragai, S. Scali, G. Cipriani, L. Depau, A. Pini, C. Falciani, S. Menichetti, L. Bracci, *Molecules* **2020**, *25*, E1088.
[51] J. Brunetti, S. Pillozzi, C. Falciani, L. Depau, E. Tenori, S. Scali, L. Lozzi, A. Pini, A. Arcangeli, S. Menichetti, L. Bracci, *Sci. Rep.* **2015**, *5*, 17736.
[52] K. Liu, J. W. Kelly, D. E. Wemmer, *J. Mol. Biol.* **2002**, *320*, 821–832.
[53] J. Oroz, J. H. Kim, B. J. Chang, M. Zweckstetter, *Nat. Struct. Mol. Biol.* **2017**, *24*, 407–413.
[54] B. I. Leach, X. Zhang, J. W. Kelly, H. J. Dyson, P. E. Wright, *Biochemistry* **2018**, *57*, 4421–4430.
[55] Y.-T. Liu, Y.-J. Yen, F. Ricardo, Y. Chang, P.-H. Wu, S.-J. Huang, K.-P. Lin, T.-Y. Yu, *Ann. Clin. Transl. Neurol.* **2019**, *6*, 1961–1970.
[56] A. Corazza, G. Verona, C. A. Waudby, P. P. Mangione, R. Bingham, I. Uings, D. Canetti, P. Nocerino, G. W. Taylor, M. B. Pepys, J. Christodoulou, V. Bellotti, *J. Med. Chem.* **2019**, *62*, 8274–8283.
[57] S. Duhr, D. Braun, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19678–19682.
[58] M. Jerabek-Willemsen, C. J. Wienken, D. Braun, P. Baaske, S. Duhr, *Assay Drug Dev. Technol.* **2011**, *9*, 342–353.
[59] S. Grzesiek, A. Bax, G. M. Clore, A. M. Gronenborn, J. S. Hu, J. Kaufman, I. Palmer, S. J. Stahl, P. T. Wingfield, *Nat. Struct. Biol.* **1996**, *3*, 340–345.

Manuscript received: March 13, 2023
Accepted manuscript online: June 5, 2023
Version of record online: June 22, 2023

6. Conclusions

The structural and biophysical characterization of biomolecules is fundamental to deepening our understanding of complex biological systems, and to set the bases for the discovery and implementation of new therapeutic strategies. Elucidating the structural basis of molecular recognition is essential for the identification of critical binding sites and binding modes, conformational changes and dynamic features of the biomolecular complexes ruling biological processes. Equally important is the characterization of drug delivery systems, which play an important role in the stability, efficacy and targeted delivery of therapeutic agents. The application of NMR spectroscopy has been an important tool in this Ph.D. project for the structural characterization of the above-mentioned systems at an atomic level. Moreover, NMR excels as the only tool in structural biology for the characterization of specific systems, in particular biomolecules embedded in complex and non-uniform matrices.

The structural characterization of Siglec-7CRD and siglec-like binding region-N revealed important details about its binding specificity towards sialoglycans. The integration of protein-based and ligand-based NMR experiments were crucial in the determination of the binding features of GD3, DGb3-α3 and DGb3-α6, as well as the OPS from *F. nucleatum ATCC 10953*. We were able to determine how this binding events influence and shape the protein binding site and ligand architecture. This provides an understanding of the biological and pathological processes mediated by Siglec-7 both in the tumor microenvironment and bacterial infections.

Furthermore, we demonstrated the preservation of the Higher Order Structure (HOS) upon erythrocyte encapsulation. This is a critical factor for ensuring functionality and stability of therapeutic biologics. The obtained NMR data confirmed the preservation of the native conformation of three different proteins, human carbonic anhydrase II, human transthyretin and asparaginase II. This supports the versatility and efficiency of the encapsulation protocol and NMR methodology applied for the study of different proteins in complex environments.

Finally, solid-state NMR proved to be a powerful technique for the structural characterization of protein-drug conjugates. Atomic-resolution insights into a non-crystallizable complex obtained by solid-state NMR, underscores the importance of the availability of multiple analytical approaches to overcome the limitations of individual techniques. This, at the same time, allows to advance in the rational design of effective drug delivery systems.

7. References

(1) Lewis, A. L.; Chen, X.; Schnaar, R. L.; Varki, A. Sialic Acids and Other Nonulosonic Acids. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2022.

(2) McDonald, N. D.; Boyd, E. F. Structural and Biosynthetic Diversity of Nonulosonic Acids (NulOs) That Decorate Surface Structures in Bacteria. *Trends Microbiol.* **2021**, *29* (2), 142–157. https://doi.org/10.1016/j.tim.2020.08.002.

(3) Ajit Varki, R. S. *Sialic Acids*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2009.

(4) Li, D.; Lin, Q.; Luo, F.; Wang, H. Insights into the Structure, Metabolism, Biological Functions and Molecular Mechanisms of Sialic Acid: A Review. *Foods* **2023**, *13* (1), 145. https://doi.org/10.3390/foods13010145.

(5) Lewis, A. L.; Lewis, W. G. Host Sialoglycans and Bacterial Sialidases: A Mucosal Perspective. *Cell. Microbiol.* **2012**, *14* (8), 1174–1182. https://doi.org/10.1111/j.1462-5822.2012.01807.x.

(6) Bärenwaldt, A.; Läubli, H. The Sialoglycan-Siglec Glyco-Immune Checkpoint – a Target for Improving Innate and Adaptive Anti-Cancer Immunity. *Expert Opin. Ther. Targets* **2019**, *23* (10), 839–853. https://doi.org/10.1080/14728222.2019.1667977.

(7) Cavalcante, T.; Medeiros, M. M.; Mule, S. N.; Palmisano, G.; Stolf, B. S. The Role of Sialic Acids in the Establishment of Infections by Pathogens, With Special Focus on Leishmania. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 671913. https://doi.org/10.3389/fcimb.2021.671913.

(8) Van De Wall, S.; Santegoets, K. C. M.; Van Houtum, E. J. H.; Büll, C.; Adema, G. J. Sialoglycans and Siglecs Can Shape the Tumor Immune Microenvironment. *Trends Immunol.* **2020**, *41* (4), 274–285. https://doi.org/10.1016/j.it.2020.02.001.

(9) Jennings, M. P.; Day, C. J.; Atack, J. M. How Bacteria Utilize Sialic Acid during Interactions with the Host: Snip, Snatch, Dispatch, Match and Attach: This Article Is Part of the Bacterial Cell Envelopes Collection and the Environmental Sensing and Cell-Cell Communication Collection. *Microbiology* **2022**, *168* (3). https://doi.org/10.1099/mic.0.001157.

(10) Dedola, S.; Ahmadipour, S.; De Andrade, P.; Baker, A. N.; Boshra, A. N.; Chessa, S.; Gibson, M. I.; Hernando, P. J.; Ivanova, I. M.; Lloyd, J. E.; Marín, M. J.; Munro-Clark, A. J.; Pergolizzi, G.; Richards, S.-J.; Ttofi, I.; Wagstaff, B. A.; Field, R. A. Sialic Acids in Infection and Their Potential Use in Detection and Protection against Pathogens. *RSC Chem. Biol.* **2024**, *5* (3), 167–188. https://doi.org/10.1039/D3CB00155E.

(11) Wang, L.; Liu, Y.; Wu, L.; Sun, X.-L. Sialyltransferase Inhibition and Recent Advances. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (1), 143–153. https://doi.org/10.1016/j.bbapap.2015.07.007.

(12) Harduin-Lepers, A.; Vallejo-Ruiz, V.; Krzewinski-Recchi, M.-A.; Samyn-Petit, B.; Julien, S.; Delannoy, P. The Human Sialyltransferase Family. *Biochimie* **2001**, *83* (8), 727–737. https://doi.org/10.1016/S0300-9084(01)01301-3.
(13) Inoue, S.; Poongodi, G. L.; Suresh, N.; Jennings, H. J.; Inoue, Y. Discovery of an A2,9-PolyNeu5Ac Glycoprotein in C-1300 Murine Neuroblastoma (Clone NB41A3). *J. Biol. Chem.* **2003**, *278* (10), 8541–8546. https://doi.org/10.1074/jbc.M212799200.
(14) Miyata, S. A Major Flagellum Sialoglycoprotein in Sea Urchin Sperm Contains a Novel Polysialic Acid, an 2,9-Linked Poly-N-Acetylneuraminic Acid Chain, Capped by an 8-O-Sulfated Sialic Acid Residue. *Glycobiology* **2004**, *14* (9), 827–840. https://doi.org/10.1093/glycob/cwh100.
(15) Miyata, S.; Sato, C.; Kumita, H.; Toriyama, M.; Vacquier, V. D.; Kitajima, K. Flagellasialin: A Novel Sulfated 2,9-Linked Polysialic Acid Glycoprotein of Sea Urchin Sperm Flagella. *Glycobiology* **2006**, *16* (12), 1229–1241. https://doi.org/10.1093/glycob/cwl036.
(16) Miyata, S.; Yamakawa, N.; Toriyama, M.; Sato, C.; Kitajima, K. Co-Expression of Two Distinct Polysialic Acids, A2,8- and A2,9-Linked Polymers of N-Acetylneuraminic Acid, in Distinct Glycoproteins and Glycolipids in Sea Urchin Sperm. *Glycobiology* **2011**, *21* (12), 1596–1605. https://doi.org/10.1093/glycob/cwr081.
(17) Hugonnet, M.; Singh, P.; Haas, Q.; Von Gunten, S. The Distinct Roles of Sialyltransferases in Cancer Biology and Onco-Immunology. *Front. Immunol.* **2021**, *12*, 799861. https://doi.org/10.3389/fimmu.2021.799861.
(18) Adak, A. K.; Yu, C.-C.; Liang, C.-F.; Lin, C.-C. Synthesis of Sialic Acid-Containing Saccharides. *Curr. Opin. Chem. Biol.* **2013**, *17* (6), 1030–1038. https://doi.org/10.1016/j.cbpa.2013.10.013.
(19) Varki, A. Glycan-Based Interactions Involving Vertebrate Sialic-Acid-Recognizing Proteins. *Nature* **2007**, *446* (7139), 1023–1029. https://doi.org/10.1038/nature05816.
(20) Pillai, S.; Netravali, I. A.; Cariappa, A.; Mattoo, H. Siglecs and Immune Regulation. *Annu. Rev. Immunol.* **2012**, *30* (1), 357–392. https://doi.org/10.1146/annurev-immunol-020711-075018.
(21) Di Carluccio, C.; Forgione, R. E.; Molinaro, A.; Crocker, P. R.; Marchetti, R.; Silipo, A. Exploring the Fascinating World of Sialoglycans in the Interplay with Siglecs. In *Carbohydrate Chemistry*; Pilar Rauter, A., Lindhorst, T. K., Queneau, Y., Eds.; Royal Society of Chemistry: Cambridge, 2020; Vol. 44, pp 31–55. https://doi.org/10.1039/9781788013864-00031.
(22) Varki, A.; Angata, T. Siglecs—the Major Subfamily of I-Type Lectins. *Glycobiology* **2006**, *16* (1), 1R-27R. https://doi.org/10.1093/glycob/cwj008.
(23) Angata, T.; Varki, N. M.; Varki, A. A Second Uniquely Human Mutation Affecting Sialic Acid Biology. *J. Biol. Chem.* **2001**, *276* (43), 40282–40287. https://doi.org/10.1074/jbc.M105926200.
(24) Mitra, N.; Banda, K.; Altheide, T. K.; Schaffer, L.; Johnson-Pais, T. L.; Beuten, J.; Leach, R. J.; Angata, T.; Varki, N.; Varki, A. SIGLEC12, a Human-Specific Segregating (Pseudo)Gene, Encodes a Signaling Molecule Expressed in Prostate Carcinomas. *J. Biol. Chem.* **2011**, *286* (26), 23003–23011. https://doi.org/10.1074/jbc.M111.244152.

(25) Avril, T.; Attrill, H.; Zhang, J.; Raper, A.; Crocker, P. R. Negative Regulation of Leucocyte Functions by CD33-Related Siglecs. *Biochem. Soc. Trans.* **2006**, *34* (6), 1024–1027. https://doi.org/10.1042/BST0341024.
(26) Müller, J.; Obermeier, I.; Wöhner, M.; Brandl, C.; Mrotzek, S.; Angermüller, S.; Maity, P. C.; Reth, M.; Nitschke, L. CD22 Ligand-Binding and Signaling Domains Reciprocally Regulate B-Cell Ca^{2+} Signaling. *Proc. Natl. Acad. Sci.* **2013**, *110* (30), 12402–12407. https://doi.org/10.1073/pnas.1304888110.
(27) Clark, E. A.; Giltiay, N. V. CD22: A Regulator of Innate and Adaptive B Cell Responses and Autoimmunity. *Front. Immunol.* **2018**, *9*, 2235. https://doi.org/10.3389/fimmu.2018.02235.
(28) Tourdot, B. E.; Brenner, M. K.; Keough, K. C.; Holyst, T.; Newman, P. J.; Newman, D. K. Immunoreceptor Tyrosine-Based Inhibitory Motif (ITIM)-Mediated Inhibitory Signaling Is Regulated by Sequential Phosphorylation Mediated by Distinct Nonreceptor Tyrosine Kinases: A Case Study Involving PECAM-1. *Biochemistry* **2013**, *52* (15), 2597–2608. https://doi.org/10.1021/bi301461t.
(29) Arquint, M.; Roder, J.; Chia, L. S.; Down, J.; Wilkinson, D.; Bayley, H.; Braun, P.; Dunn, R. Molecular Cloning and Primary Structure of Myelin-Associated Glycoprotein. *Proc. Natl. Acad. Sci.* **1987**, *84* (2), 600–604. https://doi.org/10.1073/pnas.84.2.600.
(30) Lopez, P. H. H. Role of Myelin-Associated Glycoprotein (Siglec-4a) in the Nervous System. In *Glycobiology of the Nervous System*; Yu, R. K., Schengrund, C.-L., Eds.; Advances in Neurobiology; Springer New York: New York, NY, 2014; Vol. 9, pp 245–262. https://doi.org/10.1007/978-1-4939-1154-7_11.
(31) Izquierdo-Useros, N.; Lorizate, M.; Puertas, M. C.; Rodriguez-Plata, M. T.; Zangger, N.; Erikson, E.; Pino, M.; Erkizia, I.; Glass, B.; Clotet, B.; Keppler, O. T.; Telenti, A.; Kräusslich, H.-G.; Martinez-Picado, J. Siglec-1 Is a Novel Dendritic Cell Receptor That Mediates HIV-1 Trans-Infection Through Recognition of Viral Membrane Gangliosides. *PLoS Biol.* **2012**, *10* (12), e1001448. https://doi.org/10.1371/journal.pbio.1001448.
(32) Perez-Zsolt, D.; Martinez-Picado, J.; Izquierdo-Useros, N. When Dendritic Cells Go Viral: The Role of Siglec-1 in Host Defense and Dissemination of Enveloped Viruses. *Viruses* **2019**, *12* (1), 8. https://doi.org/10.3390/v12010008.
(33) Grabowska, J.; Lopez-Venegas, M. A.; Affandi, A. J.; Den Haan, J. M. M. CD169+ Macrophages Capture and Dendritic Cells Instruct: The Interplay of the Gatekeeper and the General of the Immune System. *Front. Immunol.* **2018**, *9*, 2472. https://doi.org/10.3389/fimmu.2018.02472.
(34) Martinez-Pomares, L.; Gordon, S. CD169+ Macrophages at the Crossroads of Antigen Presentation. *Trends Immunol.* **2012**, *33* (2), 66–70. https://doi.org/10.1016/j.it.2011.11.001.
(35) Barral, P.; Polzella, P.; Bruckbauer, A.; Van Rooijen, N.; Besra, G. S.; Cerundolo, V.; Batista, F. D. CD169+ Macrophages Present Lipid Antigens to Mediate Early Activation of iNKT Cells in Lymph Nodes. *Nat. Immunol.* **2010**, *11* (4), 303–312. https://doi.org/10.1038/ni.1853.
(36) Angata, T.; Varki, A. Discovery, Classification, Evolution and Diversity of Siglecs. *Mol. Aspects Med.* **2023**, *90*, 101117. https://doi.org/10.1016/j.mam.2022.101117.

(37) Shao, J. -Y.; Yin, W. -W.; Zhang, Q. -F.; Liu, Q.; Peng, M. -L.; Hu, H. -D.; Hu, P.; Ren, H.; Zhang, D. -Z. Siglec-7 Defines a Highly Functional Natural Killer Cell Subset and Inhibits Cell-Mediated Activities. *Scand. J. Immunol.* **2016**, *84* (3), 182–190. https://doi.org/10.1111/sji.12455.
(38) Mei, Y.; Wang, X.; Zhang, J.; Liu, D.; He, J.; Huang, C.; Liao, J.; Wang, Y.; Feng, Y.; Li, H.; Liu, X.; Chen, L.; Yi, W.; Chen, X.; Bai, H.-M.; Wang, X.; Li, Y.; Wang, L.; Liang, Z.; Ren, X.; Qiu, L.; Hui, Y.; Zhang, Q.; Leng, Q.; Chen, J.; Jia, G. Siglec-9 Acts as an Immune-Checkpoint Molecule on Macrophages in Glioblastoma, Restricting T-Cell Priming and Immunotherapy Response. *Nat. Cancer* **2023**, *4* (9), 1273–1291. https://doi.org/10.1038/s43018-023-00598-9.
(39) Duan, S.; Paulson, J. C. Siglecs as Immune Cell Checkpoints in Disease. *Annu. Rev. Immunol.* **2020**, *38* (1), 365–395. https://doi.org/10.1146/annurev-immunol-102419-035900.
(40) Büll, C.; Heise, T.; Adema, G. J.; Boltje, T. J. Sialic Acid Mimetics to Target the Sialic Acid–Siglec Axis. *Trends Biochem. Sci.* **2016**, *41* (6), 519–531. https://doi.org/10.1016/j.tibs.2016.03.007.
(41) O'Reilly, M. K.; Paulson, J. C. Multivalent Ligands for Siglecs. In *Methods in Enzymology*; Elsevier, 2010; Vol. 478, pp 343–363. https://doi.org/10.1016/S0076-6879(10)78017-4.
(42) Conti, G.; Bärenwaldt, A.; Rabbani, S.; Mühlethaler, T.; Sarcevic, M.; Jiang, X.; Schwardt, O.; Ricklin, D.; Pieters, R. J.; Läubli, H.; Ernst, B. Tetra- and Hexavalent Siglec-8 Ligands Modulate Immune Cell Activation. *Angew. Chem.* **2023**, *135* (52), e202314280. https://doi.org/10.1002/ange.202314280.
(43) Murugesan, G.; Weigle, B.; Crocker, P. R. Siglec and Anti-Siglec Therapies. *Curr. Opin. Chem. Biol.* **2021**, *62*, 34–42. https://doi.org/10.1016/j.cbpa.2021.01.001.
(44) Falco, M.; Biassoni, R.; Bottino, C.; Vitale, M.; Sivori, S.; Augugliaro, R.; Moretta, L.; Moretta, A. Identification and Molecular Cloning of P75/Airm1, a Novel Member of the Sialoadhesin Family That Functions as an Inhibitory Receptor in Human Natural Killer Cells. *J. Exp. Med.* **1999**, *190* (6), 793–802. https://doi.org/10.1084/jem.190.6.793.
(45) Nicoll, G.; Ni, J.; Liu, D.; Klenerman, P.; Munday, J.; Dubock, S.; Mattei, M.-G.; Crocker, P. R. Identification and Characterization of a Novel Siglec, Siglec-7, Expressed by Human Natural Killer Cells and Monocytes. *J. Biol. Chem.* **1999**, *274* (48), 34089–34095. https://doi.org/10.1074/jbc.274.48.34089.
(46) Angata, T.; Varki, A. Siglec-7: A Sialic Acid-Binding Lectin of the Immunoglobulin Superfamily. *Glycobiology* **2000**, *10* (4), 431–438. https://doi.org/10.1093/glycob/10.4.431.
(47) 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. Discovery of a New Sialic Acid Binding Region That Regulates Siglec-7. *Sci. Rep.* **2020**, *10* (1), 8647. https://doi.org/10.1038/s41598-020-64887-4.
(48) May, A. P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y. Crystal Structure of the N-Terminal Domain of Sialoadhesin in Complex with 3' Sialyllactose at 1.85 Å

Resolution. *Mol. Cell* **1998**, *1* (5), 719–728. https://doi.org/10.1016/S1097-2765(00)80071-4.
(49) Alphey, M. S.; Attrill, H.; Crocker, P. R.; Van Aalten, D. M. F. High Resolution Crystal Structures of Siglec-7. *J. Biol. Chem.* **2003**, *278* (5), 3372–3377. https://doi.org/10.1074/jbc.M210602200.
(50) Zhuravleva, M. A.; Trandem, K.; Sun, P. D. Structural Implications of Siglec-5-Mediated Sialoglycan Recognition. *J. Mol. Biol.* **2008**, *375* (2), 437–447. https://doi.org/10.1016/j.jmb.2007.10.009.
(51) Pronker, M. F.; Lemstra, S.; Snijder, J.; Heck, A. J. R.; Thies-Weesie, D. M. E.; Pasterkamp, R. J.; Janssen, B. J. C. Structural Basis of Myelin-Associated Glycoprotein Adhesion and Signalling. *Nat. Commun.* **2016**, *7* (1), 13584. https://doi.org/10.1038/ncomms13584.
(52) Pröpster, J. M.; Yang, F.; Rabbani, S.; Ernst, B.; Allain, F. H.-T.; Schubert, M. Structural Basis for Sulfation-Dependent Self-Glycan Recognition by the Human Immune-Inhibitory Receptor Siglec-8. *Proc. Natl. Acad. Sci.* **2016**, *113* (29). https://doi.org/10.1073/pnas.1602214113.
(53) Ereño-Orbea, J.; Sicard, T.; Cui, H.; Mazhab-Jafari, M. T.; Benlekbir, S.; Guarné, A.; Rubinstein, J. L.; Julien, J.-P. Molecular Basis of Human CD22 Function and Therapeutic Targeting. *Nat. Commun.* **2017**, *8* (1), 764. https://doi.org/10.1038/s41467-017-00836-6.
(54) Miles, L. A.; Hermans, S. J.; Crespi, G. A. N.; Gooi, J. H.; Doughty, L.; Nero, T. L.; Markulić, J.; Ebneth, A.; Wroblowski, B.; Oehlrich, D.; Trabanco, A. A.; Rives, M.-L.; Royaux, I.; Hancock, N. C.; Parker, M. W. Small Molecule Binding to Alzheimer Risk Factor CD33 Promotes Aβ Phagocytosis. *iScience* **2019**, *19*, 110–118. https://doi.org/10.1016/j.isci.2019.07.023.
(55) Lenza, M. P.; Egia-Mendikute, L.; Antoñana-Vildosola, A.; Soares, C. O.; Coelho, H.; Corzana, F.; Bosch, A.; Manisha, P.; Quintana, J. I.; Oyenarte, I.; Unione, L.; Moure, M. J.; Azkargorta, M.; Atxabal, U.; Sobczak, K.; Elortza, F.; Sutherland, J. D.; Barrio, R.; Marcelo, F.; Jiménez-Barbero, J.; Palazon, A.; Ereño-Orbea, J. Structural Insights into Siglec-15 Reveal Glycosylation Dependency for Its Interaction with T Cells through Integrin CD11b. *Nat. Commun.* **2023**, *14* (1), 3496. https://doi.org/10.1038/s41467-023-39119-8.
(56) Lenza, M. P.; Atxabal, U.; Nycholat, C.; Oyenarte, I.; Franconetti, A.; Quintana, J. I.; Delgado, S.; Núñez-Franco, R.; Garnica Marroquín, C. T.; Coelho, H.; Unione, L.; Jiménez-Oses, G.; Marcelo, F.; Schubert, M.; Paulson, J. C.; Jiménez-Barbero, J.; Ereño-Orbea, J. Structures of the Inhibitory Receptor Siglec-8 in Complex with a High-Affinity Sialoside Analogue and a Therapeutic Antibody. *JACS Au* **2023**, *3* (1), 204–215. https://doi.org/10.1021/jacsau.2c00592.
(57) Hudak, J. E.; Canham, S. M.; Bertozzi, C. R. Glycocalyx Engineering Reveals a Siglec-Based Mechanism for NK Cell Immunoevasion. *Nat. Chem. Biol.* **2014**, *10* (1), 69–75. https://doi.org/10.1038/nchembio.1388.
(58) Landolina, N.; Zaffran, I.; Smiljkovic, D.; Serrano-Candelas, E.; Schmiedel, D.; Friedman, S.; Arock, M.; Hartmann, K.; Pikarsky, E.; Mandelboim, O.; Martin, M.; Valent,

P.; Levi-Schaffer, F. Activation of Siglec-7 Results in Inhibition of in Vitro and in Vivo Growth of Human Mast Cell Leukemia Cells. *Pharmacol. Res.* **2020**, *158*, 104682. https://doi.org/10.1016/j.phrs.2020.104682.

(59) Läubli, H.; Alisson-Silva, F.; Stanczak, M. A.; Siddiqui, S. S.; Deng, L.; Verhagen, A.; Varki, N.; Varki, A. Lectin Galactoside-Binding Soluble 3 Binding Protein (LGALS3BP) Is a Tumor-Associated Immunomodulatory Ligand for CD33-Related Siglecs. *J. Biol. Chem.* **2014**, *289* (48), 33481–33491. https://doi.org/10.1074/jbc.M114.593129.

(60) Yap, T. A.; Parkes, E. E.; Peng, W.; Moyers, J. T.; Curran, M. A.; Tawbi, H. A. Development of Immunotherapy Combination Strategies in Cancer. *Cancer Discov.* **2021**, *11* (6), 1368–1397. https://doi.org/10.1158/2159-8290.CD-20-1209.

(61) Brunetta, E.; Fogli, M.; Varchetta, S.; Bozzo, L.; Hudspeth, K. L.; Marcenaro, E.; Moretta, A.; Mavilio, D. 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* **2009**, *114* (18), 3822–3830. https://doi.org/10.1182/blood-2009-06-226332.

(62) Brunetta, E.; Hudspeth, K. L.; Mavilio, D. Pathologic Natural Killer Cell Subset Redistribution in HIV-1 Infection: New Insights in Pathophysiology and Clinical Outcomes. *J. Leukoc. Biol.* **2010**, *88* (6), 1119–1130. https://doi.org/10.1189/jlb.0410225.

(63) Lamprinaki, D.; Garcia-Vello, P.; Marchetti, R.; Hellmich, C.; McCord, K. A.; Bowles, K. M.; Macauley, M. S.; Silipo, A.; De Castro, C.; Crocker, P. R.; Juge, N. Siglec-7 Mediates Immunomodulation by Colorectal Cancer-Associated Fusobacterium Nucleatum Ssp. Animalis. *Front. Immunol.* **2021**, *12*, 744184. https://doi.org/10.3389/fimmu.2021.744184.

(64) Galaski, J.; Rishiq, A.; Liu, M.; Bsoul, R.; Bergson, A.; Lux, R.; Bachrach, G.; Mandelboim, O. Fusobacterium Nucleatum Subsp. Nucleatum RadD Binds Siglec-7 and Inhibits NK Cell-Mediated Cancer Cell Killing. *iScience* **2024**, *27* (6), 110157. https://doi.org/10.1016/j.isci.2024.110157.

(65) Kostic, A. D.; Gevers, D.; Pedamallu, C. S.; Michaud, M.; Duke, F.; Earl, A. M.; Ojesina, A. I.; Jung, J.; Bass, A. J.; Tabernero, J.; Baselga, J.; Liu, C.; Shivdasani, R. A.; Ogino, S.; Birren, B. W.; Huttenhower, C.; Garrett, W. S.; Meyerson, M. Genomic Analysis Identifies Association of *Fusobacterium* with Colorectal Carcinoma. *Genome Res.* **2012**, *22* (2), 292–298. https://doi.org/10.1101/gr.126573.111.

(66) Dai, Z.; Coker, O. O.; Nakatsu, G.; Wu, W. K. K.; Zhao, L.; Chen, Z.; Chan, F. K. L.; Kristiansen, K.; Sung, J. J. Y.; Wong, S. H.; Yu, J. Multi-Cohort Analysis of Colorectal Cancer Metagenome Identified Altered Bacteria across Populations and Universal Bacterial Markers. *Microbiome* **2018**, *6* (1), 70. https://doi.org/10.1186/s40168-018-0451-2.

(67) Kunzmann, A. T.; Proença, M. A.; Jordao, H. W.; Jiraskova, K.; Schneiderova, M.; Levy, M.; Liska, V.; Buchler, T.; Vodickova, L.; Vymetalkova, V.; Silva, A. E.; Vodicka, P.; Hughes, D. J. Fusobacterium Nucleatum Tumor DNA Levels Are Associated with Survival in Colorectal Cancer Patients. *Eur. J. Clin. Microbiol. Infect. Dis.* **2019**, *38* (10), 1891–1899. https://doi.org/10.1007/s10096-019-03649-1.

(68) Geller, L. T.; Barzily-Rokni, M.; Danino, T.; Jonas, O. H.; Shental, N.; Nejman, D.; Gavert, N.; Zwang, Y.; Cooper, Z. A.; Shee, K.; Thaiss, C. A.; Reuben, A.; Livny, J.; Avraham, R.; Frederick, D. T.; Ligorio, M.; Chatman, K.; Johnston, S. E.; Mosher, C. M.; Brandis, A.; Fuks, G.; Gurbatri, C.; Gopalakrishnan, V.; Kim, M.; Hurd, M. W.; Katz, M.; Fleming, J.; Maitra, A.; Smith, D. A.; Skalak, M.; Bu, J.; Michaud, M.; Trauger, S. A.; Barshack, I.; Golan, T.; Sandbank, J.; Flaherty, K. T.; Mandinova, A.; Garrett, W. S.; Thayer, S. P.; Ferrone, C. R.; Huttenhower, C.; Bhatia, S. N.; Gevers, D.; Wargo, J. A.; Golub, T. R.; Straussman, R. Potential Role of Intratumor Bacteria in Mediating Tumor Resistance to the Chemotherapeutic Drug Gemcitabine. *Science* **2017**, *357* (6356), 1156–1160. https://doi.org/10.1126/science.aah5043.

(69) Ramos, A.; Hemann, M. T. Drugs, Bugs, and Cancer: Fusobacterium Nucleatum Promotes Chemoresistance in Colorectal Cancer. *Cell* **2017**, *170* (3), 411–413. https://doi.org/10.1016/j.cell.2017.07.018.

(70) Vinogradov, E.; St. Michael, F.; Homma, K.; Sharma, A.; Cox, A. D. Structure of the LPS O-Chain from Fusobacterium Nucleatum Strain 10953, Containing Sialic Acid. *Carbohydr. Res.* **2017**, *440–441*, 38–42. https://doi.org/10.1016/j.carres.2017.01.009.

(71) Di Carluccio, C.; Cerofolini, L.; Moreira, M.; Rosu, F.; Padilla-Cortés, L.; Gheorghita, G. R.; Xu, Z.; Santra, A.; Yu, H.; Yokoyama, S.; Gray, T. E.; St. Laurent, C. D.; Manabe, Y.; Chen, X.; Fukase, K.; Macauley, M. S.; Molinaro, A.; Li, T.; Bensing, B. A.; Marchetti, R.; Gabelica, V.; Fragai, M.; Silipo, A. Molecular Insights into O-Linked Sialoglycans Recognition by the Siglec-Like SLBR-N (SLBR $_{UB10712}$) of *Streptococcus Gordonii*. *ACS Cent. Sci.* **2024**, *10* (2), 447–459. https://doi.org/10.1021/acscentsci.3c01598.

(72) Bensing, B. A.; Li, Q.; Park, D.; Lebrilla, C. B.; Sullam, P. M. Streptococcal Siglec-like Adhesins Recognize Different Subsets of Human Plasma Glycoproteins: Implications for Infective Endocarditis. *Glycobiology* **2018**, *28* (8), 601–611. https://doi.org/10.1093/glycob/cwy052.

(73) Heindel, D. W.; Figueroa Acosta, D. M.; Goff, M.; Yengo, C. K.; Jan, M.; Liu, X.; Wang, X.-H.; Petrova, M. I.; Zhang, M.; Sagar, M.; Barnette, P.; Pandey, S.; Hessell, A. J.; Chan, K.-W.; Kong, X.-P.; Chen, B. K.; Mahal, L. K.; Bensing, B. A.; Hioe, C. E. HIV-1 Interaction with an O-Glycan-Specific Bacterial Lectin Enhances Virus Infectivity and Resistance to Neutralizing Antibodies. *iScience* **2024**, *27* (8), 110390. https://doi.org/10.1016/j.isci.2024.110390.

(74) Di Carluccio, C.; Forgione, R. E.; Bosso, A.; Yokoyama, S.; Manabe, Y.; Pizzo, E.; Molinaro, A.; Fukase, K.; Fragai, M.; Bensing, B. A.; Marchetti, R.; Silipo, A. Molecular Recognition of Sialoglycans by Streptococcal Siglec-like Adhesins: Toward the Shape of Specific Inhibitors. *RSC Chem. Biol.* **2021**, *2* (6), 1618–1630. https://doi.org/10.1039/D1CB00173F.

(75) Koleva, L.; Bovt, E.; Ataullakhanov, F.; Sinauridze, E. Erythrocytes as Carriers: From Drug Delivery to Biosensors. *Pharmaceutics* **2020**, *12* (3), 276. https://doi.org/10.3390/pharmaceutics12030276.

(76) Chen, M.; Leng, Y.; He, C.; Li, X.; Zhao, L.; Qu, Y.; Wu, Y. Red Blood Cells: A Potential Delivery System. *J. Nanobiotechnology* **2023**, *21* (1), 288. https://doi.org/10.1186/s12951-023-02060-5.

(77) Muzykantov, V. R. Drug Delivery by Red Blood Cells: Vascular Carriers Designed by Mother Nature. *Expert Opin. Drug Deliv.* **2010**, *7* (4), 403–427. https://doi.org/10.1517/17425241003610633.
(78) Villa, C. H.; Anselmo, A. C.; Mitragotri, S.; Muzykantov, V. Red Blood Cells: Supercarriers for Drugs, Biologicals, and Nanoparticles and Inspiration for Advanced Delivery Systems. *Adv. Drug Deliv. Rev.* **2016**, *106*, 88–103. https://doi.org/10.1016/j.addr.2016.02.007.
(79) Zhang, X.; Luo, M.; Dastagir, S. R.; Nixon, M.; Khamhoung, A.; Schmidt, A.; Lee, A.; Subbiah, N.; McLaughlin, D. C.; Moore, C. L.; Gribble, M.; Bayhi, N.; Amin, V.; Pepi, R.; Pawar, S.; Lyford, T. J.; Soman, V.; Mellen, J.; Carpenter, C. L.; Turka, L. A.; Wickham, T. J.; Chen, T. F. Engineered Red Blood Cells as an Off-the-Shelf Allogeneic Anti-Tumor Therapeutic. *Nat. Commun.* **2021**, *12* (1), 2637. https://doi.org/10.1038/s41467-021-22898-3.
(80) Dai, J.; Wu, M.; Wang, Q.; Ding, S.; Dong, X.; Xue, L.; Zhu, Q.; Zhou, J.; Xia, F.; Wang, S.; Hong, Y. Red Blood Cell Membrane-Camouflaged Nanoparticles Loaded with AIEgen and Poly(I : C) for Enhanced Tumoral Photodynamic-Immunotherapy. *Natl. Sci. Rev.* **2021**, *8* (6), nwab039. https://doi.org/10.1093/nsr/nwab039.
(81) Wang, C.; Wang, M.; Zhang, Y.; Jia, H.; Chen, B. Cyclic Arginine-Glycine-Aspartic Acid-Modified Red Blood Cells for Drug Delivery: Synthesis and in Vitro Evaluation. *J. Pharm. Anal.* **2022**, *12* (2), 324–331. https://doi.org/10.1016/j.jpha.2021.06.003.
(82) Hou, K.; Zhang, Y.; Bao, M.; Xin, C.; Wei, Z.; Lin, G.; Wang, Z. A Multifunctional Magnetic Red Blood Cell-Mimetic Micromotor for Drug Delivery and Image-Guided Therapy. *ACS Appl. Mater. Interfaces* **2022**, *14* (3), 3825–3837. https://doi.org/10.1021/acsami.1c21331.
(83) Godfrin, Y.; Horand, F.; Franco, R.; Dufour, E.; Kosenko, E.; Bax, B. E.; Banz, A.; Skorokhod, O. A.; Lanao, J. M.; Vitvitsky, V.; Sinauridze, E.; Bourgeaux, V.; Gunter, K. C. International Seminar on the Red Blood Cells as Vehicles for Drugs. *Expert Opin. Biol. Ther.* **2012**, *12* (1), 127–133. https://doi.org/10.1517/14712598.2012.631909.
(84) Bax, B. E. Erythrocytes as Carriers of Therapeutic Enzymes. *Pharmaceutics* **2020**, *12* (5), 435. https://doi.org/10.3390/pharmaceutics12050435.
(85) Rossi, L.; Pierigè, F.; Bregalda, A.; Magnani, M. Preclinical Developments of Enzyme-Loaded Red Blood Cells. *Expert Opin. Drug Deliv.* **2021**, *18* (1), 43–54. https://doi.org/10.1080/17425247.2020.1822320.
(86) Kravtzoff, R.; Colombat, P.; Desbois, I.; Linassier, C.; Muh, J. P.; Philip, T.; Blay, J. Y.; Gardenbas, M.; Poumier-Gaschard, P.; Lamagnere, J. P.; Chassaigne, M.; Ropars, C. Tolerance Evaluation of L -Asparaginase Loaded in Red Blood Cells. *Eur. J. Clin. Pharmacol.* **1996**, *51* (3–4), 221–225. https://doi.org/10.1007/s002280050187.
(87) Hunault-Berger, M.; Leguay, T.; Huguet, F.; Leprêtre, S.; Deconinck, E.; Ojeda-Uribe, M.; Bonmati, C.; Escoffre-Barbe, M.; Bories, P.; Himberlin, C.; Chevallier, P.; Rousselot, P.; Reman, O.; Boulland, M.; Lissandre, S.; Turlure, P.; Bouscary, D.; Sanhes, L.; Legrand, O.; Lafage-Pochitaloff, M.; Béné, M. C.; Liens, D.; Godfrin, Y.; Ifrah, N.; Dombret, H.; Group for Research on Adult Acute Lymphoblastic Leukemia (GRAALL). A P Hase 2 Study of L -asparaginase Encapsulated in Erythrocytes in Elderly Patients with P

Hiladelphia Chromosome Negative Acute Lymphoblastic Leukemia: The GRASPALL/GRAALL-SA 2-2008 Study. *Am. J. Hematol.* **2015**, *90* (9), 811–818. https://doi.org/10.1002/ajh.24093.

(88) Thomas, X.; Le Jeune, C. Erythrocyte Encapsulated L -Asparaginase (Graspa) in Acute Leukemia. *Int. J. Hematol. Oncol.* **2016**, *5* (1), 11–25. https://doi.org/10.2217/ijh-2016-0002.

(89) Kosenko, E. A.; Venediktova, N. I.; Kudryavtsev, A. A.; Ataullakhanov, F. I.; Kaminsky, Y. G.; Felipo, V.; Montoliu, C. Encapsulation of Glutamine Synthetase in Mouse Erythrocytes: A New Procedure for Ammonia Detoxification. *Biochem. Cell Biol.* **2008**, *86* (6), 469–476. https://doi.org/10.1139/O08-134.

(90) Protasov, E. S.; Borsakova, D. V.; Alexandrovich, Y. G.; Korotkov, A. V.; Kosenko, E. A.; Butylin, A. A.; Ataullakhanov, F. I.; Sinauridze, E. I. Erythrocytes as Bioreactors to Decrease Excess Ammonium Concentration in Blood. *Sci. Rep.* **2019**, *9* (1), 1455. https://doi.org/10.1038/s41598-018-37828-5.

(91) Bax, B. E.; Bain, M. D.; Fairbanks, L. D.; Webster, A. D. B.; Ind, P. W.; Hershfield, M. S.; Chalmers, R. A. A 9-yr Evaluation of Carrier Erythrocyte Encapsulated Adenosine Deaminase (ADA) Therapy in a Patient with Adult-type ADA Deficiency. *Eur. J. Haematol.* **2007**, *79* (4), 338–348. https://doi.org/10.1111/j.1600-0609.2007.00927.x.

(92) Bourgeaux, V.; Lanao, J.; Bax, B.; Godfrin, Y. Drug-Loaded Erythrocytes: On the Road toward Marketing Approval. *Drug Des. Devel. Ther.* **2016**, 665. https://doi.org/10.2147/DDDT.S96470.

(93) Williamson, M. P. Chemical Shift Perturbation. In *Modern Magnetic Resonance*; Webb, G. A., Ed.; Springer International Publishing: Cham, 2018; pp 995–1012. https://doi.org/10.1007/978-3-319-28388-3_76.

(94) Williamson, M. P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. https://doi.org/10.1016/j.pnmrs.2013.02.001.

(95) Kleckner, I. R.; Foster, M. P. An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2011**, *1814* (8), 942–968. https://doi.org/10.1016/j.bbapap.2010.10.012.

(96) Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88* (7), 990–994. https://doi.org/10.1021/ed101169t.

(97) Walpole, S.; Monaco, S.; Nepravishta, R.; Angulo, J. STD NMR as a Technique for Ligand Screening and Structural Studies. In *Methods in Enzymology*; Elsevier, 2019; Vol. 615, pp 423–451. https://doi.org/10.1016/bs.mie.2018.08.018.

(98) Hall, J. L.; Sohail, A.; Cabrita, E. J.; Macdonald, C.; Stockner, T.; Sitte, H. H.; Angulo, J.; MacMillan, F. Saturation Transfer Difference NMR on the Integral Trimeric Membrane Transport Protein GltPh Determines Cooperative Substrate Binding. *Sci. Rep.* **2020**, *10* (1), 16483. https://doi.org/10.1038/s41598-020-73443-z.

(99) Fragai, M.; Luchinat, C.; Martelli, T.; Ravera, E.; Sagi, I.; Solomonov, I.; Udi, Y. SSNMR of Biosilica-Entrapped Enzymes Permits an Easy Assessment of Preservation of Native

Conformation in Atomic Detail. *Chem Commun* **2014**, *50* (4), 421–423. https://doi.org/10.1039/C3CC46896H.
(100) Giuntini, S.; Cerofolini, L.; Ravera, E.; Fragai, M.; Luchinat, C. Atomic Structural Details of a Protein Grafted onto Gold Nanoparticles. *Sci. Rep.* **2017**, *7* (1), 17934. https://doi.org/10.1038/s41598-017-18109-z.
(101) Cerofolini, L.; Giuntini, S.; Carlon, A.; Ravera, E.; Calderone, V.; Fragai, M.; Parigi, G.; Luchinat, C. Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. *Chem. – Eur. J.* **2019**, *25* (8), 1984–1991. https://doi.org/10.1002/chem.201804488.
(102) Lecoq, L.; Fogeron, M.-L.; Meier, B. H.; Nassal, M.; Böckmann, A. Solid-State NMR for Studying the Structure and Dynamics of Viral Assemblies. *Viruses* **2020**, *12* (10), 1069. https://doi.org/10.3390/v12101069.
(103) Wiegand, T.; Lacabanne, D.; Torosyan, A.; Boudet, J.; Cadalbert, R.; Allain, F. H.-T.; Meier, B. H.; Böckmann, A. Sedimentation Yields Long-Term Stable Protein Samples as Shown by Solid-State NMR. *Front. Mol. Biosci.* **2020**, *7*, 17. https://doi.org/10.3389/fmolb.2020.00017.
(104) Bax, B. E.; Bain, M. D.; Talbot, P. J.; John Parker-Williams, E.; Chalmers, R. A. Survival of Human Carrier Erythrocytes in Vivo. *Clin. Sci.* **1999**, *96* (2), 171–178. https://doi.org/10.1042/cs0960171.
(105) Xu, Y.; Matthews, S. TROSY NMR Spectroscopy of Large Soluble Proteins. In *Modern NMR Methodology*; Heise, H., Matthews, S., Eds.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 335, pp 97–119. https://doi.org/10.1007/128_2011_228.