How glycosylation turns proteins into molecular precision tools: A Molecular Dynamics study

Silvia D'Andrea

A thesis submitted to Maynooth University in
fulfilment of the requirements for the degree of

Doctor of Philosophy
by
Silvia D'Andrea

Chemistry Department

November 2025

Research supervisor: Prof. Damien Woods

External research supervisor: Dr. Elisa Fadda

Head of Department: Dr. Diego Montagner

Table of Contents

Acknowledgements

First, I would like to sincerely thank my supervisors, Prof. Damien Woods and Dr. Elisa Fadda, for giving me the opportunity to undertake this project and for welcoming me as a PhD student. I am especially grateful to Elisa, to whom I wish to express my deepest gratitude. Your guidance and support throughout these years have provided me with a profound scientific and personal growth that I will carry with me into my future career. Your steady confidence in me, together with the example you set through your dedication and approach to science, has been a constant source of motivation and strength throughout this journey.

I am also grateful to Maynooth University, the Department of Chemistry, and the Hamilton Institute for providing a welcoming and stimulating environment in which to carry out this work.

I would like to thank past and present colleagues for creating a supportive and inspiring environment in which to grow as a researcher. In particular, I am grateful to Carl, Callum, Akash and Lucy for their help and for all the discussions we shared over the years. A very special thank you goes to Ojas, for giving me the opportunity to look at science from his unique perspective, and to Beatrice, for being such a precious colleague and friend. From the very beginning, I admired the way you approached this world, your perseverance, and your sensitivity; your presence was fundamental during the most difficult times, and even more for being there to celebrate our successes together.

I am profoundly grateful to my parents, my brothers, and my lifelong friends for their unconditional support, patience, and encouragement at every stage of this journey.

My heartfelt thanks also go to my “Dublin family”, who made this experience richer, lighter, and full of memories that I will always carry with me; without them, this adventure would not have been the same. Thanks to you, I never truly felt far from home.

Finally, my deepest thanks go to Antonio, my greatest source of strength and support, for more reasons than I could ever fully express here.

Abbreviations

A2	Biantennary N-glycan structure
A2B	Bisected biantennary N-glycan structure
A2M	Alpha-2-Macroglobulin
A3	Triantennary N-glycan structure
Aβ	Amyloid beta
AF2	AlphaFold-2 protein structure prediction software
AMBER	Assisted Model Building with Energy Refinement
ApoE	Apolipoprotein E
β4GalT1	β4-galactosyltransferase 1
CD52	Glycopeptide (CAMPATH-1 antigen)
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CG	Coarse Graining
CHL	Cholesterol
CR	Cysteine-Rich repeats
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DSPC	1,2-distearoyl-sn-glycero-3-phosphocholine
ER	Endoplasmic Reticulum
FUT8	α1,6-fucosyltransferase
GAFF	General Amber Force Field
GALNT	UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase
GAP	Glycan Analysis Pipeline
GDB	GlycoShape 3D Glycan Structure Database
GHs	Glycoside Hydrolases
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	GROningen MOlecular Simulation
GTs	Glycosyltransferases
HMGB1	High Mobility Group Box Protein 1
HPC	High-Performance Computing
IMS	Ion Mobility Separation
ITIM	Immunoreceptor Tyrosine-Based Inhibitory Motif
KDE	Kernel Density Estimates

LRP1	Low-Density Lipoprotein Receptor-Related Protein 1
MAG	Myelin-Associated Glycoprotein(Siglec-4)
MD	Molecular Dynamics
MGAT	N-acetylglucosaminyltransferases
N-glycan	N-linked glycan (attached to asparagine)
ND	Nanodiscs
NMR	Nuclear Magnetic Resonance
Neu5Ac	N-acetylneuraminic acid
OST	Oligosaccharyltransferase
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PDB	Protein Data Bank
PTMs	Post-Translational Modifications
RAP	Receptor-Associated Protein
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
Rg	Radius of Gyration
SNFG	Symbol Nomenclature for Glycans
Siglecs	Sialic acid-binding immunoglobulin-like lectins
TIP3P	Three-point interacting water model (TIP3P)
VMD	Visual Molecular Dynamics
cryo-EM	Cryo-Electron Microscopy
nMS	Native Mass Spectrometry

Declaration

This thesis has not been submitted before, in whole or part, to this University for any degree, and is, except where otherwise stated, the original work of the author.

This thesis has been prepared in accordance with the PhD regulations of Maynooth University and is subject to copyright. For more information see PhD Regulations .

Silvia D'Andrea

Abstract

Glycosylation is a complex and ubiquitous post-translational modification that critically contributes to the structural and functional diversity of proteins and lipids. It occurs when glycans are enzymatically linked to specific amino acid residues on proteins, influencing folding, structural stability and function. In glycoconjugates, sialylation has been often shown to be an essential feature for molecular recognition and immune regulation, meanwhile altered sialylation patterns have been linked to cancer, infectious diseases, and neurological and inflammatory disorders. The molecular factors that determine whether sialylation has activating or inhibitory effects remain poorly defined. This thesis addresses this knowledge gap by exploring at the atomistic level of detail the impact of sialylation in a variety of molecular systems linked to different biological pathways and responses, key to human health and disease. To do this I used molecular dynamics (MD) simulations, a powerful computational approach that allows users to determine the structure, dynamics and energetics features underpinning molecular recognition and function through sampling and statistical thermodynamics principles. The results of the MD simulations I present are used to guide and/or to support experimental methods, leading to a common, clear rationale of the biological event examined. Through the studies I present in this thesis, I demonstrate that the effects of sialylation depend on its position, whether on receptors, peptide ligands, or glycolipids, and on its mode of presentation, such as monovalent or multivalent, and in soluble or in membrane-anchored constructs. I show that sialic acids can act as local inhibitors, for example by introducing steric hindrance that reshapes ligand preference at the endocytic receptor LRP1, which is relevant to neurodegenerative disease. Conversely, hypersialylation can function as a structural switch that promotes and stabilises high-avidity interactions, as observed in the CD52–HMGB1–Siglec-10 immune signalling pathway. Recognition by immunomodulatory receptors, such as Siglec-6, depends on the combined influence of the sialylated epitope and its orientation with respect to the lipid membrane. Sialylation of N-glycans is ultimately controlled by upstream modifications, such as the bisecting GlcNAc motif, which, as a stop codon, can halt further glycan maturation.

By providing an atomistic view, the work in this thesis reveals that sialylation behaves in complex and context-dependent ways. When factors such as local environment, valency and the 3D shape of the sialylated epitope are considered, sialic acids emerge as active

determinants of recognition rather than passive terminal caps. These results link the crucial role of sialylation to disease-relevant mechanisms and offer a structural framework to guide future glycan-focused therapeutic strategies.

List of Publications

- Ives, C. M., Singh, O., **D'Andrea**, S., Fogarty, C. A., Harbison, A. M., Satheesan, A., Tropea, B., Fadda, E. (2024). Restoring protein glycosylation with GlycoShape. *Nature Methods*, *21*(11), 2117-2127.

- DeBono, N. J., **D'Andrea**, S., Bandala-Sanchez, E., Goddard-Borger, E., Zenaidee, M. A., Moh, E. S., Fadda, E., Harrison, L.C., Packer, N. H. (2025). The molecular basis of immunosuppression by soluble CD52 is defined by interactions of N-linked and O-linked glycans with HMGB1 box B. *Journal of Biological Chemistry*, 301(4), 108350.

- **D'Andrea**, S., Schmidt, E. N., Bui, D., Singh, O., Han, L., Mahal, L. K., Klassen, J., S., Macauley, S.,M., & Fadda, E. (2025). Glycolipid recognition and binding by Siglec-6 hinges on interactions with the cell membrane. *bioRxiv*, 2025-06. *Communication Biology* (*in press*)

- Sobczak, K., Antoñana-Vildosola, A., Valverde, P., Travecedo, M. A., Jame-Chernaboo, Z., Schmidt, E. N., **D'Andrea, S**., Valdaliso-Díez, E., Oyenarte, I., Laugieri, M. E., Joe, M., Mozaneh, F., Lin, S.-Y., Bosch, A., Moure, M. J., Franconetti, A., Lee, S. Y., Etxaniz-Díaz de Durana, J., Pérez-Gutierrez, L., Palazón, A., Marcelo, F., Fadda, E., Corzana, F., Gimeno, A., Macauley, M. S., Jiménez-Barbero, J., & Ereño-Orbea, J. (2025). The unique molecular recognition features of Siglec-10: structural insights into sialoglycan and antibody interactions. *bioRxiv*, 2025-06. **doi:** https://doi.org/10.1101/2025.06.10.658867

- Hintze, J., Topaktas, A. B., Daugbjerg, T. M., Jebari, S., **D'Andrea, S**., de Haan, N., Hansen, L. H., Baars, R., Quittot, N., Martin, C., Yang, Z., Vakhrushev, S. Y., Miller, R. L., Strickland, D. K., Hyman, B., Fadda, E., & Schjoldager, K. T. (2025). The ligand preference of LRP1 is regulated by O-glycans. *bioRxiv*, 2025-10. doi: https://doi.org/10.1101/2025.10.23.68410

Chapter 1. Introduction

Complex carbohydrates (glycans) are one of the four major classes of biological macromolecules, together with nucleic acids, proteins, and lipids. These macromolecules are essential for life and at the core of biology and medicine[1]. Although proteins, lipids, and nucleic acids have been central to structural and molecular biology teaching and research, glycans have received less attention in the past, as they are non-templated (heterogeneous) and structurally flexible, while they are crucial to all processes regulating cellular function in health and disease[2,3].

Monosaccharides are the basic units of glycans and chemically are polyhydroxylated aldehydes or ketones, present in nature primarily in their cyclic (closed) form. Monosaccharides can be classified into two configurations, D and L, depending on the orientation of the hydroxyl group at the chiral centre farthest from the carbonyl carbon. In aqueous solution, monosaccharides mainly adopt cyclic hemiacetal or hemiketal forms, produced by intramolecular nucleophilic attack of a hydroxyl group on the carbonyl carbon (C1 in aldoses and typically C2 in ketoses). The ring closure forms a new stereocentre, known as anomeric carbon, resulting in either five-membered (furanose) or six-membered (pyranose) rings. In their cyclic form, monosaccharides can exist in two anomeric configurations, α or β, defined by the orientation of the hydroxyl group at the anomeric carbon[4].

According to the International Union of Pure and Applied Chemistry (IUPAC) definition, glycans are compounds consisting of monosaccharides linked by glycosidic bonds[5]. Each monosaccharide has several hydroxyl groups that can serve as acceptor sites for another monosaccharide, which usually donates its anomeric centre to form the glycosidic linkage. The identity of the monosaccharides, their furanose or pyranose ring forms, and the position and α/β configuration of each glycosidic bond together enable even a small set of monosaccharides to generate an enormous variety of linear and branched oligosaccharide structures. To address this structural complexity, standard symbolic notations, such as the Symbol Nomenclature for Glycans (SNFG)[6], are used to represent glycans (**Figure 1**).

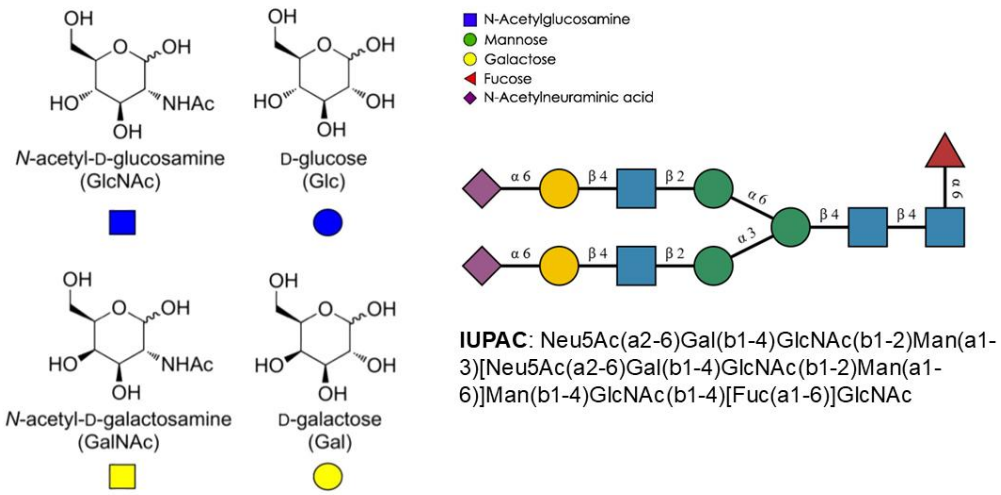


Figure 1.1 *Left*: ring structures of common mammalian monosaccharides (GlcNAc, Glc, GalNAc and Gal) alongside their corresponding SNFG symbols. *Right*: example of a biantennary sialylated N-glycan drawn using SNFG notation, with the corresponding International Union of Pure and Applied Chemistry (IUPAC)-condensed sequence shown below.

Glycans are typically found as components of larger glycoconjugates, including proteoglycans, glycoproteins containing N- and O-linked glycans, glycosphingolipids such as gangliosides, and glycosylphosphatidylinositol (GPI) anchors[7,8]. In mammals, the largest concentration of free (unlinked) glycans are found in milk with a function and diversity linked to raise immunity[9]. Glycoconjugates form the glycocalyx on the cell surface, where glycans play critical roles in regulating receptor–ligand recognition, cell adhesion, signalling, and immune surveillance[10]. Recognising glycans as a major class of biological macromolecules is therefore essential for developing accurate molecular descriptions of cellular processes[4,11].

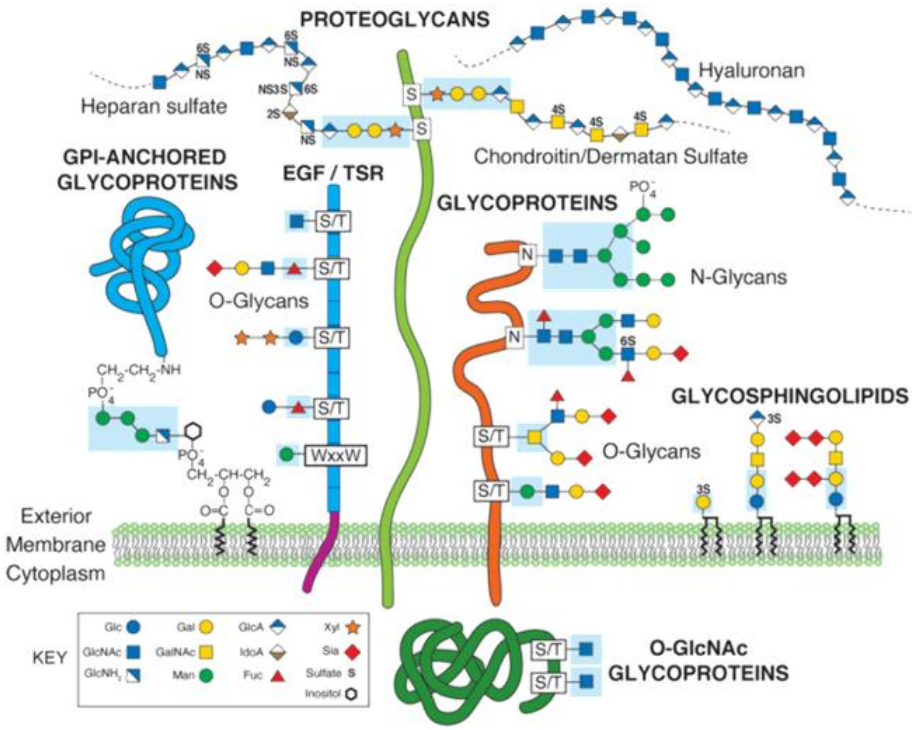


Figure 1.2. Major classes of cell-surface glycoconjugates. Proteoglycans, GPI-anchored glycoproteins, N- and O-glycoproteins, glycosphingolipids, and O-GlcNAc-modified glycoproteins together contribute to the glycocalyx at the plasma membrane. Adapted from *Essentials of Glycobiology*, 4th edition (Varki et al., 2022).

Glycosylation is among the most abundant and functionally important post-translational modifications (PTMs)[12]. Protein glycosylation is broadly classified as N-linked or O-linked, while glycolipids, including gangliosides, form a major class of glycosylated lipids[13]. In N-glycosylation, a pre-assembled oligosaccharide is transferred as a unit to asparagine residues within a specific sequon, i.e. Asn-X(any aa but Pro)-Ser/Thr, in the endoplasmic reticulum (ER)[13,14]. This immature glycan is then extensively remodeled along the ER and through the Golgi by enzymatically-driven trimming and elongation, producing a range of oligomannose, hybrid, and complex structures, see **Figure 1.3.a.** In O-GalNAc glycosylation, the most common type of O-glycosylation in mammals, monosaccharides are added stepwise to serine or threonine residues in the Golgi. O-GalNAc glycans have various core structures that can be further elongated and capped with fucose, galactose, N-acetylglucosamine, and sialic acids, see **Figure 1.3.b.**

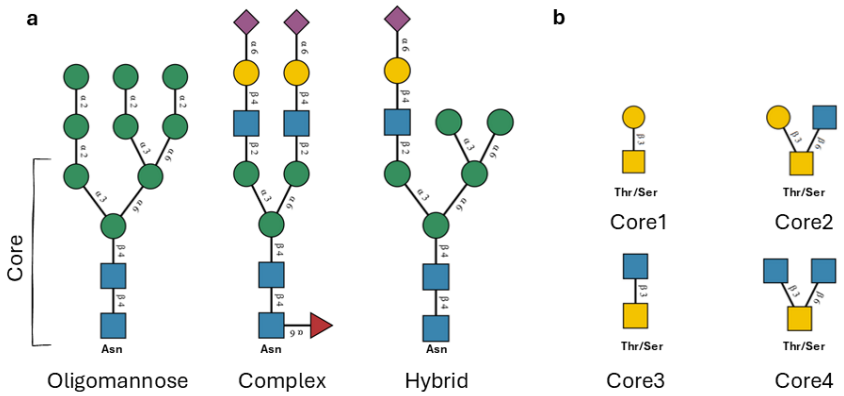


Figure 1.3. Representative N- and O-glycan core structures shown with SNFG notation. **a.** Oligomannose, complex, and hybrid N-glycans all share a conserved $GlcNAc_2Man_3$ core linked to asparagine (Asn), but differ in the composition and branching of their outer arms. **b.** Major mucin-type (O-GalNAc) glycan core structures (cores 1–4) linked to Ser/Thr residues.

Although glycans are abundant and functionally important, they are often difficult to characterise structurally. Their intrinsic flexibility and the heterogeneity make determining their 3D structure challenging[15]. The conformation of each glycosidic linkage is primarily defined by a set of torsion (dihedral) angles (φ, ψ and, when relevant, ω; **Figure 1.4**), which populate multiple low-energy minima separated by energy barriers interconvertible at room temperature. As a result, glycan chains exist as different structure (3D) arrangements, called conformational ensembles, that are difficult or impossible to capture experimentally[16].

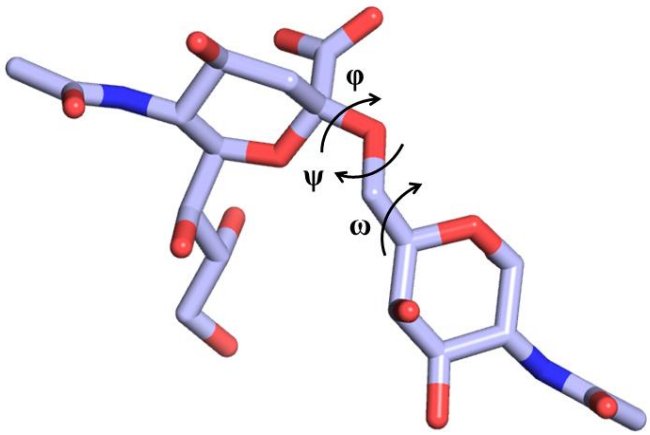


Figure1.4. Example of DNeup5Aca2-6DGalpNAc disaccharide in 3D representation illustrating the glycosidic torsion angles ϕ, ψ, and ω that define the conformation of the glycosidic linkage and underline the intrinsic flexibility of glycan chains.

Consequently, glycans present in protein structures deposited in the RCSB Protein Data Bank (PDB) are largely incomplete, truncated, poorly resolved, but more often omitted entirely. As a result, even high-resolution structures of glycoproteins often under-represent the glycan component or lack sufficient detail to capture its dynamic behaviour. However, it is this dynamic behaviour, i.e. how glycans move, reorganise, and transiently interact with neighbouring residues, lipids, or receptors, that is key to their biological functions.

All-atom molecular dynamics (MD) simulations offer a powerful approach to address these challenges[17]. MD simulations explicitly represent proteins, glycans, lipids, and solvent, tracking their motion over time through the integration of Newton's equation of motion, where the potential energy that leads to the force propelling the atoms, is represented by an appropriately chosen force field. This simulation method provides access to conformational ensembles and timescales beyond the reach of most biochemical or structural techniques[18], revealing how glycan flexibility, microheterogeneity and the local environment influence molecular recognition. In this thesis, MD simulations are the primary research tool I use for investigating how different glycosylation (and sialylation) patterns affect the structure and interactions of glycoproteins and glycolipids[19,20].

Sialic acids are central to immune regulation and cell–cell communication among vertebrate glycans[21,3]. N-acetylneuraminic acid (Neu5Ac) and related sialic acids are typically found as terminal monosaccharides in glycans, attached via α2–3, α2–6, or α2–8 linkages to galactose or other sialic acid residues[22]. As the terminal residue of N-glycans, O-glycans, and gangliosides, sialic acids serve as key determinants of molecular

recognition[23]. Sialylation is closely linked to sialyltransferase and sialidase activity in the glycosylation pathway, where adding a terminal sialic acid can halt further elongation of specific glycan arms[24]. In the Golgi, sialic acid is transferred to terminal galactose residues by specific sialyltransferases, such as ST3Gal and ST6Gal isoforms[25]. In mammalian N-glycans sialic acid can be only added to galactose, and thus the presence of a terminal galactose determines whether an arm can be sialylated. Once galactose is capped with sialic acid, that branch cannot undergo further elongation[26]. Therefore, if an N-glycan arm lacks terminal galactose, it cannot be sialylated: no galactose, no sialic acid.

1.1 Project aims

The central focus of this thesis is sialylation and its role within different glycoforms and constructs. I will show that sialylation does not play a single universal role, yet that it can inhibit or promote molecular recognition depending on where it is within a glycan structure, in how many units, i.e. one or more sialic acids in a single glycan, and which receptor surface it engages with. Sialic acids can therefore act as local inhibitors that block productive contacts, as structural switches that reshape a peptide or glycan epitope, or as multivalent, negatively charged keys that engage high-avidity binding sites on immune receptors. Collectively, the systems examined in this thesis address a central question: how does sialylation, in distinct molecular contexts, control recognition by specific receptors? After a detailed description of the computational methods in **Chapter 2**, the thesis presents a series of case studies illustrating how sialylation tunes molecular recognition in different contexts.

Chapter 3 focuses on the endocytic receptor low-density lipoprotein receptor–related protein 1 (LRP1) and examines how a core-1 sialylated O-glycan modulates receptor–ligand binding. The LRP1 is a critical multiligand receptor involved in the clearance of amyloid beta (Aβ) and the trafficking of tau protein, relevant to neurodegenerative disorders. Experimental work by our collaborators[27] revealed that the lack of GALNT11-dependent O-glycans, which are atypically hyposialylated, on LRP1 CR linkers, led to opposite ligand-specific effects: a decrease in tau uptake and an increase in Aβ uptake. Using RAP as a ligand model, we show through MD simulations of the LRP1:RAP interface that the non-sialylated Core-1 O-glycan actively modulates the structure and dynamics of the interface. This O-glycan folds back toward CR6, enabling transient intramolecular interactions (CH–π stacking with Trp63) and forming hydrogen bonds

with RAPd1 residues (Glu23/Glu30). Conversely, sialylation of core-1 O-glycan acts as a local inhibitor. The added steric hindrance and negative charge prevent these accessory contacts, thereby blocking the glycan from engaging the interface and keeping the linker solvent-exposed. This hyposialylation-driven mechanism is proposed to finely modulate the receptor's ligand preference. Here, sialylation acts as a negative modulator by introducing steric clashes that inhibit receptor and ligand recognition.

Chapter 4 explores a contrasting scenario, where an N-glycan and an O-glycan in the active form of the CD52 glycopeptide are both hyper-sialylated. In the context of the CD52–HMGB1–Siglec-10 immune signalling pathway, sialylation acts as a structural switch that enables and stabilizes high-avidity interactions. We find that hyper-sialylation of the soluble CD52 glycopeptide induces a conformational shift towards more extended states. This extended architecture promotes the direct and stable binding of the doubly sialylated core-2 O-glycan at T8 to HMGB1 Box B. This binding exhibits selectivity for Box B over the highly similar Box A domain, explained by key non-conserved residues in Box A. The formation of this complex exposes the hypersialylated N-glycan at Arg3 toward the solvent, making the epitope accessible for recognition by the inhibitory receptor Siglec-10. MD simulations further revealed that Siglec-10 uniquely utilizes a dual-arginine in the binding pocket (the canonical Arg119 and the non-canonical Arg127) to cooperatively engage multivalent sialylated epitopes, explaining its preference for complex, hypersialylated ligands.

Chapter 5 places these findings in the broader context of the Siglec family, a group of immunomodulatory receptors that recognise sialylated epitopes via a conserved arginine in their binding (V-set) domain. In this chapter, the focus shifts from glycoproteins to membrane-anchored ganglioside headgroups. By examining sialylated gangliosides as ligands for Siglec-6, I show that a terminal sialic acid is necessary but not sufficient for binding. Productive engagement requires insertion of the Trp127 side chain into the lipid bilayer, supported by Lys126, which orients the V-set domain relative to the membrane. Within this geometry, Siglec-6 selectively recognises GM1 because its terminal Gal can engage the flexible C–C′ loop while maintaining membrane contact, a configuration that GM2 cannot achieve. This chapter therefore illustrates how Siglec-6 recognition of a similar sialylated epitope is defined by the combined contributions of ganglioside headgroup orientation and the Siglec-6–membrane interface.

Chapter 6 examines the N-glycosylation pathway to assess how the presence of bisecting N-acetylglucosamine (GlcNAc) affects the localization of sialylated epitopes on N-glycans. Because sialic acids are incorporated at the final stage onto terminal galactose residues, modifications that alter galactosyltransferase accessibility indirectly shape sialylation patterns. The addition of a bisecting GlcNAc by MGAT3 restricts arm elongation in bisected N-glycans and can prevent the formation of terminal galactose residues, which are necessary for sialyltransferase activity. Thus, the bisecting GlcNAc serves as a molecular "stop codon" for sialylation on specific branches. This mechanism was clarified by comparing a bisected biantennary (A2B) N-glycan with a non-bisected triantennary (A3) structure. Structural alignment of this A2B conformer with key maturation enzymes, α1,6-fucosyltransferase (FUT8) and β4-galactosyltransferase 1 (β4GalT1), showed that the bisecting GlcNAc creates steric clashes that prevent effective substrate recognition by both enzymes. This structural constraint explains why bisection suppresses further N-glycan maturation and limits the number and distribution of terminal galactose residues available for sialylation. Overall, the results demonstrate that branching and bisection act upstream of sialylation to define where terminal sialic acids can be displayed on the N-glycans.

In the final chapter, **Chapter 7**, I introduce the GlycoShape 3D database (https://glycoshape.org), which I contributed to create and has supported my PhD research from its initial creation through ongoing development and refinement. GlycoShape offers curated 3D collections of biologically important glycans from the analysis of enhanced sampling MD simulation and through the tool Re-Glyco serves as the main tool for building, visualizing, and analyzing the glycan structures explored in my works.

By combining all-atom MD simulations across these systems, this thesis demonstrates that sialylation does not serve a single universal role. Instead, it acts in distinct molecular contexts, functioning either as a local inhibitor that blocks productive contacts or as a structural modulator that reshapes an epitope to promote and stabilize high-avidity binding. These findings show that the role of sialylation in recognition depends on its location, relative 3D orientation, number of units, and the specific receptor involved.

References

1. Böhm M, Bohne-Lang A, Frank M, Loss A, Rojas-Macias MA, Lütteke T. Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update). Nucleic Acids Res. 2019 Jan 8;47(D1):D1195–201.
2. Structural insights into host–microbe glycointeractions. Current Opinion in Structural Biology. 2022 Apr 1;73:102337.
3. Varki A. Biological roles of glycans. Glycobiology. 2017 Jan;27(1):3–49.
4. Seeberger PH. Monosaccharide Diversity. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
5. Nomenclature of Carbohydrates: Recommendations 1996.
6. Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lütteke T, et al. Updates to the Symbol Nomenclature for Glycans guidelines. Glycobiology. 2019 Aug 20;29(9):620–4.
7. Varki A, Cold Spring Harbor Laboratory Press. Essentials of Glycobiology. 2022.
8. Zielinska DF, Gnad F, Wiśniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell. 2010 May 28;141(5):897–907.
9. Slater AS, Hickey RM, Davey GP. Interactions of human milk oligosaccharides with the immune system. Front Immunol. 2025 Jan 14;15:1523829.
10. Hudak JE, Canham SM, Bertozzi CR. Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. Nat Chem Biol. 2014 Jan;10(1):69–75.
11. Bowman K, Friedman D. Glycoscience: integrating a key macromolecule more fully into the curriculum. CBE Life Sci Educ. 2013 Spring;12(1):5–8.
12. Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. Cell. 2006 Sept 8;126(5):855–67.
13. Schnaar RL. Gangliosides of the Vertebrate Nervous System. J Mol Biol. 2016 Aug 14;428(16):3325–36.
14. Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
15. Grant OC, Woods RJ. Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. Curr Opin Struct Biol. 2014 Oct;28:47–55.
16. Alibay I, Burusco KK, Bruce NJ, Bryce RA. Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. J Phys Chem B. 2018 Mar 8;122(9):2462–74.
17. Alocci D, Mariethoz J, Gastaldello A, Gasteiger E, Karlsson NG, Kolarich D, et al. GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical. J Proteome Res. 2019 Feb 1;18(2):664–77.
18. Dror RO, Jensen MØ, Borhani DW, Shaw DE. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. J Gen Physiol. 2010 June 1;135(6):555–62.
19. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018 Sept 19;99(6):1129–43.
20. Fadda E, Woods RJ. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. Drug Discov Today. 2010 Aug;15(15-16):596–609.

21. Varki A, Gagneux P. Multifarious roles of sialic acids in immunity. Ann N Y Acad Sci. 2012 Apr;1253(1):16–36.
22. Varki A, Schauer R. Sialic Acids. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009.
23. Soares CO, Grosso AS, Ereño-Orbea J, Coelho H, Marcelo F. Molecular Recognition Insights of Sialic Acid Glycans by Distinct Receptors Unveiled by NMR and Molecular Modeling. Front Mol Biosci. 2021 Nov 15;8:727847.
24. Zhu W, Zhou Y, Guo L, Feng S. Biological function of sialic acid and sialylation in human health and disease. Cell Death Discov. 2024 Sept 30;10(1):415.
25. Lopez Aguilar A, Meng L, Hou X, Li W, Moremen KW, Wu P. Sialyltransferase-Based Chemoenzymatic Histology for the Detection of N- and O-Glycans. Bioconjug Chem. 2018 Apr 18;29(4):1231–9.
26. Ramya TNC, Weerapana E, Cravatt BF, Paulson JC. Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. Glycobiology. 2013 Feb;23(2):211–21.
27. Hintze J, Topaktas AB, Daugbjerg TM. , Shifa Jebari, D'Andrea S, Noortje de Haan, Lasse H. Hansen, Rob Baars, Noé Quittot, Cesar Martin, Zhang Yang, Sergey Y. Vakhrushev, Rebecca L. Miller, Dudley K. Strickland, Bradley Hyman, Elisa Fadda, Katrine T. Schjoldager The Ligand Preference of LRP1 is Regulated by O-glycans.doi: https://doi.org/10.1101/2025.10.23.684103 bioRxiv. (2025)

Chapter 2. Computational methods

The work presented in this thesis is based primarily on molecular dynamics (MD) simulations to examine how N- and O-glycosylation influence protein structure and ligand–receptor recognition, focusing on immune receptors and related signalling pathways. The findings offer insight into how glycan and glycopeptide recognition affect immunity and other biological processes.

Glycans are highly flexible and chemically heterogeneous molecules, and their conformational ensembles are notoriously difficult to characterise experimentally[1,2]. Techniques such as X-ray crystallography and cryo-EM typically yield only static snapshots, in which glycans are often only partially resolved or not visible[3]. NMR and other spectroscopic methods yield averaged data that can be challenging to interpret structurally[4].

Classical MD simulations can circumvent these limitations by explicitly modelling atomic motion over time and across phase space, offering atomistic views of glycoproteins and glycolipids in near-physiological conditions[1]. MD is well-suited to study glycan conformational preferences, their coupling to protein dynamics, and the mechanisms of glycan-mediated recognition[1].

2.1 Theory Foundation of Molecular Dynamics Simulations

MD simulations have become increasingly common and a useful computational method for studying the structure, dynamics, and energetics of biomolecular interactions at the atomistic level of details[5]. MD approaches largely make use of classical mechanics as it allows to cast very complex molecular systems in a relatively simple framework, which in turn allows us to perform simulations and collect statistical sampling on thousands to millions of atoms, with the technology available to date. In classical mechanics, atoms are approximated as hard spheres with precise masses, radii, and finite charges and dispersion potentials. Their motion is governed by Newton's equations of motion[6]. For each atom, the force, F, acting on it determines its acceleration a according to Newton's second law (**Eq. 2.1**):

$$\vec{F} = m\vec{a} \quad (1)$$

where *m* is the mass of the atom. The force can also be expressed as the negative gradient of the potential energy with respect to the atomic coordinates (**Eq. 2.2**):

$$\vec{F} = -\frac{\partial V(r)}{\partial \vec{r}} \tag{2}$$

where *V(r)* is the potential energy of the system as a function of the positions *r* of all atoms. Within the classical mechanics approximation, the potential energy can be defined by an empirical force field, generally expressed by the functional form in **Eq. 3:**

$$\begin{aligned} V(r) = & \sum_{\text{bonds}} k_b (r - r_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 \\ & + \sum_{\text{torsions}} \sum_{n} \frac{V_n}{2} \left[1 + \cos(n\omega - \gamma)\right] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\} \end{aligned} \tag{3}$$

The potential energy of the molecule, V(r), depends on the positions of the *N* atoms in the system and is calculated as the sum of potential energy functions from bonded (covalent) and non-bonded (non-covalent) interactions, as shown in **Eq. 4**.

F

$$E_{system} = E_{bonded} + E_{non-bonded} \tag{4}$$

The bonded contribution, E_{bonded} includes potentials for bond stretching, angle bending, and torsional (dihedral) rotations, usually represented by harmonic and periodic functions.

The non-bonded term, $E_{non-bonded}$, accounts for van der Waals interactions, typically modelled with a Lennard-Jones potential, and electrostatic interactions, modelled with a Coulomb potential. The quality of an MD simulation therefore depends critically on both the functional form of the force field and the accuracy of its parameters.

Additive, or non-polarisable, force fields are the most widely used in MD simulations[6,7]; in these, partial charges are fixed for each atom. In polarisable force fields, partial charges are adjusted to explicitly account for electron polarisation effects from the surrounding molecular environment. Several general biomolecular force fields have been developed, including AMBER[8], CHARMM[9], and GROMOS[9,10], each with distinct parameterisation

strategies and balances of bonded and non-bonded terms. For glycans and glycoconjugates, the most widely used parameter sets are AMBER/GLYCAM06[11] and CHARMM/CHARMM36[12]. GLYCAM06 is specifically designed for carbohydrates and, while complementary to AMBER, can be used independently with other force fields[11]. CHARMM36 provides specialised parameters for glycans, including glycosidic and glycoprotein linkages, and is compatible with the broader CHARMM all-atom force fields, supporting the modelling of glycoproteins, glycolipids, and other glycan-containing systems. Other carbohydrate force fields include GROMOS[13], which uses a united-atom approach, OPLS-AA, an all-atom parameter set, and DRUDE, a polarisable force field that extends CHARMM to include electronic polarisation effects[13,14]. The parameterisation protocols for these force fields are summarised in **Figure 2.1.**

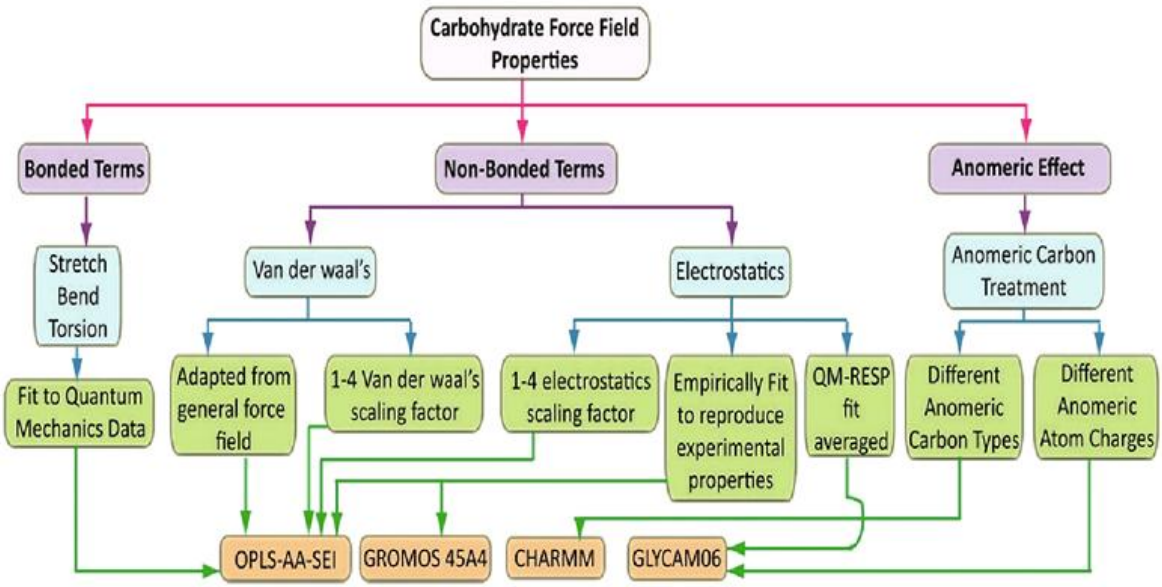


Figure 2.1: Parameterisation protocols for various carbohydrate forcefields. Adapted from reference (Perez *et al*. 2021).

In this thesis, I primarily used AMBER type force fields for soluble glycoprotein systems and CHARMM36 for membrane-embedded systems. Proteins are described with the AMBER ff14SB or AMBER19SB force fields, while glycans are represented using GLYCAM06, which is optimised for N-, O-, and glycolipid-linked carbohydrates. For all simulations that involved a membrane, I used the CHARMM36 parameter set for all species.

Although additive force fields are highly successful, they have limitations. For example, they may underestimate weak interactions, such as CH–π stacking, and can over-stabilise

glycan–glycan contacts[1]. It is important to be aware of these limitations when running simulations as these can be a source of bias.

Once the force field has been selected, the other key ingredient of an MD simulation is time. The potential energy and forces tell us how atoms tend to move, but we still need a practical way to follow how their positions change. Because Newton's equations of motion cannot be solved exactly for many-atom systems, time is divided into small steps Δt and the trajectory is computed step by step using numerical integration schemes. In practice, MD trajectories are generated by integrating Newton's equations of motion (**Eq. 2.2**) with Verlet-type integrators (Verlet, 1967). Two commonly used variants are the basic Verlet integrator and its leapfrog form, shown in **Eqs. 5** and **6**, respectively.

(5)

$$r(t+\delta t) = r(t) + v(t)\,\delta t + \frac{1}{2}a(t)\,\delta t^2$$

$$r(t-\delta t) = r(t) - v(t)\,\delta t + \frac{1}{2}a(t)\,\delta t^2$$

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\,\delta t^2$$

(6)

$$r(t+\delta t) = r(t) + v\left(t+\frac{1}{2}\delta t\right)\delta t$$

$$v\left(t+\frac{1}{2}\delta t\right) = v\left(t-\frac{1}{2}\delta t\right) + a(t)\,\delta t$$

The Verlet integrator shown in **Eq. 5** updates atomic positions at time t+δt using the current acceleration at time *t* and previous positions at time *t* - δt. The basic Verlet integrator does not directly compute or store velocities, which can lead to inaccuracies over time. The leapfrog integrator shown in **Eq. 6**, addresses this issue by computing velocities at half-time steps $t + ½\ \delta t$ and using them to compute positions at full-time steps $t + \delta t$. This makes velocities directly available at each step, which is convenient for evaluating kinetic energy and controlling the temperature of the system.

The choice of integration time step Δt is critical, because it must be much shorter than the fastest motion in the system. In biomolecules, the fastest motions are bond vibrations involving hydrogen atoms (e.g. C–H), which have characteristic periods of about 10 fs.

Without any constraints, this limits the time step to around 1 fs. By constraining all bonds involving hydrogen atoms using algorithms such as SHAKE[15] or LINCS[16], these high-frequency vibrations are effectively removed. This allows the time step to be increased to 2 fs without loss of accuracy, which in turn reduces the number of integration steps needed to reach microsecond timescales and makes the simulations computationally more efficient.

2.2 Simulation set-up

In this section, I outline the general workflow for setting up MD simulations of glycoproteins under near-physiological conditions on high-performance computing (HPC) resources. The procedure is summarised schematically in **Figure 2.2** and consists of four main stages: (i) structure generation, (ii) parametrisation, (iii) system building, and (iv) minimisation, equilibration, and production.

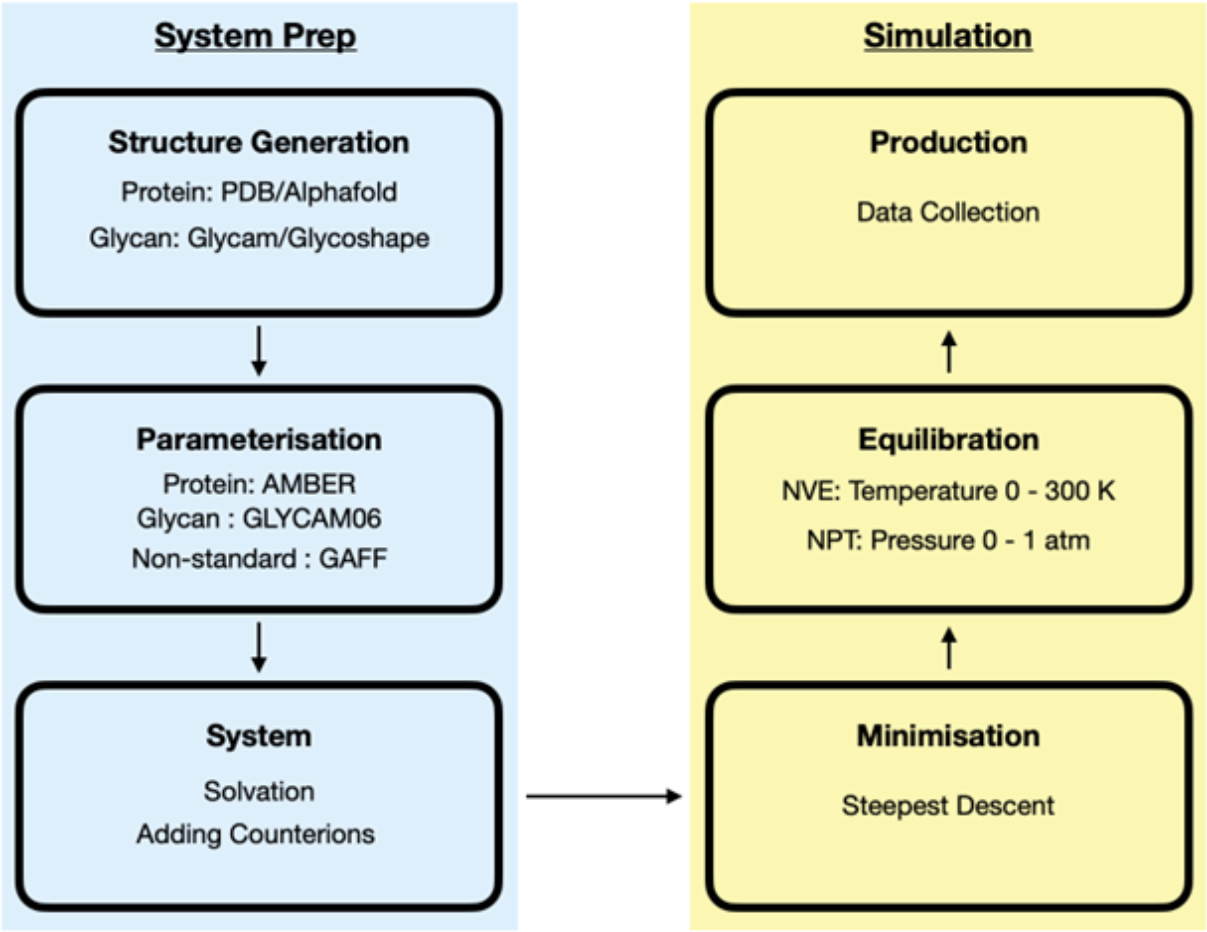


Figure 2.2: Workflow for setting up and running an MD simulation of a glycoprotein.

2.2.1 Structure generation

The first step is to generate reliable three-dimensional structures for both the protein and the glycan(s) of interest. For proteins, experimental structures can be sourced from the RCSB Protein Data Bank (PDB)[17]. When suitable experimental structures are unavailable or incomplete, 3D models can be generated from the amino acid sequence using AlphaFold[18,19]. In both cases, the quality of the starting structure must be critically evaluated in the context of the biological question the simulation addresses. For crystal structures, this includes checking the resolution and B-factors, inspecting the completeness of loops and termini, and verifying that key functional residues and domains are present. For AlphaFold models, local confidence scores (pLDDT) are examined, and, where possible, the models are compared with experimentally determined structures of the same or homologous proteins to assess structural plausibility and identify flexible or uncertain regions.

Glycan structures are obtained from two main sources. When available, pre-equilibrated conformers are downloaded from the GlycoShape 3D structure database[20], which our group has developed and curated, and contains MD-derived ensembles for more than 850 glycans. Alternatively, initial glycan geometries can be generated with the GLYCAM Carbohydrate Builder (www.glycam.org). In this case, it is important to note that the single conformer produced does not necessarily correspond to a representative state of the glycan in solution, and MD simulations in explicit solvent are required to characterise its equilibrium conformational ensemble. Glycan conformers with populations greater than 10 % in solution are deposited in GlycoShape; the development and application of this resource are discussed in detail in **Chapter 7**.

Once individual protein and glycan structures are available, the next step is to link the glycans to the protein at the relevant glycosylation sites. For simple cases, graphical tools such as PyMOL (www.pymol.org) can be used to manually create the covalent bond between the glycan and the side chain (e.g. Asn for N-glycans, Ser/Thr for O-glycans), possibly using a glycan fragment present in the crystal structure as an alignment template. In most of the systems described in this thesis, however, I use Re-Glyco, a tool of the GlycoShape that automates glycoprotein building (https://glycoshape.org/reglyco). Re-Glyco attaches pre-equilibrated glycan conformers from GlycoShape to specified sequons, systematically optimises their orientation, and minimises steric clashes with the protein. The resulting glycoprotein structures are exported in PDB format with atom names consistent with AMBER/GLYCAM nomenclature, thus ready for parametrisation.

2.2.2 Parametrisation and system building

The next step is to assign force-field parameters to each atom in the glycoprotein and embed the molecule in an explicit solvent box. A lipid bilayer can be included if needed.

For solvated glycoproteins, I used the AMBER software. Before parametrisation, the PDB file can be processed with *pdb4amber* to correct common format issues and ensure compatibility. Parameters are then assigned using the *tleap* program, which loads the selected force-field libraries and builds the complete system. Throughout this thesis, protein atoms are described using the ff14SB protein force fields; carbohydrate atoms are described using GLYCAM06; non-standard residues (e.g. modified amino acids, small-molecule ligands, or unusual glycan fragments not covered by GLYCAM) are parametrised with the General AMBER Force Field (GAFF)[20,21]. The final output of *tleap* is a pair of files: a topology file, containing atom types and all bonded and non-bonded parameters, and a coordinate file, containing the initial positions of all atoms. For systems where membrane interactions are mechanistically important, such as Siglec-10 with GT1b and Siglec-6 with GM1 or GM2 (discussed in **Chapters 4 and 5**), the glycoprotein is embedded in a lipid bilayer using CHARMM-GUI Membrane Builder. The bilayer composition mimics a simple plasma membrane, typically a symmetric bilayer of approximately 130 Å × 130 Å, composed of 60% DSPC and 40% cholesterol. CHARMM-GUI generates coordinate and topology files, as well as an initial equilibration protocol, which is converted into AMBER input as needed.

After parametrising the system is solvated with explicit TIP3P water molecules and neutralised with counterions. Additional salt is added as needed to achieve physiological ionic strength (~0.15 M NaCl)

2.3 Equilibration, production and sampling protocols

Running MD simulations requires careful system setup, which can be automated by generating the necessary input files and submission scripts for each simulation stage. Input files specify control parameters such as the time step, simulation length, ensemble type (NVT, NPT), restraints or constraints, temperature and pressure coupling, and output frequency. Template files and tutorials are available on the AMBER website (https://ambermd.org/), along with a comprehensive manual for parameter optimisation.

Submission files are bash scripts that manage job execution on HPC clusters by requesting computing resources, including CPU or GPU allocation and wall time, and listing the commands for each simulation step. These scripts also handle input execution, output generation, and error logging for monitoring and troubleshooting. Computing resources differ across HPC clusters and should always be selected based on thorough benchmarking.

In a canonical ensemble (NVT), the number of particles (N), volume (V), and temperature (T) are kept constant. NVT simulations require temperature regulation at equilibrium, typically achieved by coupling the system to a thermostat that rescales or perturbs velocities, such as the Berendsen[22], Nosé–Hoover[23], or Langevin thermostats[24]. The NVT ensemble is mainly used when temperature must be controlled while keeping the simulation volume fixed. The isothermal–isobaric (NPT) ensemble extends the NVT ensemble by introducing a barostat to regulate the pressure (P), allowing the volume to fluctuate. Common barostats include Berendsen, Parrinello–Rahman[25], and Monte Carlo[26] schemes. The NPT ensemble is widely used for solvated systems, as it mimics experimental conditions by maintaining both temperature and pressure close to laboratory or physiological values.

While specific details are tailored to each system, all simulations in this thesis follow a general protocol. The process begins with energy minimisation to remove steric clashes and high-energy contacts introduced during system building. Steepest-descent minimisation, typically up to 500,000 steps, is performed, often with positional restraints on heavy atoms in the initial stages. After minimisation, the system is equilibrated to room temperature (298–300 K) and atmospheric pressure (1 atm or 1 bar). In AMBER, equilibration is carried out in two main stages: first, NVT equilibration at constant temperature, followed by NPT equilibration at constant pressure to match laboratory or physiological conditions.

During NVT heating, the system is gradually warmed from 0 K to the target temperature (usually 300 K) using Langevin dynamics. Heating is performed in two stages of several hundred picoseconds each, with restraints on solute heavy atoms maintained to prevent large structural distortions. The ensemble is then switched to NPT, and the system is equilibrated until a stable density and box size are achieved. For membrane systems, semi-isotropic pressure coupling is used so that the lateral and normal dimensions of the simulation box can relax independently. During NPT equilibration, positional restraints on proteins, glycans and, for membrane systems, lipids are gradually reduced in several steps.

For example, restraints of 10 kcal mol^{-1} $Å^{-2}$ on heavy atoms may be decreased to 5 and then 2.5 kcal mol^{-1} $Å^{-2}$ before being fully removed for production runs.

Unrestrained production simulations are then conducted in the NPT ensemble using a 2 fs time step and SHAKE constraints on all bonds involving hydrogen atoms. Depending on the system, trajectories range from hundreds of nanoseconds to several microseconds. To ensure thorough sampling of conformational space, especially for flexible glycans, multiple independent replicas are run. Each replica starts from the same equilibrated structure but uses different initial velocity seeds to explore distinct regions of phase space. In some cases, simulations are restarted from selected snapshots (for example, at 900 ns, as described in Chapter 3) with reassigned velocities to extend sampling of specific conformational states or to help overcome barriers related to weak interactions such as CH–π stacking.

In all simulations presented in this thesis, the number of particles remains constant, as required for standard statistical ensembles in biomolecular MD (NVT and NPT). Introducing hard walls to confine the system would create artificial interfaces and unwanted interactions between the solute and the wall. Instead, periodic boundary conditions (PBC) are used to address the finite size of the system[27]. With PBC, the simulation box is replicated in all directions to fill space, allowing each particle to interact with atoms in both the primary box and its periodic images in neighbouring boxes. When a particle exits one face of the unit cell, it re-enters from the opposite face at the corresponding position. This approach keeps the number of particles constant, eliminates surface effects, and enables the system to behave as if it were in an effectively infinite bulk environment (**Figure 2.3**).

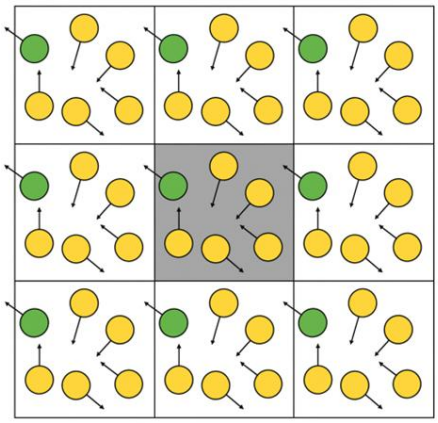


Figure 2.3. 2D representation of Periodic Boundary Conditions. In the diagram, the unit cell (grey) is replicated in all directions. If a particle leaves the unit cell (green), it will be replaced by its

periodic image on the opposite side. This ensures a constant number of particles within the unit cell.

2.4 System-specific sampling schemes

The general protocol described above was applied, with minor adaptations, to all studies presented in this thesis. A key aspect common to all projects is the use of extended and replicated sampling to adequately explore the conformational energy landscape of highly flexible glycoproteins and glycans[28].

In classical MD, the potential energy function V(*r*) can be visualized as a landscape of hills and valleys, with each point representing a possible atomic arrangement. Valleys, or *local minima*, are stable conformations where small atomic changes increase the energy. The aim of the simulations is therefore not to find a single "best" structure, but to sample as many relevant low-energy minima as possible and build an ensemble of conformations. This ensemble identifies the most favorable states for lowering system energy, such as through ligand–receptor binding, multiple interaction networks, or protein adaptation to the environment.

In **Chapter 3**, to examine the role of sialylation on the LRP1 receptor's recognition of the RAP ligand, I used an extended multi-replica sampling scheme for both sialylated and non-sialylated core-1 systems. Two independent 1 μs replicas were run for each glycoform, starting from the same equilibrated structure with different initial velocity seeds. Additionally, a 1 μs restart trajectory was initiated from a 900 ns snapshot of a non-sialylated replica with reassigned velocities. This approach addressed limitations of force fields in modeling weak interactions, such as CH–π stacking between the core-1 Gal residue and Trp63.

In **Chapter 4**, to characterise the conformational effects of hyper-sialylation and desialylation on the CD52 glycopeptide, five independent 1 μs simulations were run for each glycoform, starting from mutually decorrelated snapshots. The cumulative sampling time for CD52 systems is therefore 10 μs. This extended sampling allowed us to demonstrate how sialylation shifts the conformational ensemble towards more extended states, as quantified by the radius of gyration (Rg) and other structural descriptors.

In the Siglec systems described in **Chapters 5** and **6**, receptors were simulated in complex with membrane-embedded gangliosides using CHARMM-GUI-generated bilayers. After the standard six-step Membrane Builder equilibration protocol, long production simulations were performed. For each complex, a weak distance restraint between the

Neu5Ac carboxylate and the conserved arginine in the Siglec binding site was applied during the initial simulation stage to stabilize a biologically realistic binding mode. In the Siglec-6/GM1 system, after equilibration, it was shown that Siglec-6 binding does not depend solely on this salt bridge; a single long 2.5 μs trajectory was performed to capture the detailed dynamics of the stabilized complex.

Finally, free O-glycans and ganglioside headgroups were simulated in explicit solvent as isolated molecules. Multiple 500 ns replicas were run for each glycan, and the resulting trajectories were analysed using torsional clustering (ϕ/ψ angles and relevant exocyclic dihedrals) to identify the most populated conformers. Conformations with populations above 10 % were deposited into the GlycoShape database and used as starting structures or reference states in the protein- and membrane-bound simulations described in the later chapters of this thesis.

References

1. Fadda E. Molecular simulations of complex carbohydrates and glycoconjugates. Curr Opin Chem Biol. 2022 Aug;69:102175.
2. Woods RJ. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. Chem Rev. 2018 Sep 12;118(17):8005–24.
3. Angulo J, Zimmer J, Imberty A, Prestegard JH. Structural Biology of Glycan Recognition. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
4. Fontana C, Widmalm G. Primary Structure of Glycans by NMR Spectroscopy. Chemical Reviews [Internet]. 2023 Jan 9 [cited 2025 Nov 14]; Available from: https://pubs.acs.org/doi/full/10.1021/acs.chemrev.2c00580
5. Understanding glycobiology through multiscale molecular dynamics simulations: From basic principles to case studies. In: Translational Glycobiology in Human Health and Disease. Academic Press; 2024. p. 379–96.
6. Kaminski WA. Issac Newton's "Philosophiae Naturalis Principia Mathematica." World Scientific; 1988. 236 p.
7. Lin FY, MacKerell AD Jr. Force Fields for Small Molecules. Methods Mol Biol. 2019;2022:21–54.
8. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. 2002 May 1 [cited 2025 Nov 14]; Available from: https://pubs.acs.org/doi/abs/10.1021/ja00124a002
9. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. 1998 Apr 14 [cited 2025 Nov 14]; Available from: https://pubs.acs.org/doi/abs/10.1021/jp973084f
10. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force

field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem. 2004 Oct;25(13):1656–76.
11. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008 Mar;29(4):622–55.
12. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, et al. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. J Chem Theory Comput. 2011 Oct 11;7(10):3162–80.
13. Nester K, Gaweda K, Plazinski W. A GROMOS Force Field for Furanose-Based Carbohydrates. J Chem Theory Comput. 2019 Feb 12;15(2):1168–86.
14. Aytenfisu AH, Yang M, MacKerell AD Jr. CHARMM Drude Polarizable Force Field for Glycosidic Linkages Involving Pyranoses and Furanoses. J Chem Theory Comput. 2018 Jun 12;14(6):3132–43.
15. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977 Mar 1;23(3):327–41.
16. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J Chem Theory Comput. 2008 Jan;4(1):116–22.
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.
18. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
19. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan 7;50(D1):D439–44.
20. Ives CM, Singh O, D'Andrea S, Fogarty CA, Harbison AM, Satheesan A, et al. Restoring protein glycosylation with GlycoShape. Nat Methods. 2024 Nov;21(11):2117–27.
21. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004 Jul 15;25(9):1157–74.
22. Berendsen HJC, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984 Oct 15;81(8):3684–90.
23. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys. 1984 Jul 1;81(1):511–9.
24. Adelman SA. Generalized Langevin theory for many-body problems in chemical dynamics: General formulation and the equivalent harmonic chain representation. J Chem Phys. 1979 Dec 1;71(11):4471–86.
25. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981 Dec 1;52(12):7182–90.
26. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. Chemical Physics Letters. 2004 Jan 26;384(4-6):288–94.
27. De Leeuw SW, Perram JW, Smith ER. Simulation of Electrostatic Systems in Periodic Boundary Conditions: A reprise. III. 1982. 26 p.
28. Mayes HB, Broadbelt LJ, Beckham GT. How sugars pucker: electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis. J Am Chem Soc. 2014 Jan 22;136(3):1008–22.

Chapter 3. Structural modulation of the LRP1-RAP: The role of O-glycan sialylation revealed by molecular dynamics simulations

The work in this chapter is part of the following submitted manuscript:

Hintze J, Topaktas AB, Daugbjerg TM. , Shifa Jebari, **D'Andrea S**, Noortje de Haan, Lasse H. Hansen, Rob Baars, Noé Quittot, Cesar Martin, Zhang Yang, Sergey Y. Vakhrushev, Rebecca L. Miller, Dudley K. Strickland, Bradley Hyman, Elisa Fadda, Katrine T. Schjoldager "*The Ligand Preference of LRP1 is Regulated by O-glycans*". *bioRxiv* (2025) https://doi.org/10.1101/2025.10.23.684103

3.1 Introduction

Endocytosis is a fundamental cellular mechanism that allows cells to internalise macromolecules from the extracellular environment[1]. The low-density lipoprotein receptor (LDLR) family and LDLR-related proteins (LRPs) are central to this process[2]. LRPs are also crucial to cholesterol clearance, protein reabsorption, and for the transport of neuronal and cerebrospinal fluid proteins[1]. Mutations in LRP genes are linked to conditions such as dyslipidemia, Alzheimer's disease, migraine, and kidney failure, highlighting their physiological importance[3–6].

Within the LRP family, the low-density lipoprotein receptor-related protein 1 (LRP1) stands out as a large multiligand receptor (~600 kDa)[7] expressed in different tissues, including liver, lung, and brain[8]. LRP1 is also referred to as the Alpha-2-Macroglobulin (A2M) receptor or Apolipoprotein E (ApoE) receptor.[8] Its α-subunit, exceeding 500 kDa[9], containing four ligand-binding cysteine-rich (CR) clusters (I-IV) with a total of 31 CR repeats, enabling interaction with more than 50 structurally and functionally diverse ligands[8] (**Figure 3.1)**. LRP1 is critical in physiological processes relevant to neurodegenerative disorders, particularly in the clearance of amyloid beta (Aβ) and the trafficking of tau protein[10,11] . A key factor influencing the function of these receptors is glycosylation, particularly the presence of O-glycans found within the short linker regions connecting their ligand-binding CR domains[12]. Among the approximately 20 GALNT isoenzymes, only GALNT11 transfers the initiating GalNAc to the linker $C6X_{3\text{-}5}TC_1$ sequence, and GALNT11 knockout abolishes these linker O-glycans[13–16], with a significant impact on the binding of LDLR to LDL and LRP2[17] to albumin, with implications to renal function[16].

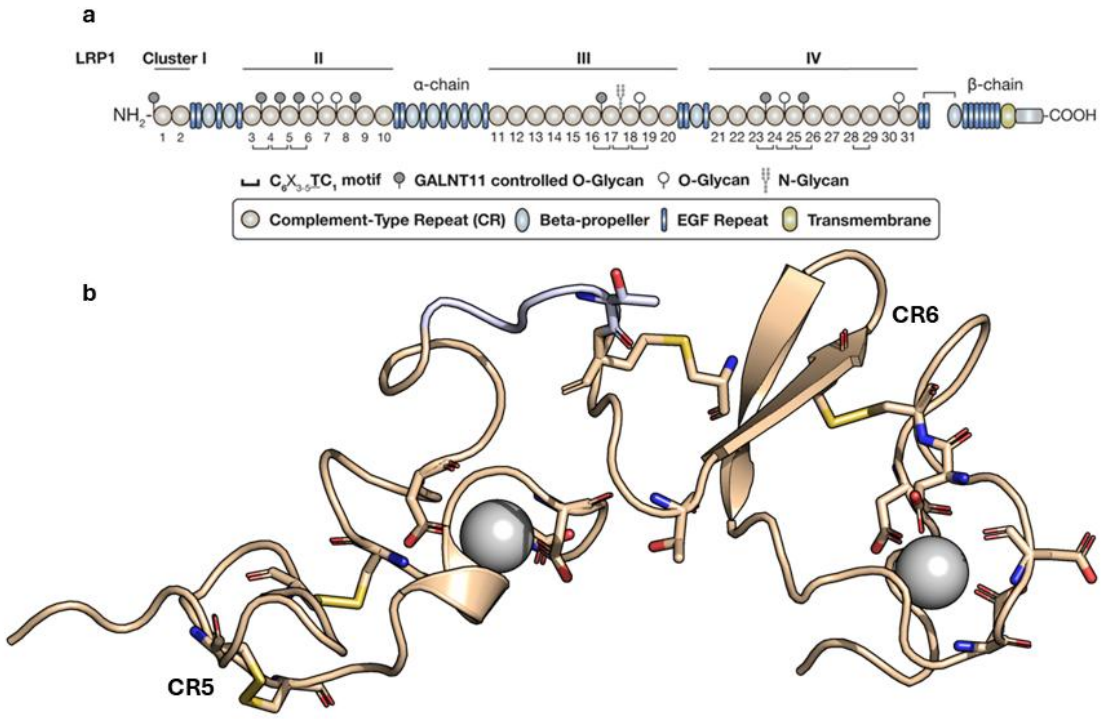


Figure 3.1. a) Schematic representation of full-coding LRP1, illustrating the ligand-binding cysteine-rich (CR) domains (numbered 1-31) within clusters I-IV. All adjacent CRs are separated by short linkers and those linker sequences that adhere to the $\mathbf{C}_6X_{3\text{-}5}\underline{\mathbf{T}}\mathbf{C}_1$ motif have been indicated by horizontal brackets. Linker O-glycans identified by O-glycoproteomic analyses of various mammalian tissues and cell lines of wild-type and *GALNT11*$^{-/-}$ genotypes are marked by a filled (GALNT11-specific) or open (unkown specificity) circle. (Adapted from Hintze *et al*, *submitted*) **b)** NMR structure of CR5 CR6 are shown in tan cartoon, the $C_6X_{3\text{-}5}\underline{T}C_1$ motif is highlighted in ice-blue, with Thr44 displayed in sticks. Ca^{2+} ions are shown as grey spheres, and the acidic Ca^{2+}coordinating residues (Asp/Glu) are depicted in sticks. Disulfide bonds are visible along the CR modules in yellow.

This study aims to determine how GALNT11-initiated O-glycosylation affects the function and specificity of the multiligand receptor LRP1. Using isogenic cell models with and without GALNT11 enzyme, quantitative endocytosis assays, and glycoprofiling, our collaborators, showed that the lack of GALNT11-dependent O-glycans on CR linkers has ligand specific effects on LRP1 mediated endocytosis, specifically they observed a decrease in tau uptake, and an increase in Aβ uptake. The uptake of the other ligands tested, namely A2M, ApoE, and RAP remained unaffected to O-glycosylation. This observation provides the rationale for using RAP as a controlled model system to isolate the role of sialylation at the LRP1 interface, as its binding is not dependent on the presence of O-glycans. In this way, the effect of

sialylation can be specifically evaluated without shifting the focus to glycan-dependent binding mechanisms.

This different behaviour suggested an investigation on the role of the O-glycan sequence and structure and in particular on the dependence on the degree of sialylation of the O-glycan on the CR linkers. It was discovered that the O-glycans introduced by GALNT11 on LRP1 are selectively hyposialylated (undersialylated)[18], which is an atypical characteristic compared to most secreted glycoproteins[12]. In this study, we used RAP (Receptor-Associated Protein) as a competitive receptor antagonist in the uptake assays to suppress LRP mediated endocytosis, and structurally, as ligand in complex with the LRP1 CR5-CR6 domains for MD simulations RAP is an ER chaperone that binds the CR domains with high affinity. RAP acts *in vitro* as a competitive antagonist, occupying the binding sites and selectively blocking LRP1-dependent endocytosis[19]. Because the uptake of RAP remains unchanged, in the isogenic cell models without GALNT11, the primary receptor sites function independently of the O-glycan on the LRP1 CR linker. Meanwhile, experimental evidence suggests that variations observed for tau and Aβ uptake are determined by the glycosylation. Co-expression of RAP allowed an increase in secretion of the soluble ectodomain (sLRP1) needed for glycoprofiling; this revealed a predominance of non-sialylated core-1 O-glycans on the CR linkers, consistently with previous studies reporting hyposialylation of the CR linkers of LRP1 and LRP2 compared with typical secreted glycoproteins[16]. To elucidate the structural and functional implications of this hyposialylation and its potential influence on LRP1 interactions, particularly at its interface with RAP, we performed MD simulations. More specifically, we explored how the presence of O-glycans, both sialylated and non-sialylated, would affect the conformation of the CR modules and their interactions with RAPd1. Results from the MD simulations suggest that a non-sialylated core1 O-glycan could form stable interactions with the Trp63 residue of CR6 and with RAPd1. Conversely, the presence of a sialic acid residue sterically hindered such interactions. Based on these findings we propose a mechanism by which the nature of the O-glycan actively modulates the LRP1:RAP interface.

In this Chapter I present and discuss the analysis of the MD simulations and the mechanism by which the sialylation state of the LRP1 O-glycan impacts inter- and intramolecular interactions at the LRP1:RAP interface, providing a structural understanding of the crucial role these glycans in modulating the receptor's binding properties and overall function.

3.2 Computational Methods

We used the 3D structure of non-glycosylated CR5-CR6 in complex with RAP domain 1 (RAPd1) obtained from NMR-guided molecular docking (PDB 2fyl) as the starting point for all MD simulations. Core-1 (GlyTouCan ID G00032MO) and sialylated core-1 O-glycans (GlyTouCan ID G72008IY) were added at Thr44 using GlycoShape and Re-Glyco[20]. Three independent MD simulations were performed for the core-1-glycosylated complex and two for the sialyl-core-1 glycosylated complex. Data was collected from deterministic sampling over 1 μs trajectories, totaling 5 μs of cumulative sampling. All simulations were performed under identical conditions using the protocol below.

We first prepared the system by running 500k steps of steepest descent energy minimization with all heavy atoms on the protein and glycan restrained. This was followed by a two-step heating phase in the NVT ensemble, from 0 K to 300 K over 1 ns using Langevin dynamics. A 500 ps equilibration in the NPT ensemble followed, for the system to reach an equilibrium pressure of 1 atm with a Berendsen barostat. All restraints on the protein heavy atoms were then removed, and production MD was run at 300 K/1 atm. For the core-1 system, we performed two independent 1μs replicas with distinct velocity seeds, plus a third 1μs restart trajectory initiated from the 900ns snapshot of replica 1 to extend sampling along that conformational state. For the sialyl-core-1 system, we performed two independent 1μs replicas with different initial velocity seeds. The simulations were run with AMBER 2018 on GPUs. The SHAKE algorithm restrained hydrogen bonds, enabling a 2 fs time step. Long-range electrostatics were treated with the Particle Mesh Ewald method, and an 11 Å cutoff was applied for electrostatic and van der Waals interactions. Topology files were generated with *tleap*[21], using GLYCAM_06j-1 to represent carbohydrates, ff14SB[22] to represent protein atoms and counterions, and TIP3P as a water model. Backbone root means square deviation (RMSD) values and distances between Gal (C1 atom) and W63 (CZ2 atom) were measured using Visual Molecular Dynamics (VMD)[23].

3.3 Results

In the NMR-resolved structure of the non-glycosylated CR5-CR6 unit, which includes the $C_6X_{3\text{-}5}TC_1$ motif, in complex with RAPd1 (PDBid 2fyl), each CR domain has the canonical complement-type repeat fold[24], defined by three disulfide bonds and Asp/Glu carboxylates that coordinate Ca^{2+}, as shown in **Figure 3.1b.** The non-glycosylated 3D structure shows that the LRP1:RAPd1 complex is stabilized by three salt-bridge networks involving seven interfacial residue pairs between CR5–CR6 and RAPd1: D27-K24, D28-K24, D27-K60, D28-K60, D66-K93, D68-K93, and D73-R34. The $C_6X_{3\text{-}5}TC_1$ linker, which connects CR5–CR6 and contains the Thr44 O-glycosylation site, is disordered and extends away from the CR5-CR6:RAPd1 binding interface into the solvent, as shown in **Figure 3.2**.

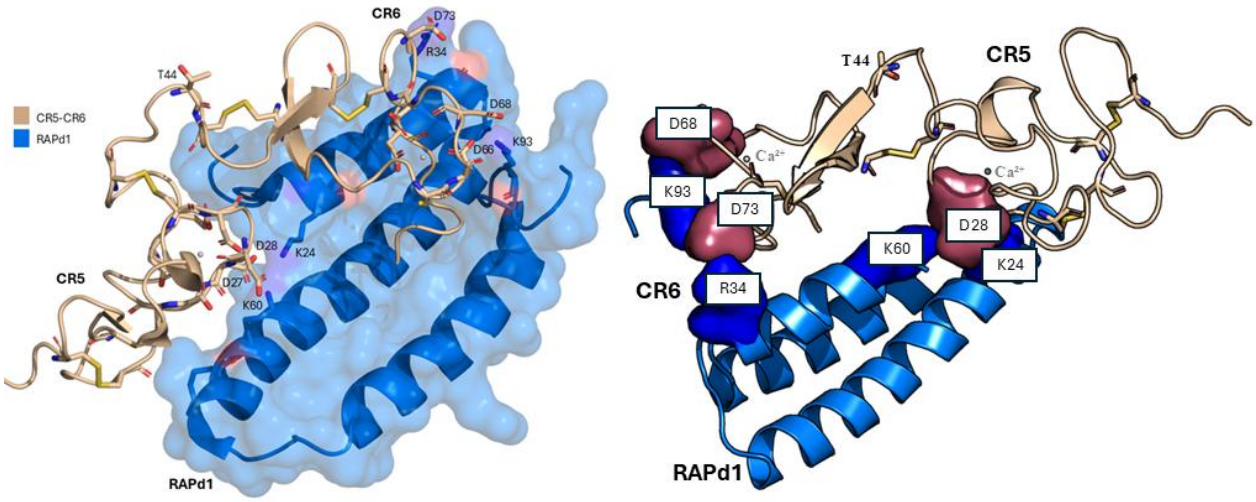


Figure 3.2. *Left:* LRP1 CR5–CR6 is shown with cartoon rendering in tan with disulphide bonds highlighted in yellow; RAPd1 is shown in blue with a semi-transparent surface. The $C_6X_{3\text{-}5}TC_1$ linker that hosts the O-glycosylation site at Thr 44 is labelled. *Right:* Residues that define the Ca^{2+}-dependent acidic site and the seven interfacial salt-bridge pairs are displayed as sticks and labelled as D27-K24, D28-K24, D27-K60, and D28-K60; D66-K93, D68-K93 and D73-R34. Molecular rendering with VMD (https://www.ks.uiuc.edu/Research/vmd/).

The result of glycoproteomics and glycomics analysis shows that the O-glycans on the CR linkers of LRP1 are hyposialylated. To assess the impact of sialylation of the O-glycan at the Thr44 site on ligand interactions, we built systems with two LRP1

glycoforms, one with a non-sialylated core-1 (GlyTouCan ID G00032MO) at Thr44 and the other with a sialyl-core-1 (GlyTouCan ID G72008IY) at Thr44. We ran independent MD simulations on these two systems in replicates using the same protocol.

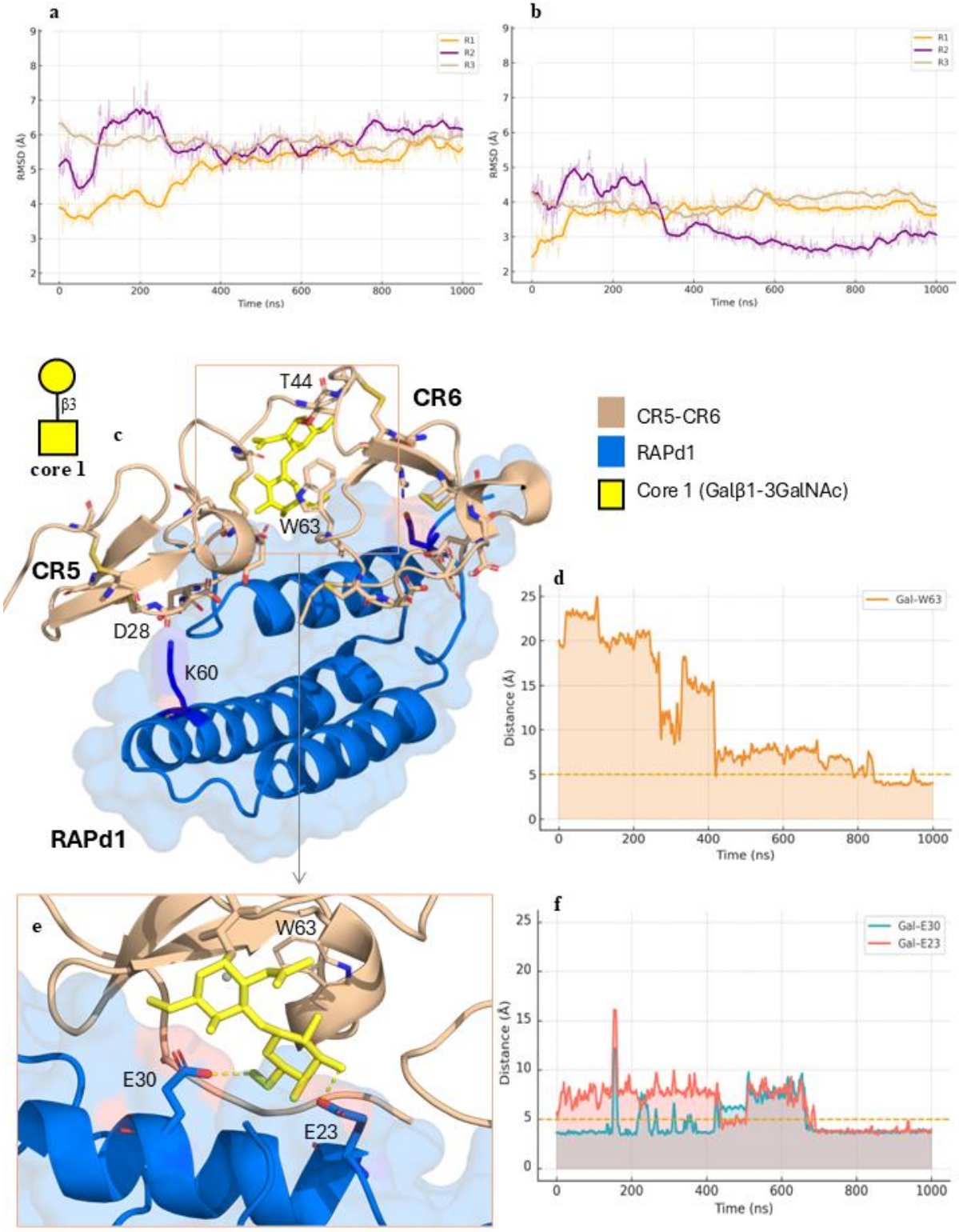


Figure 3.3 a) Backbone RMSD of CR5–CR6 for three independent replicas (two started from the NMR structure with different random velocities; one restarted from a 900 ns snapshot). **b)** Backbone root mean square deviation (RMSD) values obtained for RAPd1 from the same three

replicas. In **panels a–b** the traces reach a plateau, indicating global stability of the complex; RMSD values computed on Cα atoms and plotted. **c)** Overall view of the CR5–CR6:RAPd1 complex (PDB 2fyl) at frame 1000 ns. CR5–CR6 is shown with cartoon rendering in tan, RAPd1 in blue with a semi-transparent surface. The linker carries a core-1 O-glycan (GlyTouCan ID G00032MO) in yellow. *Top-left*: SNFG symbol for core-1 (yellow circle, Gal, yellow square, GalNAc). The boxed region highlights the orientation of Thr44, where the linker folds back toward CR6 positioning the glycan for CH–π stacking against Trp63. **d)** Time graphic of the Gal–Trp63 distance for the NMR-start trajectory (0–1 μs). The trace drops below the 5 Å contact threshold (dashed line) and remains in the stacked interaction. **e)** Close-up from a replica initiated from a snapshot with the Trp63 stacked pose; in this trajectory the glycan shifts away from Trp63 and instead forms hydrogen bonds (dashed lines) with Glu23 and Glu30 on RAPd1. **f)** Time graphic for the velocity-reassigned replica showing Gal–Glu30 (teal) and Gal–Glu23 (coral) distances, same processing as in **panel d**.

Analysis of three independent simulation replicas for the core-1 (Galβ1-3GalNAc) system at Thr44 demonstrates that the backbone RMSD of CR5–CR6 and of RAPd1 of CR5–CR6 and RAPd1 rapidly stabilises, as shown in **Figure 3.3 a, b.** In the presence of a core-1 glycan (GlyTouCan ID G00032MO), the linker adopts conformations in which it folds toward CR6, allowing the terminal Gal to approach the edge of the CR:RAP interface, as illustrated in **Figure 3.3 c.** Monitoring the distance between Gal and Trp63 on CR6 shows a shift from a solvent-exposed position to an intramolecular contact of less than 5Å, as illustrated in **panel d**. This finding supports a transient intramolecular CH–π stacking interaction between the terminal Gal of the core-1/T O-glycan and Trp63. Because CH–π contacts are weak and may be underrepresented in additive force fields, we increased sampling by restarting from this pose and reassigning velocities. In this trajectory, we observed a state shift: the Gal–Trp63 distances increase (**panel d**), while Gal–Glu30/Gal-Glu23 distances decrease (**panel f**), consistent with hydrogen bond formation with Glu23/Glu30 residues on RAPd1, we observed a state shift from intramolecular stacking (Gal–Trp63) to enhanced ligand contacts (Gal–Glu30/Gal-Glu23), as shown in **Figure 3.3 panel e.**

MD simulations with the sialylated core-1 at Thr44 show the opposite behavior. In two independent 1μs replicas, the terminal Neu5Ac prevents the linker from engaging with the interface, preventing Gal–Trp63 stacking and stable RAPd1 contacts, as steric hindrance and the negative charge block the glycan insertion. The analysis of the RMSD values confirms that CR5/CR6 remains stable (**Figure 3.4a**) and that RAPd1 maintains the same conformation in both runs, with contacts limited to CR5/CR6 and no linker involvement, as illustrated in **Figure 3.4.**

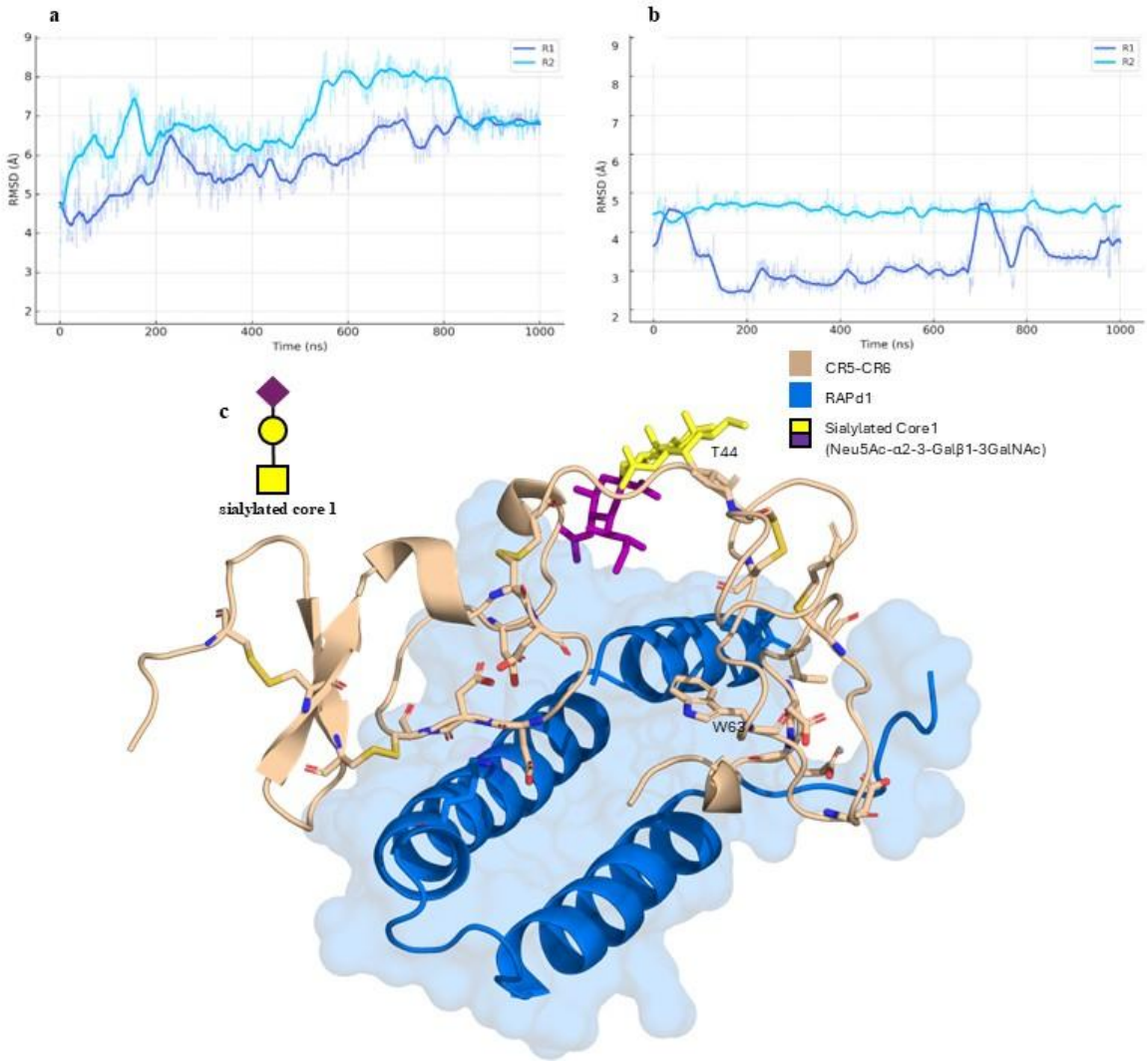


Figure 3.4 a) Backbone RMSD of CR5–CR6 for two independent replicas started from the NMR structure with different random velocities **b)** Backbone RMSD values of RAPd1 for the same two replicas. **c)** CR5–CR6 are shown as a tan cartoon, RAPd1 in blue with a semi-transparent surface. Thr44 with sialylated core-1 O-glycan (GlyTouCan ID G72008IY), depicted with SNFG at top left (purple diamond, Neu5Ac, yellow circle, Gal, yellow square, GalNAc). In this snapshot the glycan remains solvent-facing and does not insert into the interfacial region no CH–π stacking to Trp63 and no stabilising H-bonds to RAPd1, the principal CR–RAP binding site unchanged.

3.4 Discussion

The LRP1 receptor is a key multi-ligand receptor involved in cholesterol clearance and protein trafficking in the cerebrospinal fluid. Its role in clearing beta-amyloid (Aβ) and trafficking tau protein makes it particularly relevant to neurodegenerative disorders[8].

Glycosylation, especially the presence of O-glycans in the linker regions between cysteine-rich (CR) binding domains, is a major factor influencing LRP1 function.

In Hintze et al., our collaborators used isogenic cell models (HEK and SH-SY5Y) with and without GALNT11, along with quantitative endocytosis assays. They found that the absence of GALNT11-dependent O-glycans on CR linkers produced ligand-specific and opposing effects on LRP1 mediated endocytosis. A significant decrease in tau uptake was observed. In cells lacking GALNT11 (HEK-LRP1ΔT11), tau uptake was consistently reduced by 20% compared to isogenic clones with an unaltered GALNT11 O-glycosylation capacity. This reduction was even greater (up to 37%) for hyperphosphorylated tau variants. In contrast, Aβ uptake increased by up to 30% in GALNT11 knockout cells. The uptake of A2M, ApoE, and RAP remained unchanged in GALNT11 knockout cells. The differential behavior observed between the ligands indicated the need to examine the influence of O-glycan sequence and structure, with particular attention on the degree of sialylation on the CR linkers. To enable glycoprofiling, RAP was co-expressed to enhance the secretion of the soluble LRP1 ectodomain (sLRP1), which is required for mass spectrometry analysis

Glycoprofiling showed that non-sialylated core-1 O-glycans are predominant on the CR linkers. Specifically, mass spectrometry revealed that the non-sialylated core-1 is the most abundant on sLRP1, with 30% relative abundance. This unexpected preference against sialic acid capping prompted further structural and functional analysis of hyposialylation at the LRP1:RAP interface. If these sites are poorly sialylated *in vitro*, it suggests that the lack of sialic capping may allow specific accessory contacts that finely modulate interactions between CR domains and ligands. We used MD simulations to clarify the structural and functional effects of this atypical sialylation and to explain the observed differences in tau and Aβ uptake. MD trajectories of 1 μs each revealed a potential regulatory mechanism. The non-sialylated Core1 O-glycan folded to form a stable intra-molecular interaction with Trp63 of CR6 through CH-π stacking between the terminal Gal and Trp63. This conformation allowed the Gal residue to shift from stacking to form hydrogen bonds with RAPd1 residues Glu23 and Glu30, enabling inter-molecular interaction with the ligand. In contrast, sialylation of the Core1 O-glycan introduced steric hindrance, preventing these conformational changes, blocking stacking with Trp63, and inhibiting contact with RAPd1. As a result, the linker remained disengaged. These MD findings indicate that O-linked glycosylation regulates the conformation of ligand-binding CR modules. We propose that the Core1 O-glycan transferred by GALNT11 enables the linker to interact with the LRP1:RAP interface. This interaction may prevent further glycan extension or sialylation while RAP remains bound to the receptor. Our results highlight that low sialylation of linker O-glycans is a key factor in modulating

LRP1 ligand preference. The non-sialylated O-glycan alters CR domain conformation, which explains the observed differences in neurotoxic ligand uptake: tau uptake increases in the presence of the glycan, while Aβ uptake decreases. These findings suggest that O-glycans can differentially affect the progression of neurodegenerative diseases.

3.5 Conclusions

Our MD simulations indicate that O-glycans actively modulate the LRP1:RAP interface. Hyposialylation of LRP1 CR linkers, resulting in a predominance of non-sialylated Core-1 O-glycans, is essential. In this state, Core-1 acts as an accessory modulator, marginally stabilizing the interface through secondary interactions such as CH–π stacking with Trp63 on CR6 or hydrogen bonding with Glu23/Glu30 on RAPd1. These contacts restrict ligand movement, leading to a more stable RAPd1 structure compared to the sialylated system. In contrast, sialylation introduces steric hindrance and negative charge, preventing these interactions and keeping the linker solvent-exposed. The sialylated interface relies on primary contacts, including salt-bridge networks and the Ca^{2+}-dependent core that stabilize the CR5/CR6 unit. This structural insight helps explain ligand-specific effects observed for LRP1, such as decreased tau uptake and increased Aβ uptake. While RAP uptake remains unchanged, allosteric modulation at the interface may account for selective effects on other ligands.

In conclusion, sialylation of the LRP1 O-glycan linker locally modulates micro-contacts at the interface without affecting the overall stability of the complex. This chapter highlights the regulatory role of Neu5Ac in molecular recognition. In the next chapter, we will apply this concept to the CD52/HMGB1 system. We will directly compare sialylated and non-sialylated forms on N- and O-glycans and assess how sialic acid influences complex formation and downstream interactions, such as with Siglec-10.

References

1. May P, Woldt E, Matz RL, Boucher P. The LDL receptor-related protein (LRP) family: an old family of proteins with new physiological functions. Ann Med. 2007;39(3):219–28.
2. Herz J. The LDL receptor gene family: (un)expected signal transducers in the brain. Neuron. 2001 Mar;29(3):571–81.
3. Defesche JC, Gidding SS, Harada-Shiba M, Hegele RA, Santos RD, Wierzbicki AS. Familial hypercholesterolaemia. Nat Rev Dis Primers. 2017 Dec 7;3:17093.
4. Shinohara M, Tachibana M, Kanekiyo T, Bu G. Role of LRP1 in the pathogenesis of Alzheimer's disease: evidence from clinical and preclinical studies. J Lipid Res. 2017 Jul;58(7):1267–81.
5. Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, Terwindt GM, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. Nat Genet. 2012 Jun 10;44(7):777–82.
6. Kantarci S, Al-Gazali L, Hill RS, Donnai D, Black GCM, Bieth E, et al. Mutations in LRP2, which encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. Nat Genet. 2007 Aug;39(8):957–9.
7. Loganathan R, Little CD, Rongish BJ. Extracellular Matrix Dynamics in Biology, Bioengineering, and Pathology. Frontiers Media SA; 2020. 205 p.
8. Herz J, Strickland DK. LRP: a multifunctional scavenger and signaling receptor. J Clin Invest. 2001 Sep;108(6):779–84.
9. Yamamoto K, Scilabra SD, Bonelli S, Jensen A, Scavenius C, Enghild JJ, et al. Novel insights into the multifaceted and tissue-specific roles of the endocytic receptor LRP1. J Biol Chem. 2024 Aug;300(8):107521.
10. Jarosz-Griffiths HH, Noble E, Rushworth JV, Hooper NM. Amyloid-β Receptors: The Good, the Bad, and the Prion Protein. J Biol Chem. 2016 Feb 12;291(7):3174–83.
11. Rauch JN, Luna G, Guzman E, Audouard M, Challis C, Sibih YE, et al. LRP1 is a master regulator of tau uptake and spread. Nature. 2020 Apr;580(7803):381–5.
12. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTBG, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 2013 May 15;32(10):1478–88.
13. Lauc G, Trbojević-Akmačić I. The Role of Glycosylation in Health and Disease. Springer Nature; 2021. 392 p.
14. Wang S, Mao Y, Narimatsu Y, Ye Z, Tian W, Goth CK, et al. Site-specific -glycosylation of members of the low-density lipoprotein receptor superfamily enhances ligand interactions. J Biol Chem. 2018 May 11;293(19):7408–22.
15. Narimatsu Y, Joshi HJ, Schjoldager KT, Hintze J, Halim A, Steentoft C, et al. Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. Mol Cell Proteomics. 2019 Jul;18(7):1396–409.
16. Tian E, Wang S, Zhang L, Zhang Y, Malicdan MC, Mao Y, et al. Galnt11 regulates kidney function by glycosylating the endocytosis receptor megalin to modulate ligand binding. Proc Natl Acad Sci U S A. 2019 Dec 10;116(50):25196–202.
17. Charlton JR, Tan W, Daouk G, Teot L, Rosen S, Bennett KM, et al.

Beyond the tubule: pathological variants of , encoding the megalin receptor, result in glomerular loss and early progressive chronic kidney disease. Am J Physiol Renal Physiol. 2020 Dec 1;319(6):F988–99.
18. Varki NM, Varki A. Diversity in cell surface sialic acid presentations: implications for biology and disease. Lab Invest. 2007 Sep;87(9):851–7.
19. Bu G, Marzolo MP. Role of rap in the biogenesis of lipoprotein receptors. Trends Cardiovasc Med. 2000 May;10(4):148–55.
20. Ives CM, Singh O, D'Andrea S, Fogarty CA, Harbison AM, Satheesan A, et al. Restoring protein glycosylation with GlycoShape. Nat Methods. 2024 Nov;21(11):2117–27.
21. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005 Dec;26(16):1668–88.
22. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.
23. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14(1):33–8, 27–8.
24. Jensen GA, Andersen OM, Bonvin AMJJ, Bjerrum-Bohr I, Etzerodt M, Thøgersen HC, et al. Binding site structure of one LRP-RAP complex: implications for a common ligand-receptor binding motif. J Mol Biol. 2006 Sep 29;362(4):700–16.

Chapter 4. Sialylation regulates CD52–HMGB1 binding and Siglec-10 recognition

The work presented in this chapter is part of the following papers:

DeBono NJ, **D'Andrea S**, Bandala-Sanchez E, Goddard-Borger E, Zenaidee MA, Moh ESX, Fadda E, Harrison LC, Packer NH. *The molecular basis of immunosuppression by soluble CD52 is defined by interactions of N-linked and O-linked glycans with HMGB1 box B. Journal of Biological Chemistry* (2025); 301(4):108350. doi:10.1016/j.jbc.2025.108350

Sobczak K, Antoñana-Vildosola A, Valverde P, Travecedo MA, Jame-Chernaboo Z, Schmidt EN, **D'Andrea S**, Valdaliso-Díez E, Oyenarte I, Laugieri ME, Joe M, Mozaneh F, Lin S-Y, Bosch A, Moure MJ, Franconetti A, Lee SY, Etxaniz-Díaz de Durana J, Pérez-Gutiérrez L, Palazón A, Marcelo F, Fadda E, Corzana F, Gimeno A, Macauley MS, Jiménez-Barbero J, Ereño-Orbea J. *The unique molecular recognition features of Siglec-10: structural insights into sialoglycan and antibody interactions. bioRxiv* (2025) doi: 10.1101/2025.06.10.658867

4.1 Introduction

The immune response is the result of an equilibrium between activating and inhibitory signals that control the activity of immune cells. Advances in oncology have led to the identification and characterization of immune checkpoint receptors, which offer opportunities for new strategies to enhance anti-tumor immunity. More recently, research has increasingly focused on how glycans influence these immune processes[1].

Aberrant glycosylation is a hallmark of cancer[2]. A prominent example is hypersialylation, characterized by increased global expression of sialic acid in cancer cells, which profoundly alters interactions between tumour cells and their environment[3,4]. The human sialic acid-binding immunoglobulin-like lectins (Siglecs) are a family of 14 distinct carbohydrate-binding proteins that play a key role in immune regulation. These receptors recognize sialic acids through a conserved arginine in the N-terminal V-set Ig domain. Members of the family share an overall similar architecture: an N-terminal V-set domain responsible for epitope recognition, followed by a variable number of C2-set Ig domains, and a cytoplasmic tail containing ITIM or ITIM-like sequences that, when phosphorylated, recruit the SHP1 and SHP2 phosphatases to suppress immune-cell signalling[5]. As a result, Siglec-mediated sialic-acid recognition is considered a novel class of immune checkpoint[5]. Siglec-10 is one such inhibitory receptor and is emerging as a therapeutic target, although its molecular mechanisms for ligand recognition and immune

modulation are not yet fully understood. The Siglec-10 structure is shown in **Figure 4.1.a**, highlighting the V-set Ig domain, the C2-set domains and the cytoplasmic ITIM/ITIM-like motifs.

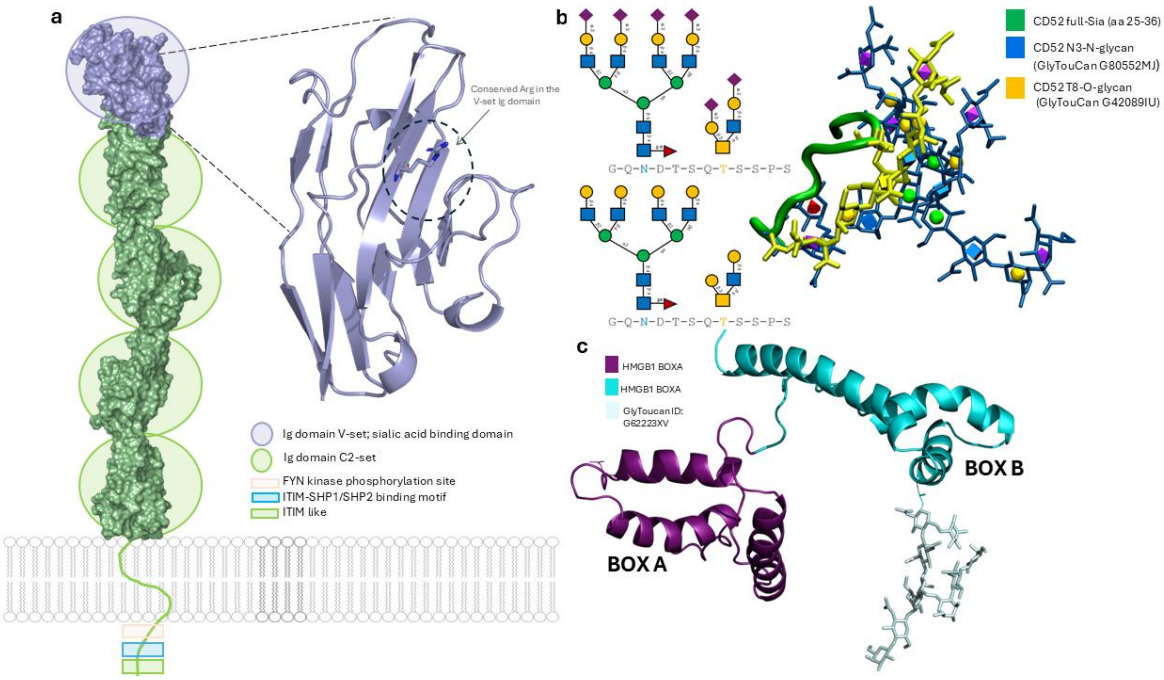


Figure 4.1. a. Surface representation of full-length Siglec-10 with the N-terminal V-set Ig domain highlighted in lightblue and the C2-set Ig domains in green. On the right, a close up view of the V-set Ig domain represented as cartoon, with the conserved arginine side chain indicated. **b.** *Left*, SNFG schematics: the 12-residue peptide CD52 (GQNDTSQTSSPS) is shown with the N-glycosylation site N3 and the O-glycosylation site T8 highlighted in the sequence. Top: hyper-sialylated glycoform with terminal Neu5Ac (purple diamonds) capping a complex, tetra-antennary N-glycan at N3 and a core-2 O-glycan at T8. Bottom: corresponding not sialylated reference scaffolds shown for comparison. Symbols follow SNFG convention. Right, atomistic representation: a representative model of soluble CD52 carrying the same hyper-sialylated glycans: the peptide backbone is rendered in green cartoon; the N3 N-glycan is in blue sticks (GlyTouCan G80552MJ); the T8 O-glycan is shown in yellow sticks (GlyTouCan G42089IU); terminal Neu5Ac residues appear in purple. **c**. Cartoon representations of the HMGB1 Box A (magenta) and Box B (cyan) domains with the N-glycan (GlyToucan ID: G62223XV) represented in light cyan as sticks.

Siglec-10 is expressed on antigen-presenting cells (APCs) and tumour-associated macrophages (TAMs)[6]. Blocking its interactions with monoclonal antibodies (mAbs), such as S10A, can enhance anti-tumor responses and is being investigated in several cancers[7]. Understanding ligand recognition is essential for developing targeted Siglec-10 therapies.

Siglec-10 binds sialylated glycans with α2-3 and α2-6 linkages[8] through its N-terminal V-set domain. Studies suggest that it can also interact with heavily O-glycosylated, GPI-anchored surface proteins such as CD24 and CD52[9,10]. In Siglec-10, the V-set domain hosts the canonical sialic-acid-binding pocket, where the conserved arginine engages the negatively charged carboxylate of Neu5Ac.

Soluble CD52 is a short glycopeptide consisting of only 12 amino acids (GQNDTSQTSSPS)[8,11] (**Figure 4.1.b**). In its membrane bound form, CD52 is presented as a GPI-anchored protein, whereas the soluble form released by activated T cells functions as an immune regulator by suppressing immune responses. Despite its small size, soluble CD52 is heavily glycosylated, carrying a complex, highly sialylated N-glycan at N3 and an O-glycan localised to T8, with both glycans contributing to its immunosuppressive activity[10,12]. Activated T cells initiate this pathway by releasing soluble CD52, which then sequesters the pro-inflammatory protein High Mobility Group Box Protein 1 (HMGB1)[12–14].

HMGB1 is composed of two structurally similar HMG-box domains, Box A and Box B, see **Figure 4.1.c,** CD52 interacts specifically with Box B², which is pro-inflammatory when uncomplexed. Structural studies have shown that Box A and Box B adopt closely related three-helix bundle folds, with a backbone RMSD of ~2.3 Å over 54 Cα atoms and ~25% sequence identity for human HMGB1 (UniProt P09429)[15]. This high structural similarity, together with the availability of the HMGB1-BoxA/p53 transactivation domain complex, provided a template for building our working model of the CD52/HMGB1-BoxB complex. In particular, we exploited the overlap between the p53 transactivation domain and the CD52 backbone around the O-glycosylation site at T8 to guide the positioning of CD52 on Box B.

Following HMGB1 sequestration, the complex binds to the inhibitory receptor Siglec-10, which is expressed on T cells. This crucial interaction leads to the exertion of broad immunosuppressive properties. The overall mechanism makes CD52 a strong candidate for development as a potential immunotherapeutic agent. However, the structural basis by which CD52 glycosylation controls HMGB1 recognition and Siglec-10 binding, and in particular how O-glycan sialylation contributes to this process, has not been defined.

In this chapter, I will show how the CD52 sialylation drives selectivity and affinity. We show that the doubly sialylated core-2 O-glycan at T8 enables selection of the HMGB1 Box B, and orientating the silaylated N-glycan appropriately for binding Siglec-10. Molecular dynamics (MD) simulations revealed that hyper-sialylation of both N-linked

and O-linked glycans influences the glycopeptide conformation and facilitates a specific interaction with HMGB1 Box B, but not Box A. The doubly sialylated core-2 O-glycan at T8 binds directly and stably to HMGB1 BoxB. This O-glycan-mediated interaction induces a conformational change in the CD52/HMGB1 complex, exposing the sialylated arms of the N-linked glycan at N3 for recognition by Siglec-10.

In parallel, I will provide key structural insight into Siglec-10, whose atomistic recognition mechanism was previously undefined. We identified two critical arginines in the V-set pocket that engage the sialic-acid carboxylate: a canonical R119 (conserved across Siglecs) and a non-canonical R127 (distinctive of Siglec-10). The results of MD simulations indicate cooperative engagement by multiple arginines, which explains the receptor's preference for hyper-sialylated, multivalent epitopes like those found in the CD52/HMGB1 complex.

4.2 Computational Methods

4.2.1 CD52/HMGB1 system preparation

To investigate how HMGB1-BoxB recognizes the CD52 glycopeptide and the role of sialic acid in this process, we conducted a series of all-atom, classical MD simulations. Simulations were first performed on the isolated glycopeptide, then on the CD52/HMGB1 box B complex, which was constructed using data from the initial simulations. The starting structure for the CD52 peptide was based on amino acids 25 to 36 (GQNDTSQTSSPS) from the Uniprot P31358 CAMPATH-1 antigen, obtained from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/). This sequence is recognized by HMGB1 box B and matches the peptide used in our experimental assays. Glycan structures were selected according to our glycome and glycoproteome analyses. To model the fully sialylated CD52 glycopeptide, a core-fucosylated tetra-antennary α2-3 tetra-sialylated N-glycan (GlyTouCan ID G80552MJ) was attached at N3, and an α2-3 di-sialylated core 2 oligosaccharide (GlyTouCan ID G42089IU) at T8. Glycan 3D structures were sourced from the GlycoShape Glycan Database (https://glycoshape.org), and the ReGlyco tool was used to link the N- and O-glycans[16]. Two sets of MD simulations were performed: one with fully sialylated N- and O-glycans, and one without terminal sialic acids. For each set, five independent 1 μs simulations were run, totaling 10 μs of sampling for both sialylated and non-sialylated forms. The first simulation in each set used the initial structures described above; subsequent simulations started from

uncorrelated snapshots from these trajectories. The system was first energy-minimized for 500,000 steps using the steepest descent method. It was then heated to room temperature in two 500 ps steps within the NVT ensemble, using a 2 fs integration time step. The temperature was increased from 0 K to 100 K, then from 100 K to 300 K, using Langevin dynamics with a friction coefficient of 1.0 ps^{-1}. This was followed by a 500 ps NPT equilibration at 1 atm. Pressure was controlled with a Berendsen barostat in AMBER 2018 and the Parrinello-Rahman method in GROMACS 2021[17]. Long-range electrostatics were handled with Particle Mesh Ewald and an 11 Å cutoff; dispersion interactions were also truncated at 11 Å. The GLYCAM_06j-1 force field[18] was used for carbohydrates, TIP3P for solvent[18,19], and ff14SB for peptides and counterions[20]. Simulations were performed with AMBER 2018[21] on GPU[22] and GROMACS[23] 2021.4 on Oracle Cloud Infrastructure. The complex between fully sialylated CD52 glycopeptide and HMGB1 was built by aligning the dominant CD52 backbone conformation, determined by Ramachandran analysis, to the 12-residue p53 transactivation domain (PDB-ID 2LY4) using PyMOL (www.pymol.org). After alignment, HMGB1 box A was replaced with box B from PDB 2GZK, after removing the bound DNA and box A domain. The MD simulation of this complex followed the same protocol, with a 1 μs production run.

4.2.2 MD Simulations of the Siglec-10 in complex with GT1b.

The 3D structure of the full-length Siglec-10 ECD was generated with AlphaFold-2 (AF2)(AlphaFold Protein Structure Database number AF-Q96LC7). GT1b ganglioside (GlyTouCan-ID G08648UJ) was used. The equilibrium structure of the GT1b was sourced from the GlycoShape server[16]. The starting conformation of the complex was built to complement the 3D structure of GT1b with an orientation selected to maximize the number of arginine residues engaging the trisialylated GT1b headgroup. Accordingly, the terminal α2-8-linked Neu5Ac was positioned to interact with the canonical R119, while R127 and R80 aligned to interact with the internal and terminal α2-3-linked Neu5Ac residues, respectively. The complex was embedded in a symmetric 130 Å × 130 Å lipid bilayer composed of 60% 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and 40% cholesterol (CHL1) using the CHARMM-GUI Membrane Builder tool[24]. All MD simulations were conducted with AMBER 22, using the CHARMM36m force field to parameterize the protein, lipids, and glycans. The system was first energy minimized for 5000 steps and the equilibration was performed following the CHARMM-GUI protocol[25] consisting of six steps. Positional restraints of 10 kcal/mol·$Å^2$ were initially

applied to the system and were gradually reduced during the NPT equilibration phase, progressively lowered to 5 kcal/mol·Å² for the protein and 2.5 kcal/mol·Å² for the membrane until they were completely removed. Temperature was maintained at 315.15 K using Langevin dynamics with a friction coefficient (gamma_ln) set to 1.0 ps^{-1} and pressure was controlled semi-isotropically at 1 atm using the Berendsen barostat. Periodic boundary conditions were applied throughout. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME) method with a cutoff of 11 Å. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm, allowing for a 2 fs integration time step. Two independent 500ns production runs were performed, resulting in a total sampling time of 1 μs. Trajectory analysis was conducted using Python3 scripts. Distances between key arginine side chains and the Neu5Ac residues of GT1b were calculated frame by frame using VMD. Occupancy was defined as the percentage of frames in which the atomic distance remained below 5 Å. All plots were generated using the python packages Matplotlib and Seaborn.

4.3 Results

We analysed two conformational sets of the CD52 glycopeptide (12 amino acids, sequence GQNDTSQTSSPS): a hyper-sialylated form and a desialylated form. The hyper-sialylated form included a core-fucosylated tetra-antennary α2-3 tetra-sialylated N-glycan (GlyTouCan ID G80552MJ) at N3 and an α2-3 di-sialylated core 2 oligosaccharide (GlyTouCan ID G42089IU) at T8. The desialylated form of the peptide carried the same glycan structures without terminal sialic acids (N-glycan GlyTouCan ID G56655CC and O-glycan GlyTouCan ID G96017QA). We selected these glycoforms based on the experimental glycoproteomics analysis from our collaborators (DeBono et al., J. Biol. Chem., 2025) where high-resolution mass spectrometry analysis assigned the O-glycosylation site to T8 with high confidence.

The results of the MD trajectories indicate that the CD52 glycopeptide adopts more extended conformations when glycans are sialylated and more compact ones when sialic acid is absent. To quantify this, we calculated the radius of gyration (Rg) for the peptide backbone across all replicas and compared their distributions using kernel density estimates (KDE), see **Figure 4.2**. The analysis was done on a cumulative sampling of 10 μs. In hyper-sialylated systems, the Rg distributions shifted toward larger and narrower values. In non-sialylated systems, the distributions are centered on lower values and are

broader or multimodal. These findings confirm that sialylation shifts the average conformational landscape of CD52 toward more elongated structures.

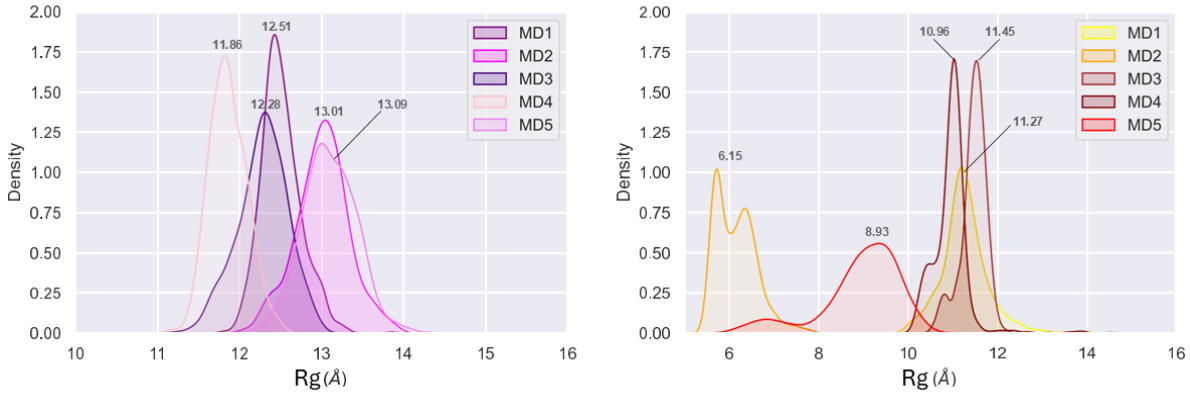


Figure 4.2 Kernel density estimates (KDE) of the backbone radius of gyration (Rg) for the isolated CD52 glycopeptide from five independent 1-μs MD replicas per condition (total 10 μs per panel). Left: hypersialylated (+Sia) ensemble (N3 tetra-sialylated N-glycan; T8 core-2 di-sialylated O-glycan). Right: non sialylated (−Sia) ensemble, which used the same glycan structures without terminal Neu5Ac. Sia increases the distribution peak to 12–13 Å, indicating more elongated conformations. Without Sia, values cluster at 10–11 Å, reflecting more compact states. Sialylation therefore shifts CD52 toward a more extended conformation.

The Ramachandran analysis of the conformation of the CD52 glycopeptide shows significant changes in the conformational propensity at bonds 2 and 8, corresponding to the peptide bonds centered at N3 and S9, respectively, see **Figure 4.3**. In the case of N3 (bond 2), the conformational preference changes from an extended β strand conformation in the presence of a core-fucosylated tetra-antennary and fully sialylated N-glycan (GlyTouCan ID G80552MJ) to a 3 to 10 helical turn when the same N-glycan structure does not have terminal sialic acids (GlyTouCan ID G56655CC).

The absence of sialylation on the extended core 2 O-glycan (GlyTouCan ID G96017QA) at T8 stabilizes the localized propensity for a left-handed α helical conformation, which extends to S9. The increased helicity of the CD52 glycopeptide in the absence of sialylation ultimately leads to a more compact structure, also shown by the analysis of the Rg, see **Figure 4.2**. The Ramachandran plots for all the peptide bonds in the presence and absence of sialylation are shown in **Figure 4.3.**

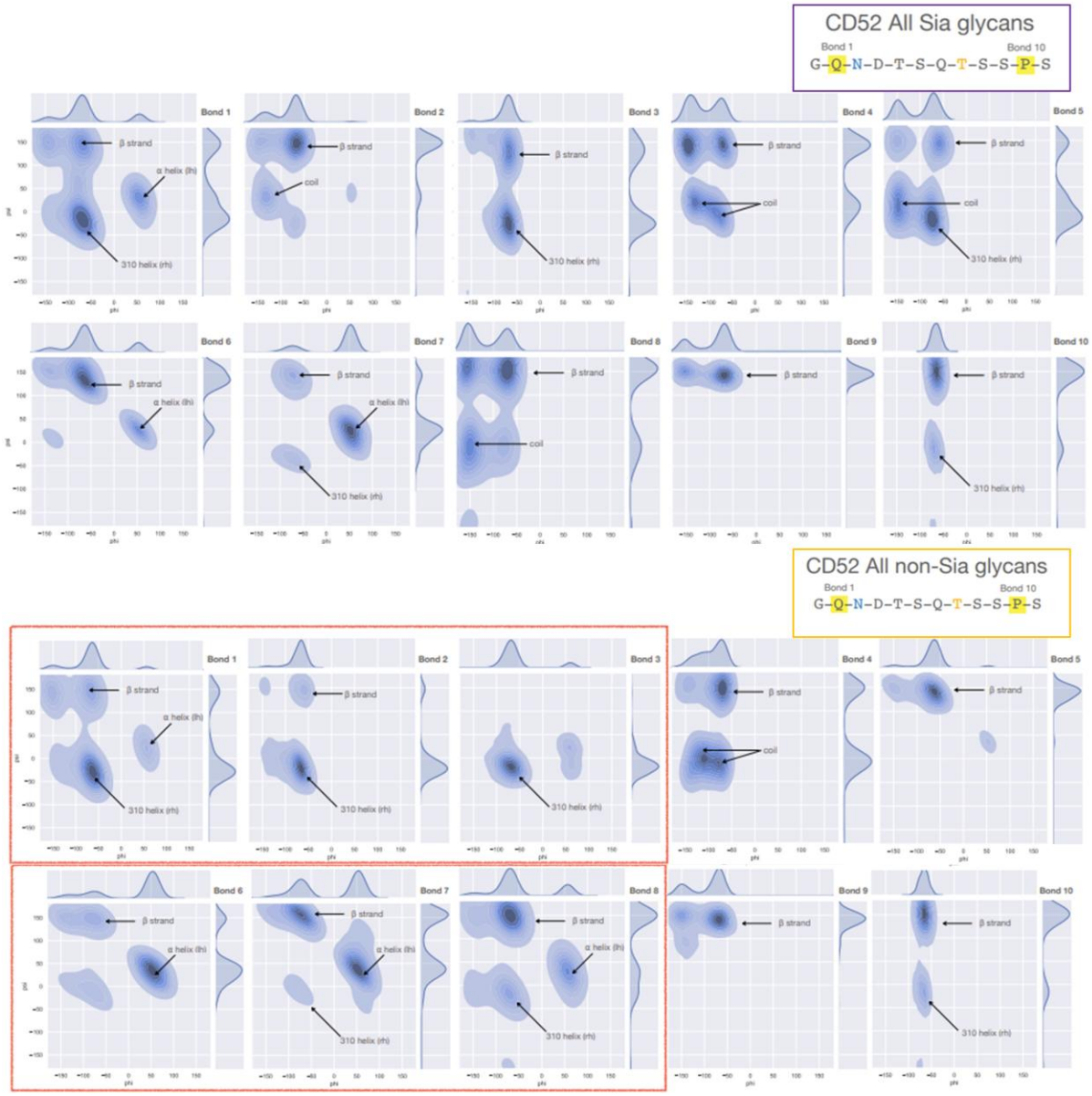


Figure 4.3. Top: Ramachandran plots of all peptide bonds (Bond 1–10) for the +Sia ensemble, obtained by pooling 5 × 1-μs replicas (5 μs total). Secondary-structure basins associated with ϕ/ψ (β-strand, 3_{10} helix, α/α^{L}, coil) are annotated; KDE traces are shown along the top and right margins of each panel. Bottom: Ramachandran plots of all bonds for the −Sia ensemble, with red boxes highlighting Bond 2 (centered on N3) and Bond 8 (centered on S9) selected for direct comparison. These panels show a shift from β/extended states (+Sia) toward 3_{10} helix (Bond 2) and an increased left-handed α/3_{10} propensity around T8/S9 (Bond 8) when sialic acids are absent. All plots were generated using Python with Seaborn.

To our knowledge, to date there is no available structural information on the complex between CD52 and HMGB1-BoxB, nor a structural explanation for the preference of the CD52 glycopeptide for BoxB over the similar BoxA domain. To gain insight into the molecular basis of this selectivity, we analyzed the structure of the complex in silico between the HMGB1 box A and a section of the p53 transactivation domain[26] (aa 36–48).

Although the functions of these two complexes are entirely unrelated, the structural features of the HMGB1 box A/p53 complex provide important clues that ultimately led us to build a working model for our complex. The rationale behind choosing the HMGB1 box A/p53 as a template is based on two main considerations. First, as discussed above, HMGB1 BoxA and BoxB share a high degree of structural similarity in both sequence and fold. Second, the structural alignment of representative snapshots from the MD simulations of the CD52 hyper-sialylated glycopeptide to the box A-bound p53 conformation indicates a keen similarity between the two peptides around the O-glycosylation site on CD52 at T8, see **Figure 4.4**

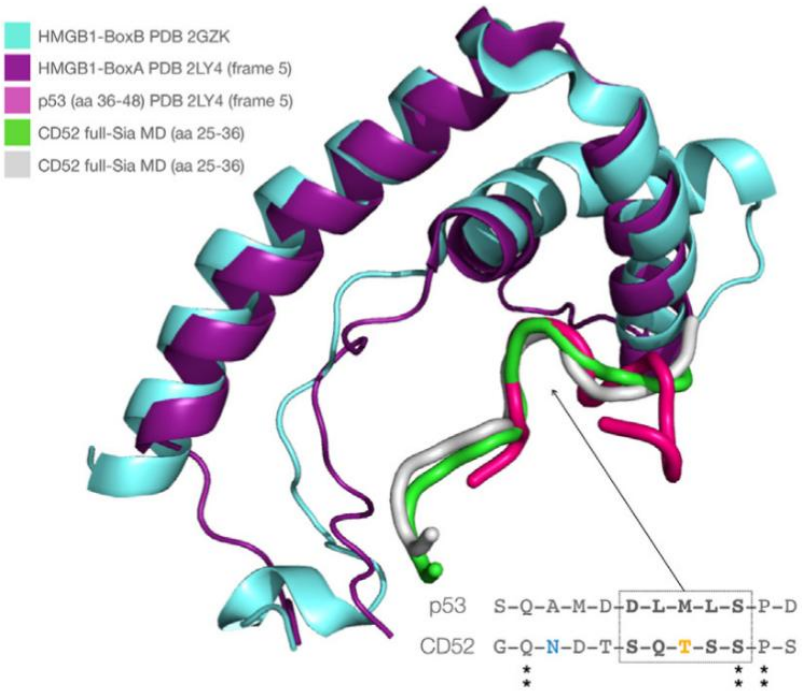


Figure 4.4 Structural alignment of the HMGB1 box A (purple cartoon representation, PDB 2LY4) bound to p53 (pink cartoon representation, PDB 2LY4) to the HMGB1 box B (cyan cartoon representation, PDB 2GZK) with two selected representative snapshots (green and white cartoon representation) from the cumulative 5 μs MD simulation of the hyper-sialylated CD52 glycopeptide. Sequence alignment of the HMGB1 box A-bound p53 peptide to the active CD52 glycopeptide where the sequence corresponds to the helical turn is highlighted in the dotted box. The T8 of CD52 is highlighted in yellow. Graphical rendering and sequence alignment with pymol (https://pymol.org/2/).

Guided by these considerations, we generated a 3D model of the HMGB1-BoxB in complex with the CD52 glycopeptide, in which glycosylation on the peptide and the HMGB1-BoxB at N134 was restored with GlycoShape[16] to complement the tight steric

requirements. The most striking and unique feature of this 3D model is that the main contacts occur between the HMGB1-BoxB and the sialylated core 2 O-glycan, which directly faces the BoxB surface, and not between HMGB1-BoxB and the CD52 peptide, see **Figure 4.5**. Also, in this model, the large tetra-antennary N-glycan at N3 is oriented with the arms directed towards the solvent, and thus accessible for binding, potentially exposing the terminal sialic acid to Siglec 10[27]. We assessed the stability of this complex through a single MD trajectory with a 1 μs production phase with no restraints in place. This simulation allowed us to identify a set of key residues of HMGB1 box B that make specific contacts with the fully sialylated core 2 at T8, namely R97, K146, K152 and R163, see **Figure 4.5**. Interestingly, all of these residues, except K146, are not conserved in HMGB1-BoxA, where R97 is G, K152 is D and R163 is I, providing a rationale for the preference of the hyper-sialylated peptide for BoxB and not BoxA. Ultimately, aside from other considerations and based on this 3D model of the complex alone, lack of sialylation of the core 2 O-glycan at T8 would prevent binding of CD52 to HMGB1-BoxB as most contacts with the protein directly involve the terminal sialic acids, see **Figure 4.5**. The MD simulation also shows that the terminal sialic acids on the CD52 N-glycan are mostly oriented toward the solvent and that the N-glycan on HMGB1-BoxB wraps around the CD52 N3 N-glycan. As a note of caution, additive force fields tend to enhance glycan-glycan interactions[28], so the extent of these contacts may be overestimated.

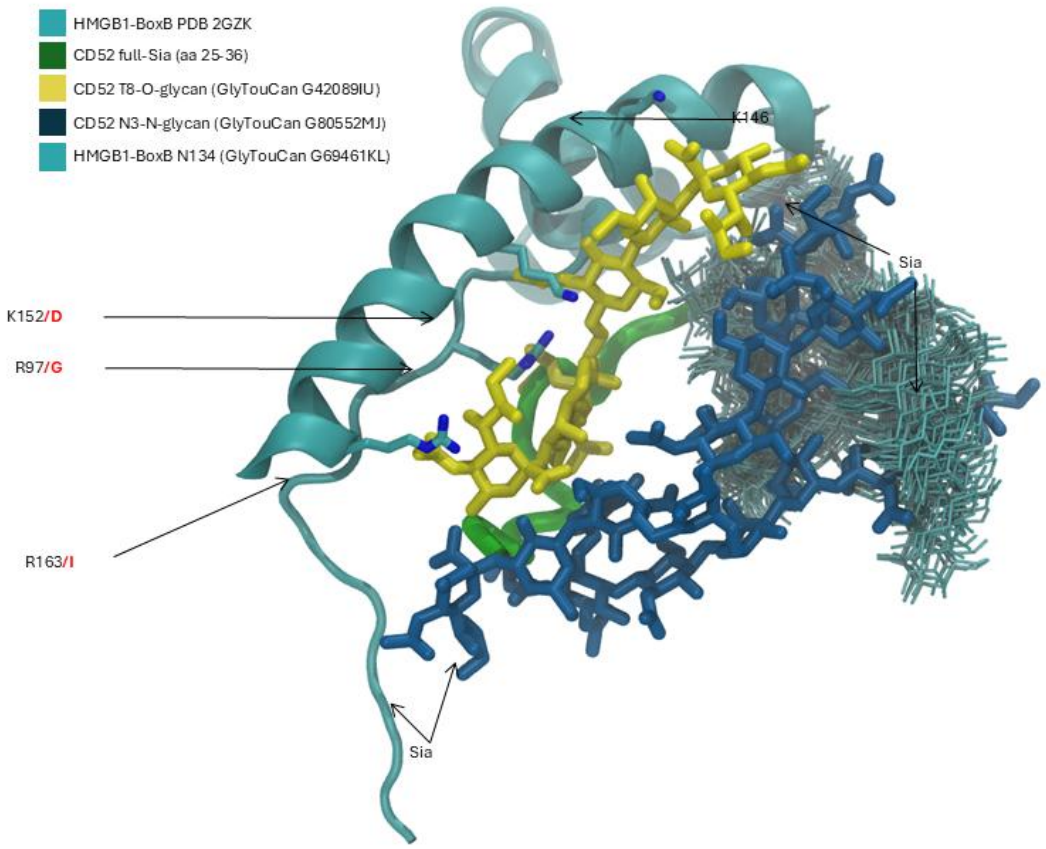


Figure 4.5 Snapshot (frame 200 ns) from the 1 μs MD trajectory of the reconstructed 3D model of HMGB1 box B (cyan) in complex with the hyper-sialylated CD52 glycopeptide (green). The N- and O-glycans are shown with sticks in blue and yellow, respectively. The key protein residues identified to interact with the O-glycan at T8 are indicated with labels and shown with sticks, corresponding residues in Box A are shown in red labels. The positions of the terminal sialic acids of the CD52 N-glycan are also highlighted with labels. The conformation of the HMGB1-BoxB N-glycan is shown with multiple snapshots collected every 20 frames (1 ns each) through the duration of the trajectory. Graphical rendering with VMD (https://www.ks.uiuc.edu/Research/vmd/).

Given the results on the CD52/HMGB1 complex, where hyper-sialylation promotes multivalent exposure of sialic acid to the solvent, the next step is to define how the Siglec-10 receptor may recognize and stabilize these epitopes.

Although the physiological ligand of Siglec-10 in the CD52 pathway is the tetra-antennary N-glycan at N3, we did not use this glycan as a model epitope in our simulations. The sialic acids in a tetra-antennary N-glycan are located at the ends of flexible arms that move independently, making the epitope highly dynamic. This flexibility would have made it difficult to determine the specific roles of arginines in the binding pocket in recognising and stabilising individual sialic acid residues. Instead, we selected the trisialylated ganglioside GT1b, whose three sialic acid residues form a more

compact and defined headgroup, while still providing a hypersialylated, multivalent epitope.

At the time when these simulations were carried out the crystal structure of Siglec-10 was not available, so as a starting structure for the MD we built the 3D model of the complex with the GT1b ganglioside (GlyTouCan G08648UJ) by using the full length AF2 structure of Siglec-10 deposited in the AlphaFold Protein Structure Database (AF-Q96LC7)[29]. The AF2-predicted model shows a high degree of similarity with the experimental structure, particularly across the V-set domain, with a backbone RMSD value of 0.429 Å, calculated for the residues not in dynamic loops (i.e. 648 to 648 atoms). Furthermore, the root mean square fluctuation (RMSF) profile closely mirrored the experimental β-factor distribution, see **Figure 4.6.**

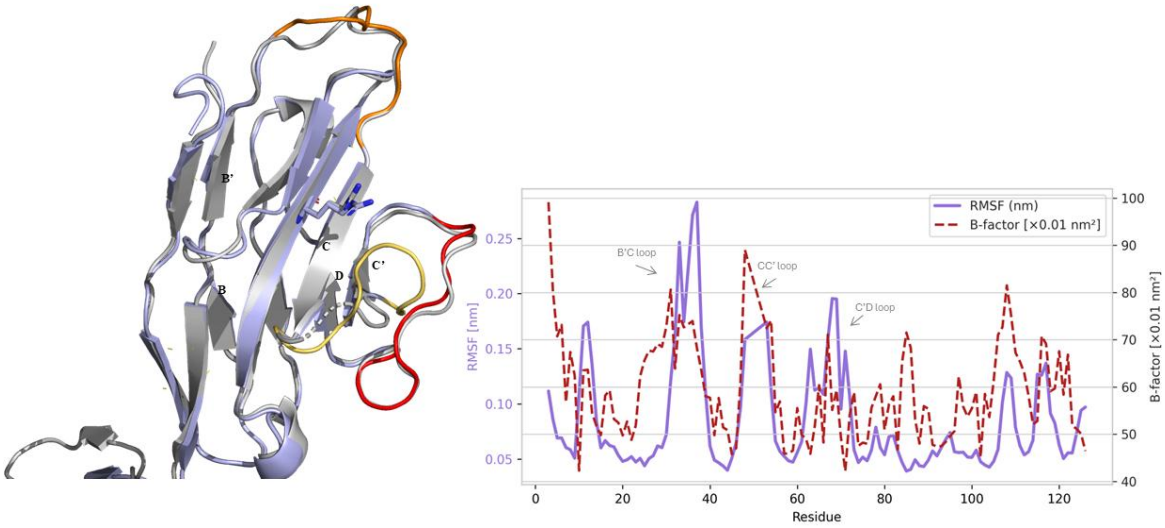


Figure 4.6. Structural alignment and RMSF analysis of the Siglec-10 V-set domain. *Left:* Cartoon representation of the V-set domain of Siglec-10 (AF-Q96LC7) is shown in light blue, aligned with the X-ray structure reported by Sobczak *et al.*[19] shown in grey. The B′C, CC′, and C′D loops are highlighted in orange, yellow and red, respectively. *Right*: RMSF from the MD simulation (purple line) compared with crystallographic β-factors (red dashed line) plotted as a function of residue number. Figure adapted from Sobczak *et al.* (2025) .

GT1b is a trisialylated ganglioside containing two α2-3-linked sialic acid residues, one terminal and one internal, and one α2-8-linked sialic acid that bridges the internal α2-3 sialic acid to galactose. Experimental work in liposomal systems[31] identified GT1b as the epitope with the strongest binding affinity for Siglec-10 among a panel of tested gangliosides.

The equilibrium structure of the GT1b was sourced from the GlycoShape[16] database. The starting conformation of the complex was built to complement the 3D structure of GT1b, with an orientation selected to maximize the number of Arginine residues engaging the

trisialylated GT1b headgroup. Accordingly, the terminal α2-8-linked sialic acid was positioned to interact with the canonical Arg119, while Arg127 and Arg80 aligned to interact with the internal and terminal α2-3-linked sialic acid residues, respectively, see **Figure 4.7.**

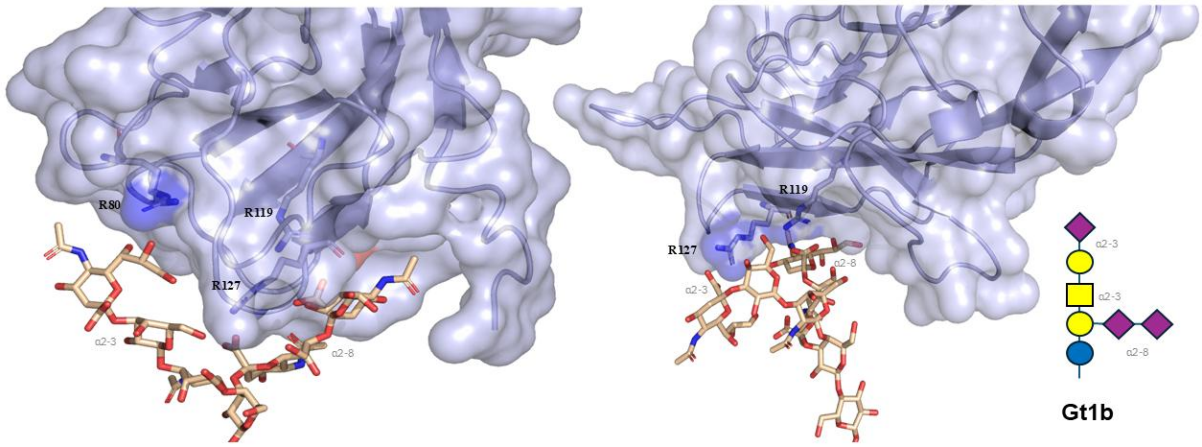


Figure 4.7. GT1b ganglioside in complex with Siglec-10. Left panel: 3D model of the Siglec-10/GT1b complex used for the MD simulations. The protein is rendered with a transparent surface and cartoon in light blue to highlight secondary structure motifs. The GT1b ganglioside is rendered with sticks with C atoms in tan, O in red, and N in blue. The three NeuNAc are labeled according to their linkages (α2-3 terminal, α2-3 internal, and α2-8). Right panel: Representative frame from the MD simulation showing a rearranged binding conformation. The Arg127 shifts its interaction to the terminal α2-3 sialic acid, while Arg119 maintains its interaction with the α2-8 sialic acid.

To understand the molecular determinants for the stability and specificity of the interactions between GT1b and the Arginine residues in the Siglec-10 binding site, both canonical and non-canonical, we ran two independent MD replicas of 500 ns each with the Siglec-10 and GT1b complex embedded in a lipid bilayer. Our results show that in both replicas Arg127 plays a synergistic role, stabilizing the interaction with GT1b. The simulations reveal that GT1b forms a stable interaction with Arg119 through α2-8 sialic acid, while Arg127 interacts with α2-3 sialic acid. More specifically, in the first replica, Arg119 keeps its interaction with the α2-8 sialic acid throughout the entire 500 ns trajectory, showing an occupancy of 95%, while Arg127 exhibited a high occupancy (77%) with the terminal α2-3 Neu5Ac, see **Figure 4.8**. In the second replica, Arg119 switches its interaction from the α2-8 sialic acid to the terminal α2-3 sialic acid, suggesting that the canonical binding Arginine is capable of engaging differently linked terminal Neu5Ac. Meanwhile, Arg127 consistently retained interaction with the terminal α2-3 sialic acid during both simulations. These results support that recognition and

binding of sialylated epitopes by Siglec-10 can involve multiple Arginine residues in addition to the canonical Arg119, and that these contribute to increased specificity and affinity for different sialic acid linkages, particularly in the context of polysialylated epitopes.

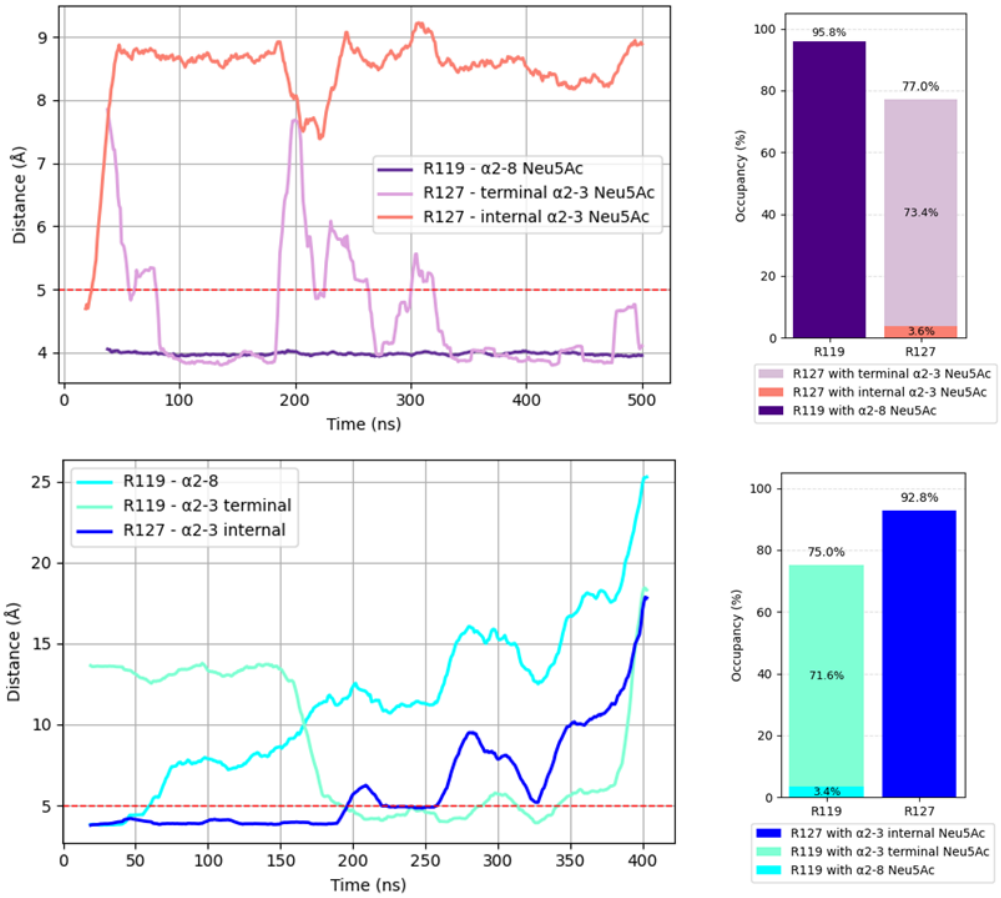


Figure 4.8. *Left*, *top and bottom*: Distance plots showing the interactions between Siglec-10 Arginine residues and the sialic acid moieties of GT1b over 500 ns (MD1). The red dashed line at 5 Å indicates the threshold distance for interaction. *Right, top and bottom*: Occupancy analysis for the interactions between Arg and sialic acid residues in MD1 and MD2, respectively. Occupancy is defined as the percentage of simulation frames where the interatomic distance remains below 5 Å. MD1 shows dominant occupancy for R119 α2-8 (95.8%) and R127 terminal α2-3 (77.0%). In MD2, R119 shows dominant occupancy with 71.6% toward the terminal α2-3, and R127 shows 92.8% for internal α2-3.

Our results indicate that CD52 hypersialylation stabilizes its assembly with HMGB1 (Box B) and exposes multivalent sialic acid on the N-glycan. This creates an epitope that the Siglec-10 V-set domain can engage through multiple interactions (R119/R127).

4.4 Discussion

Our results indicate that sialylation regulates both the assembly of the CD52/HMGB1 complex and its recognition by Siglec-10. Hyper-sialylation of CD52 changes its conformation, promotes extended peptide states, and enables direct interaction with HMGB1-BoxB through the di-sialylated core-2 O-glycan at T8. This positions the N-glycan at N3 toward the solvent, exposing Neu5Ac moieties for Siglec-10 binding. In Siglec-10, two arginines (R119 and R127) in the V-set pocket stabilise sialic acid contacts, explaining the receptor's preference for multivalent, hyper-sialylated CD52 epitopes when bound to HMGB1-BoxB.

The dual-arginine structure of Siglec-10 is a unique feature within the Siglec family. The conserved Arg119 anchors the carboxylate group of sialic acid, while Arg127 provides an additional anchoring point. This arrangement explains avidity for multivalent sialylation, tolerance to micro-heterogeneity in glycan presentation, and the functional readouts seen in cellular assays where blocking Siglec-10 restores T-cell activity. Ligand multivalency and receptor bivalency (Arg119/Arg127) together support sustained immune inhibitory signaling.

Our mechanistic model predicts that disrupting T8 O-glycan sialylation will weaken (or negate) HMGB1-BoxB binding, reducing Siglec-10 engagement, even where the N-Glycan at position N3 remains sialylated. In turn, reducing the degree of sialylation of the N3 glycan alone, while retaining the O3 glycan sialylation, is expected to decrease Siglec-10 avidity without eliminating HMGB1-BoxB capture. Altering Arg127, through mutation or targeted antibodies, is expected to lower preference for multivalent sialic epitopes more than altering Arg119. Additionally, since CD52 is GPI-anchored *in vivo*, but soluble in some regulatory contexts, membrane clustering may further increase multivalency and enhance Siglec-10 signaling.

These findings identify sialylation as a modifiable factor. Increasing sialylation or stabilizing the CD52/HMGB1-BoxB complex may enhance immune suppression where reduced inflammation is needed, such as in autoimmunity. In contrast, blocking Siglec-10 could restore cytotoxic lymphocyte activity, consistent with this, Sobczak *et al. (2025)* showed that antibody-mediated Siglec-10 blockade with the S10A monoclonal antibody restores CAR-T cell cytotoxicity against Siglec-10–expressing target cells[30]. Both ligand engineering, by adjusting N3 or T8 sialylation, and receptor targeting, using antibodies or small molecules for the Arg119/Arg127 pocket, represent complementary strategies.

4.5 Conclusions

In summary, sialylation organizes the entire CD52 → HMGB1 → Siglec-10 axis, acting first as a structural switch that extends and orients the ligand and then as a chemical key read by Siglec-10's dual-arginine pocket. In the broader context of this thesis, these findings complement the previous chapter on low-density lipoprotein receptor-related protein 1 (LRP1). In that context, sialylation functioned as a local inhibitor of glycan-mediated micro-contacts at the RAP interface, while here, sialylation enables and stabilizes interactions with the receptor, by extending CD52 and facilitating productive contacts with HMGB1, by selecting of a specific subdomain (BoxB) and subsequently by binding Siglec-10. This comparison highlights a central theme: the effect of sialylation is context dependent, determined by the location of sialic acids (linker vs epitope), their presentation (monovalent vs multivalent), and the specific protein surface that recognizes them (CR repeat domain vs Siglec V-set domain).

Together with the LRP1 case study, the results in this Chapter illustrate the versatility of sialic acid as a biological modulator, that can inhibit glycan-assisted contacts in one protein, while enabling high-avidity recognition with another. In the next Chapter, we will place these findings in the broader context of the Siglec family, linking binding site structure to preferences for different sialic acid presentations.

References

1. Wei SC, Duffy CR, Allison JP. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. Cancer Discov. 2018 Sep;8(9):1069–86.
2. Munkley J, Elliott DJ. Hallmarks of glycosylation in cancer. Oncotarget. 2016 Jun 7;7(23):35478–89.
3. van Houtum EJH, Büll C, Cornelissen LAM, Adema GJ. Siglec Signaling in the Tumor Microenvironment. Front Immunol. 2021 Dec 13;12:790317.
4. Jiang KY, Qi LL, Kang FB, Wang L. The intriguing roles of Siglec family members in the tumor microenvironment. Biomark Res. 2022 Apr 13;10(1):22.
5. Duan S, Paulson JC. Siglecs as Immune Cell Checkpoints in Disease. Annu Rev Immunol. 2020 Apr 26;38:365–95.
6. Barkal AA, Brewer RE, Markovic M, Kowarsky M, Barkal SA, Zaro BW, et al. CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature. 2019 Aug;572(7769):392–6.
7. Meyer SJ, Linder AT, Brandl C, Nitschke L. B Cell Siglecs-News on Signaling and Its Interplay With Ligand Binding. Front Immunol. 2018 Dec 3;9:2820.
8. Blixt O, Collins BE, van den Nieuwenhof IM, Crocker PR, Paulson JC. Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein. J Biol Chem. 2003 Aug 15;278(33):31007–19.
9. Chen GY, Tang J, Zheng P, Liu Y. CD24 and Siglec-10 selectively repress tissue damage-induced immune responses. Science. 2009 Mar 27;323(5922):1722–5.
10. Bandala-Sanchez E, G Bediaga N, Goddard-Borger ED, Ngui K, Naselli G, Stone NL, et al. CD52 glycan binds the proinflammatory B box of HMGB1 to engage the Siglec-10 receptor and suppress human T cell function. Proc Natl Acad Sci U S A. 2018 Jul 24;115(30):7783–8.
11. Treumann A, Lifely MR, Schneider P, Ferguson MA. Primary structure of CD52. J Biol Chem. 1995 Mar 17;270(11):6088–99.
12. Shathili AM, Bandala-Sanchez E, John A, Goddard-Borger ED, Thaysen-Andersen M, Everest-Dass AV, et al. Specific Sialoforms Required for the Immune Suppressive Activity of Human Soluble CD52. Front Immunol. 2019 Aug 27;10:1967.
13. Rashidi M, Bandala-Sanchez E, Lawlor KE, Zhang Y, Neale AM, Vijayaraj SL, et al. CD52 inhibits Toll-like receptor activation of NF-κB and triggers apoptosis to suppress inflammation. Cell Death Differ. 2018 Feb;25(2):392–405.
14. Bandala-Sanchez E, Zhang Y, Reinwald S, Dromey JA, Lee BH, Qian J, et al. T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10. Nat Immunol. 2013 Jul;14(7):741–8.
15. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct 11;7:539.
16. Ives CM, Singh O, D'Andrea S, Fogarty CA, Harbison AM, Satheesan A, et al. Restoring protein glycosylation with GlycoShape. Nat Methods. 2024 Nov;21(11):2117–27.
17. Kutzner C, Kniep C, Cherian A, Nordstrom L, Grubmüller H, de Groot BL, et al. GROMACS in the Cloud: A Global Supercomputer to Speed Up Alchemical Drug Design. J Chem Inf Model. 2022 Apr 11;62(7):1691–711.
18. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force

field. Carbohydrates. J Comput Chem. 2008 Mar;29(4):622–55.
19. Nayar D, Agarwal M, Chakravarty C. Comparison of Tetrahedral Order, Liquid State Anomalies, and Hydration Behavior of mTIP3P and TIP4P Water Models. J Chem Theory Comput. 2011 Oct 11;7(10):3354–67.
20. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.
21. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications. 1995 Sep 2;91(1-3):1–41.
22. Lee TS, Cerutti DS, Mermelstein D, Lin C, LeGrand S, Giese TJ, et al. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. J Chem Inf Model. 2018 Oct 22;58(10):2043–50.
23. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015 Sep 1;1-2:19–25.
24. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008 Aug;29(11):1859–65.
25. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. 2015 Dec 3 [cited 2025 Oct 2]; Available from: https://pubs.acs.org/doi/full/10.1021/acs.jctc.5b00935
26. Rowell JP, Simpson KL, Stott K, Watson M, Thomas JO. HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail. Structure. 2012 Dec 5;20(12):2014–24.
27. Forgione RE, Di Carluccio C, Guzmán-Caldentey J, Gaglione R, Battista F, Chiodo F, et al. Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. iScience. 2020 Aug 21;23(8):101401.
28. Fadda E. Molecular simulations of complex carbohydrates and glycoconjugates. Curr Opin Chem Biol. 2022 Aug;69:102175.
29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Jul 15;596(7873):583–9.
30. Website [Internet]. Available from: doi: https://doi.org/10.1101/2025.06.10.658867
31. Schmidt EN, Lamprinaki D, McCord KA, Joe M, Sojitra M, Waldow A, et al. Siglec-6 mediates the uptake of extracellular vesicles through a noncanonical glycolipid binding pocket. Nature Communications. 2023 Apr 22;14(1):1–17.

Chapter 5. Glycolipid recognition and binding by Siglec-6 hinges on interactions with the cell membrane

The following text is adapted from the paper:

D'Andrea, S., Schmidt, E. N., Bui, D., Singh, O., Han, L., Mahal, L. K., Klassen, J., S., Macauley, S.,M., & Fadda, E. (2025). Glycolipid recognition and binding by Siglec-6 hinges on interactions with the cell membrane. *bioRxiv*, 2025-06. *Communication Biology* (*in press*) doi**:**https://doi.org/10.1101/2025.06.21.660838

5.1 Introduction

Sialic acid-binding immunoglobulin (Ig)-like lectins (Siglecs) are immunoregulatory transmembrane receptors that recognise sialylated glycoconjugates. Siglecs modulate immune responses by triggering inhibitory or activating signals, maintaining immune homeostasis[1–4]. The ability to distinguish self from non-self through sialic acid recognition makes Siglecs critical regulators of immune function and attractive therapeutic targets for autoimmune diseases, cancer, and infections[5], as well as for the development of chimeric antigen receptor (CAR) T-cell immunotherapeutic strategies[6–8]. Humans Siglecs count 14 distinct members that are commonly classified into two groups, see Figure 1. Evolutionarily conserved Siglecs[9], here labelled as Group 1, are present in all mammalian species[10] and have a low degree of homology, i.e. 25–30% sequence identity, namely sialoadhesin (Siglec-1), CD22 (Siglec-2), myelin-associated glycoprotein (MAG or Siglec-4), and Siglec-15. Group 2 comprises CD33-related Siglecs (CD33rSiglecs), which are not conserved across species, but share a higher sequence similarity (~50–99%)[1], suggesting that diversification could have occurred in response to species-specific selective pressures[10].

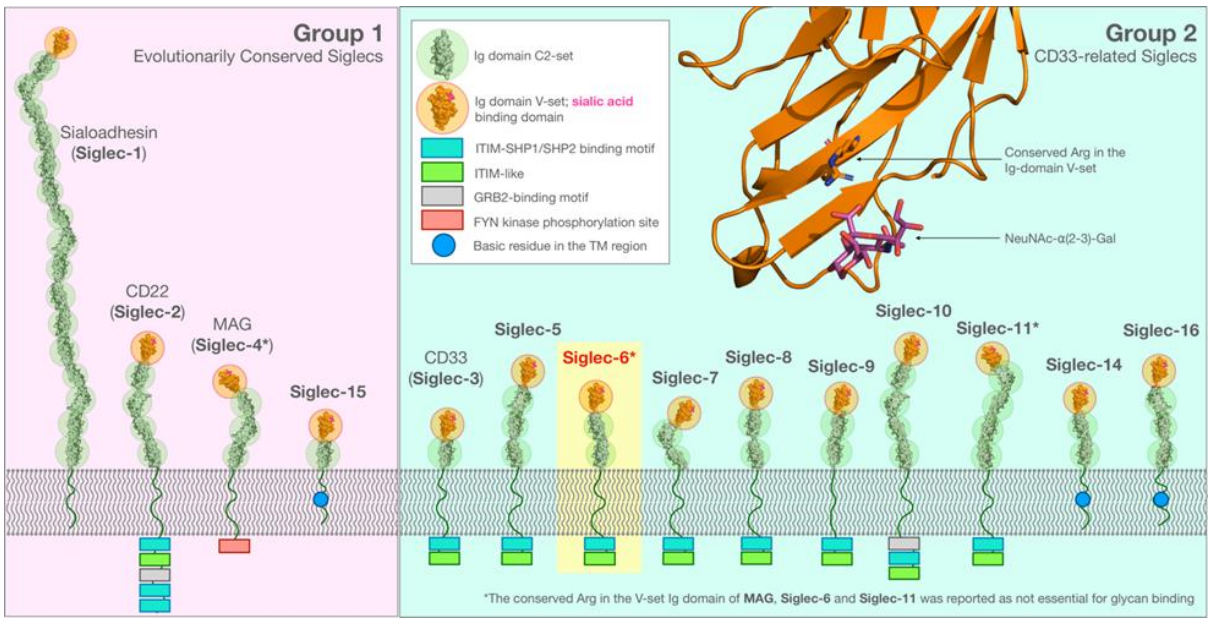


Figure 1. Schematic representation of human Siglecs. Evolutionarily conserved Siglecs (Group 1) are shown on the left-hand side panel (pink box). The CD33-related Siglecs (Group 2) are shown on the right-hand side panel, (light-green box). The Siglecs extracellular domains are represented by a composition of Ig domains (C2-set) rendered in grey as a solvent accessible surface, highlighted within green circles. The terminal Ig V-set binding domain is rendered in orange and highlighted within an orange circle with a bound sialic acid in magenta (PDB 1OD9). Cytoplasmic domains are indicated with boxes and labelled according to the legend. Siglec-6 is highlighted within a yellow box as it is the focus of this work. All structural elements in this image were rendered with pymol (www.pymol.org) and incorporated in an original design based on Figure 1 in ref.[2]. In the insert on the top right-hand side, the conserved Arg is shown in the 3D structure of the Ig domain V-set of Siglec-6 (orange cartoons) bound to the NeuNAc-a(2-3)-Gal fragment (sticks with purple C atoms, red O atoms and blue N atoms) as an example.

In terms of 3D structure, human Siglecs are characterized by a different number of Ig C2-set domains terminating with an Ig V-set domain that contains the canonical binding site, where a conserved Arg binds sialosides, see **Fig. 1**. The V-set domain includes a structurally flexible C-C' loop, characterised by a high sequence variability across Siglecs. Both of these features allow the C-C' loop to play a critical role in modulating recognition[1] and contribute to the Siglecs' binding specificity[5,11].

Target sialosides can be presented to Siglecs within a wide variety of different structural contexts, namely functionalising glycans linked to membrane bound or soluble glycoproteins, terminating the chains of membrane-embedded glycolipids or on secreted oligosaccharides[12,13]. In most Siglecs, the loss of the conserved Arg in the V-set domain results in loss of binding, except in Siglec-6, where binding is only weakened by the loss of the canonical Arg, as demonstrated in recent work by our collaborators[14]. Earlier work

indicates that MAG and Siglec-11[15–18] can also operate independently of the canonical Arg, suggesting an evolution of alternative binding mechanisms in selected members of the Siglec family. To this end, previous work[14] by our collaborators suggested the involvement of a non-canonical Arg in the V-set domain in the binding mechanism of Siglec-6, spearheading further investigation.

In this work, we used computational and molecular biology approaches to investigate the recognition and binding specificity of Siglec-6 for monosialylated gangliosides, namely GM1, GM2 and GM3, which are known Siglecs ligands[12,14].

Molecular dynamics (MD) simulations have been used extensively and successfully to investigate at the atomistic and molecular levels of details the distribution and dynamics of gangliosides, and other glycolipids, in lipid bilayers[19–21]. Coarse graining (CG) methods have been particularly useful to explore the enormous complexity of such highly dynamic systems at biologically relevant timescales, where membranes can bear a wide range of chemically diverse lipids, in different concentrations, with different structure flexibilities and diffusion properties[22]. Although the CG force fields sophistication has been greatly improved recently[23], allowing users to obtain important insights into membrane composition, structure and biology[24,25], in this work we needed an atomistic approach for our simulation to characterise the Siglec-6 recognition mechanism and how this facilitates a discrimination between structurally and chemically similar monosialylated gangliosides. The membrane used in the simulations is set to match the experimental composition, selected based on the results of extensive testing in earlier work[14] with varied lipid compositions and concentration of cholesterol.

We first gathered a detailed understanding of the binding mechanism of Siglec-6 at the atomistic-level of detail through extensive molecular dynamics (MD) simulations. Based on structural analysis and the results of MD simulations, we were able to exclude the role of any other Arg residues in the V-set domain of Siglec-6. Instead, we found that the interaction between the Siglec-6 and the gangliosides still involves the canonical binding site, with the conserved Arg122 coordinating the carboxylic group of the sialic acid, but not continuously. This occurs because the interaction with Arg122 is structurally stabilised and energetically supplemented by an additional contact that Siglec-6 establishes with the lipid bilayer through residues Trp127 and Lys126, see **Fig. 2.a**. Mutagenesis and cell binding assays, both on liposomes and on nanodisks, support the atomistic-scale insight, also demonstrating a unique interaction between Siglec-6 and the

bilayer devoid of GM1. Additionally, our collaborators show through Concentration Independent (COIN)-Catch-and-Release (CaR)-native mass spectrometry (nMS) assays26 that the binding of free (not membrane-linked) oligosaccharides hinges entirely on the canonical Arg122. In the following sections we present the results and discuss the biological implications of the unique specialization of Siglec-6 in the recognition and binding of membrane-bound glycan epitopes with a direct comparison to other members of the Siglecs family.

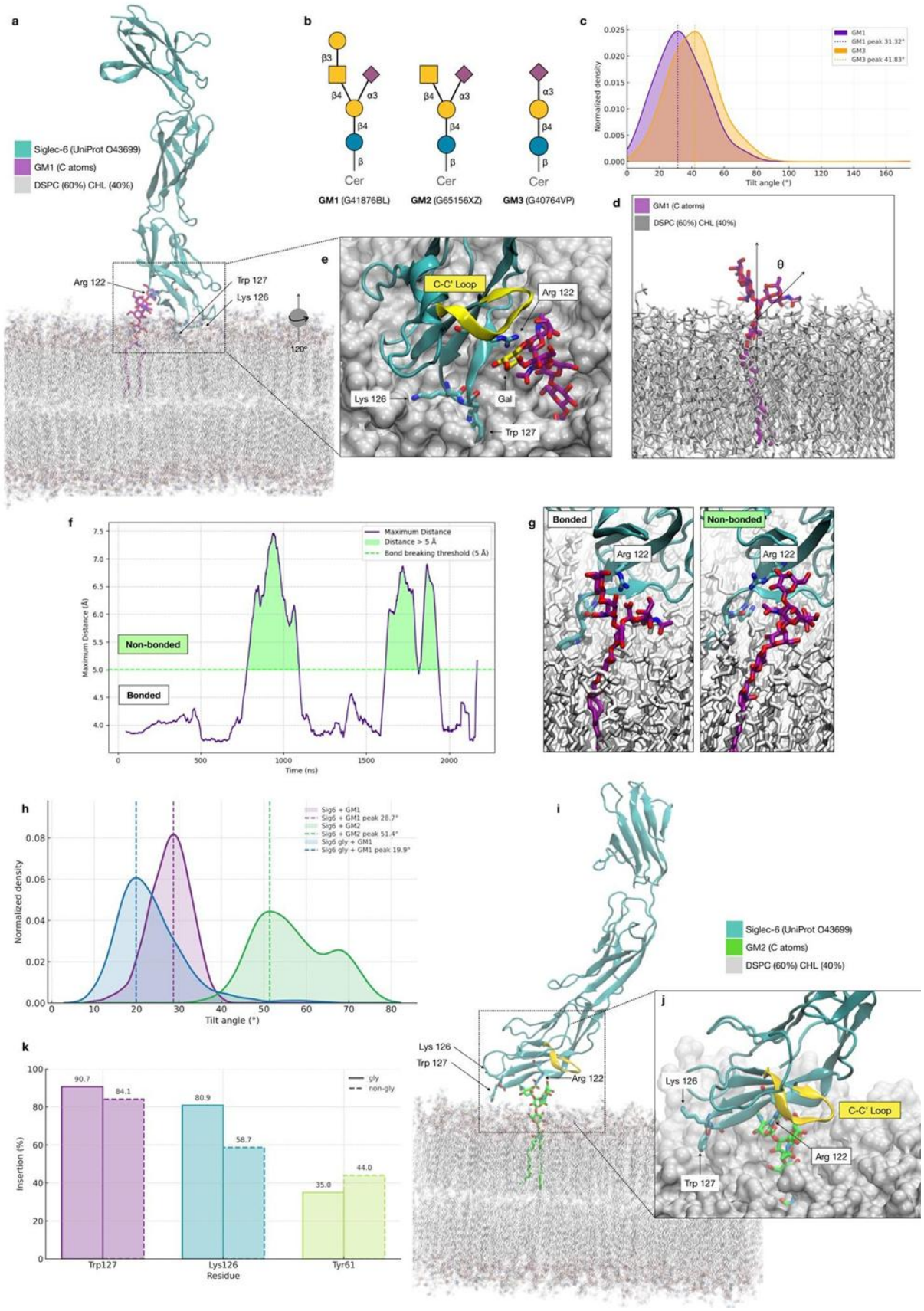


Figure 2. a. 3D structure of Siglec-6 in complex with GM1 embedded in a lipid bilayer with composition distearoylphosphatidylcholine (DSPC) 60% and cholesterol 40%. In this representative snapshot from the MD trajectory, collected at 0.715 μs from the equilibrated MD ensemble, Arg 122 is engaged in a salt bridge with the Neu5Ac of GM1. The protein is represented

with cyan cartoon rendering, the GM1 with sticks and C atoms in purple, O in red and N in blue. DSPC and cholesterol are rendered with semi-transparent sticks, with C atoms in grey, O in red, N in blue and P in yellow. Key residues are labelled with numbering corresponding to the human Siglec-6 (UniProtID O43699). b. Structures of GM1, GM2 and GM3 represented with the SNFG nomenclature. Labels below each structure include the oligosaccharides GlyTouCan IDs. **c**. Kernel Density Estimates (KDE) analysis of the tilt angle values measured though the 1.0 μs MD trajectories ran for isolated GM1 (purple) and GM3 (orange). KDE maxima are 31.32º and 41.83º measured for GM1 and GM3, respectively. **d**. 3D structure of an isolated GM1 molecule (sticks with C atoms in purple, O in red and N in blue) embedded in one of the bilayer leaflets (sticks with all atoms in grey) used to represent the axes used to measure the tilt angle (θ) indicating the orientation relative to the bilayer of the Neu5Ac and thus its accessibility. **e**. Close-up view on the Siglec-6 binding site obtained through a counterclockwise rotation of approximately 120º relative to the structure in panel a). Key residues are labelled, while the embedding of the Trp127 sidechain in the bilayer is highlighted by a surface rendering of the lipids. The terminal Gal of GM1 is highlighted with C atoms in yellow for his role in orienting the C-C' loop in the bound complex, also shown in yellow. **f.** Time evolution along the MD trajectory of the distance (Å) between the Arg122 and the Neu5Ac carboxylate group. Data points correspond to the largest distance value calculated between four pairs of atoms R122-NH1(2) and O11(12)-Neu5Ac. A distance of 5 Å was chosen as a threshold for the formation of the salt bridge. **g**. Comparison of the canonical binding site of Siglec-6 in complex with GM1, the image on the left shows a snapshot from the MD simulation when Arg122 is engaged in the salt bridge, collected at 0.715 μs, while the image on the right shows a snapshot when the salt bridge is broken, collected at 1.028 μs. **h**. KDE analysis of the tilt angle (º) indicating the orientation of the Ig V-set domain relative to the membrane plane, in purple and in green for the complexes with the non-glycosylated Siglec-6 and GM1 and GM1, respectively, and in blue for the complex with the glycosylated Siglec-6 and GM1. The normal line passes through the Ig V-set centre of mass (COM), while the Ig V-set axis is set to pass between the COM and an auxiliary point located at 10 Å from the COM obtained from the principal component 1 from the dynamics of the C*a* atoms, representative of the spread along the eigenvector's direction. **i**. 3D structure of Siglec-6 in complex with GM2 (sticks with C atoms in green) embedded in a lipid bilayer with composition distearoylphosphatidylcholine (DSPC) 60% and cholesterol 40%. In this representative snapshot from the MD trajectory, collected at 1.0 μs from the equilibrated MD ensemble, Arg 122 is still engaged in a salt bridge with the Neu5Ac of GM2, but the Siglec-6 is not embedded in the bilayer through the Trp 127 and Lys 126 due to a conformational change of the C-C' loop (yellow cartoon section). Colouring described in the legend and rendering style as in panel a). **k**. Insertion (%) of the sidechain COM of the Trp127, Lys126 and Tyr61 below the membrane surface calculated along the MD trajectories for the complex with GM1 with the non-glycosylated Siglec-6 (full line) and the glycosylated Siglec-6 (dashed line). **j**. Close-up view on the Siglec-6 binding site with key residues labelled. Molecular rendering done with Visual Molecular Dynamics27 (VMD; https://www.ks.uiuc.edu/Research/vmd/). Tilt angle analysis and other graphs done with matplotlib (https://matplotlib.org/) and seaborn libraries in python (https://seaborn.pydata.org/).

5.2 Computational Methods

5.2.1 MD simulations of Siglec-6 with multiple GM1os.

The 3D structure of Siglec-6 was obtained from the AlphaFold database (AF-O43699-F1) [6,7], and the structures of GM1os were retrieved from the GlycoShape database[5]. One GM1os molecule was manually placed at the canonical sialic acid-binding site near Arg122, while six additional GM1os molecules were randomly distributed around the protein surface to investigate for potential alternative binding sites. The protein was parameterized using the AMBER ff14SB force field, and the glycans with GLYCAM06j-1. Simulations were performed using AMBER18 . Following 500,000 steps of steepest descent minimization, the system was heated in two stages (0–100 K and 100–300 K, 500 ps each) in the NVT ensemble using Langevin dynamics ($\gamma\_ln = 1.0\ ps^{-1}$). Equilibration was carried out in the NPT ensemble for 500 ps at 1 atm using a Berendsen barostat. All restraints on the protein heavy atoms were then removed, and three independent 300 ns production runs were performed from different initial velocities. Due to intrinsic limitations of the force field in modeling systems with multiple free glycans, the simulations were halted prematurely[3]. Residue specific occupancies were calculated by monitoring persistent contacts (within 5 Å) between the carboxyl group of the sialic acid and the guanidinium group of ARG side chains over time. The results of these simulations are included in Supplementary Material as Appendices.

5.2.2 MD simulations of isolated GM1 and GM3 in a lipid bilayer.

Two independent systems were prepared, each embedding either a GM1 and GM3 ganglioside within a symmetric 130 Å × 130 Å lipid bilayer composed of 60% 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and 40% cholesterol (CHL1), using the CHARMM-GUI Membrane Builder tool[8]. All molecular dynamics (MD) simulations were performed with AMBER18, using the CHARMM36m force field to parameterize the gangliosides and lipids. Each system underwent energy minimization for 5000 steps (2500 steps of steepest descent followed by 2500 steps of conjugate gradient minimization). Equilibration followed the six-step CHARMM-GUI protocol [9], during which positional restraints of 10 kcal/mol·Å² were initially applied and then gradually reduced to 5 kcal/mol·Å² for the protein and 2.5 kcal/mol·Å² for the membrane, before being completely removed. Temperature was maintained at 315.15 K using Langevin dynamics ($\gamma\_ln = 1.0\ ps^{-1}$), and pressure was controlled semi-isotropically at 1 atm using the Berendsen barostat. Periodic boundary conditions were applied throughout. Long-

range electrostatics were treated using the Particle Mesh Ewald (PME) method with an 11 Å cutoff. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm, enabling a 2 fs integration time step. Each system was simulated for 1 μs. The orientation of the sialic acid headgroup (Neu5Ac) was quantified by calculating the tilt angle (θ), defined as the angle between the vector connecting atom C1 of the terminal galactose and atoms C2 and C3 of the Neu5Ac residue, relative to the axis perpendicular to the membrane plane. Tilt angle distributions were computed over the entire MD trajectories for both GM1 and GM3.

5.2.3 MD simulations of GM1 and GM2 in complex with Siglec-6.

The Siglec-6/GM1 complex was constructed by aligning the conformation of GM1, obtained from the membrane embedded simulation, to the sialic acid moiety in the crystal structure of Siglec-3 bound to a sialoside analogue (PDB: 7AW6). The structure of Siglec-6 (AF-O43699) was validated by structural alignment with Siglec-3, resulting in a backbone RMSD of 0.704 Å. A second system, containing GM2, was generated by removing the terminal Gal residue from GM1. Both complexes were embedded in the same DSPC/cholesterol bilayer described above, and the simulations were carried out using AMBER v.22[10], following the same multistep equilibration and production protocol. In the GM1 system, a distance restraint (5 kcal/mol·Å²) was applied between the side chain of Arg122 and the carboxyl group of the Neu5Ac to stabilize the salt bridge during the initial 400 ns of the 2.5 μs production simulation. The same approach was used for the GM2 complex; in both 1-μs replicas, removal of the restraint led to destabilization of the complex, indicating a loss of stable interaction between Siglec-6 and GM2 in the absence of the terminal Gal residue. All MD trajectories were analysed using Python3 scripts written in-house. Distances, occupancies, and tilt angles were calculated frame-by-frame using the *cpptraj* module in AMBER v.22 and the graphical user interface VMD[11]. Kernel Density Estimates (KDE) and all other plots were generated using the *matplotlib* (https://matplotlib.org/) and *seaborn* (https://seaborn.pydata.org/) libraries.

5.2.4 Modelling of the glycosylated Siglec-6

The AlphaFold-predicted structure of Siglec-6 (AF-O43699-F1) was scanned using GlcNAc Scan tool[5], which identified seven putative *N*-glycosylation sites across the extracellular region. The glycan 3D structures were retrieved from the GlycoShape database and added to each predicted sequon using Re-Glyco, which optimizes the

orientation of the attached N-glycans (GlyTouCan ID: G99129GB) to minimize steric clashes and ensure proper linkage geometry. The glycosylated PDB output from Re-Glyco was then used as input in CHARMM-GUI Membrane Builder to generate the membrane system. MD simulations were performed as described in the Methods section "MD simulations of GM1 and GM2 in complex with Siglec-6".

5.3 Results

5.3.1 Orientation and accessibility of the ganglioside epitopes in the bilayer.

We used all-atom classical MD simulations to investigate the mechanism that regulates ganglioside recognition and binding in Siglec-6 and to determine its ligand preference. Earlier work by our collaborators[14] shows that while Siglec-6 binds GM1 in an Arg-independent manner, it does not bind GM2 or GM3, which differ from GM1 by the lack of the terminal Gal and Gal-β(1-3)-GalNAc, respectively, see **Fig. 2.b**. To verify that this effect was not determined by a different degree of exposure of the sialic acid through the membrane in the different gangliosides, we ran 1 μs MD simulations of isolated GM1 and GM3, as the two extremes in the monosialylated ganglioside series, in duplicates, with one ganglioside embedded in the upper and one in the lower leaflet of the bilayer, see **Fig. 2.d**. The analysis of the tilt angle (θ) describing the orientation of the Neu5Ac along the trajectory relative to an axis perpendicular to the plane of the bilayer, see **Fig. 2.c**, shows that the Neu5Ac is equally exposed in GM1 and GM3, and thus equally accessible for recognition and binding. These results are in excellent agreement with previous work[20].

5.3.2 Structure of the Siglec-6/GM1 complex and binding mechanism.

From a representative conformer of the GM1 embedded in the membrane obtained from the MD simulations described above, we built the 3D structure of the complex with Siglec-6, with the canonical Arg122 involved in a salt bridge with the Neu5Ac of the ligand. To date, no experimental 3D structure of Siglec-6 is available in public repositories, however, the common architecture and high degree of sequence and structure homology across Siglecs support confident structure predictions by machine learning (ML) leveraging on existing structural data in the RCSB PDB. In this work we used the AlphaFold (AF)[21] model AF-O43699-F1 of Siglec-6 deposited in the EBI-EMBL AF protein structure database[22]. , with 69% of the residues predicted with very high confidence (pLDDT > 90) and 26% with high confidence (pLDDT > 70), **see Fig. S.1**.

Structural alignment of the AF model with the high-resolution crystal structure of Siglec-3 (PDB 7AW6 at 1.95 Å), which shares 60% sequence identity with Siglec-6, results in an RMSD value of 0.704 Å calculated over backbone atoms. The Siglec-6/GM1 complex was built by structural alignment of the membrane bound GM1 from the MD simulations to the Neu5Ac in the sialoside analogue bound to Siglec-3 in the crystal structure (PDB 7AW6), to which the Siglec-6 was aligned. As an important note, the orientation of the Siglec-6 was selected to complement the chosen equilibrium conformation of the GM1 ligand embedded in the membrane and no alteration to that was made to enhance structure complementarity. Comparison between the resulting Siglec-6/GM1 complex and the Siglec-3/sialoside-analogue complex shows that the glycan in the sialoside analogue is rotated by approximately 180° relative to the membrane bound GM1. This suggests that free and bound sialosides could potentially bind Siglecs in different orientations, due to the C2V symmetry of the carboxylate group and the broad accessibility of the Siglecs canonical binding site, discussed further in the following sections.

The Siglec-6 sequence (453 aa) carries seven *N*-glycosylation sequons, with only one in the V-set domain at N103. Although we have no information on the glycosylation of Siglec-6 from mast cells or memory B-cells, earlier work[23] indicates that occupancy of these sites is low or partial in normal (non-aberrant) syncytiotrophoblasts, while it appears to be higher in recombinant constructs (317 aa). Molecular weights of 50 kDa are reported for Siglec-6 from normal pregnancy and preeclampsia placenta cells lysates[23], which corresponds to a low level of *N*-glycosylation. In preeclampsia placental lysates, molecular weights of up to 70 kDa are found, which suggests *N*-glycosylation at potentially all seven sites of Siglec-6. Molecular weights from 57 up to 90 kDa are reported on commercial sites for recombinant products. In the absence of more detailed insight on the nature and occupancy of the sites, we ran simulations on two 3D models of the Siglec-6/GM1 complex, one with a non-glycosylated Siglec-6 and the other with a fully glycosylated Siglec-6 as a control, with all sites occupied with a biantennary mono galactosylated complex *N*-glycan (GlyTouCan ID G99129GB; 1478.5 Da). Below we focus our attention on the results obtained for the non-glycosylated system, as it corresponds more closely to Siglec-6 expression in normal placental syncytiotrophoblasts. The details on the structure and dynamics of the complex with the fully N-glycosylated Siglec-6 are included in the Supplementary Material. A comparison between the results obtained shows that N-glycosylation of Siglec-6 does not affect binding selectivity nor the mechanism.

After a multi-stage structure equilibration, see Methods section in the Supplementary Material for details, the conformational dynamics of the Siglec-6/GM1 complex was analysed with MD simulations ran with a deterministic sampling scheme. We ran a single trajectory of 2.4 μs with the non-glycosylated Siglec-6, and two uncorrelated replicas of for 1,1 μs and 1,0 μs with the N-glycosylated Siglec-6 to explore the stability of the interactions and to monitor functionally relevant conformational changes, energetically accessible at room temperature. The results obtained indicate that, 1) the Neu5Ac of GM1 forms a salt bridge with the canonical Arg122, and 2) this interaction is structurally and energetically supplemented by the insertion in the membrane of the Siglec-6 V-set domain through the Trp127 indole sidechain and through ancillary binding of the phospholipid heads by the adjacent Lys126, see **Fig. 2.a.** Furthermore, the terminal Gal of GM1 interacts with the C-C' loop (aa 70 to 75) through hydrogen bonds with Asp70 and Glu73. In the context of this interaction network, the salt bridge between the canonical Arg122 and the Neu5Ac of GM1 appears to be energetically dispensable at room temperature. Indeed, along the MD trajectory we observe that Arg122 interrupts its contact with Neu5Ac for significant amounts of sampling time, see **Fig. 2.f** and **2.g**. Population analysis shows that the salt bridge is formed only 68.5% of the total simulation time. After these intervals, the salt bridge gets restored with no disruption of the Siglec-6 bound conformation, as shown by orientation of the Ig V-set domain relative to the membrane, see **Fig. 2.h**, which is primarily stabilised by its interactions with the membrane, supported by contacts between the C-C' loop and the terminal Gal of GM1. The analysis of the tilt angle in **Fig 2.h** shows that the orientation of the Ig V-set in the complex with the fully glycosylated Siglec-6 is slightly more orthogonal relative to the non-glycosylated, except when contact with GM1is lost, indicated by the tail in the KDE distribution. The results of the analysis of the degree of insertion of the Ig V-set residues in the membrane is shown in the bar chart in **Fig.2.k**, where we measured the (%) of the frames in which the centre of mass of the sidechain was below the membrane surface. The results clearly indicate the contributions of Lys 126 and Trp127, in both glycosylated and non-glycosylated Siglec-6 complexes. A minor contribution to binding could be given by Tyr61, whose centre of mass is only sporadically found below the membrane surface when the Ig V-set is in a bound conformation. Yet its sidechain inserts into the bilayer to a larger extent when the Siglec-6 detaches from the ligand see **Fig. S.2** for the timeframe trace.

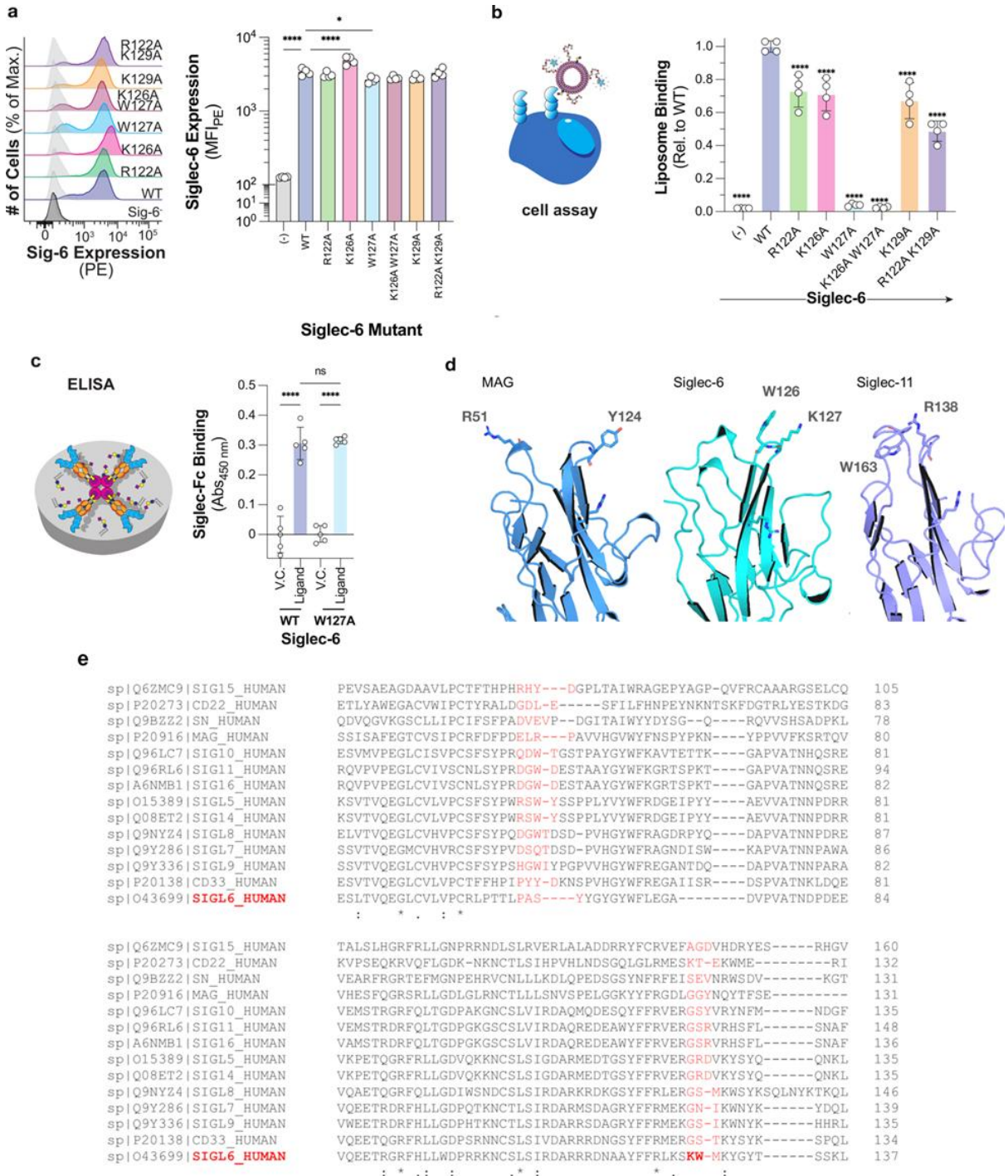
5.3.3 The structure of the ganglioside determines Siglec-6 binding selectivity.

Earlier work by our collaborators[14] shows that Siglec-6 binds GM1 selectively when embedded in a membrane, but not GM2 or GM3. Our results show that the terminal Gal of GM1 likely contributes to the stability of the complex through interactions with the C-C' loop. To test if our binding model supports the earlier experimental data and if the interaction between the C-C' loop and the terminal Gal is indeed critical, we built the 3D structure of the Siglec-6/GM2 complex from a conformation selected from the equilibrated ensemble of the Siglec-6/GM1 complex, by removing the terminal Gal of GM1, and thus biasing the new model system towards a stable conformation. We ran two independent MD trajectories with production runs of 1.0 μs and 0.5 μs each from uncorrelated starting structures. Both simulations show that the absence of the terminal Gal triggers a conformational change in the flexible C-C' loop, which can extend into the space that was occupied by the Gal in the complex with GM1, see **Fig. 2.g** and **2.i**. This shift contributes to pulling the Ig V-set domain away from the membrane, lifting out of the bilayer the embedded Trp127 and Lys126 sidechains, ultimately compromising the stability of the complex. The change of the orientation of the Ig V-set domain in the complex with GM2 is highlighted by the analysts of the tilt angle in **Fig. 2.k.** Because GM3 is even shorter, missing the terminal Gal-β(1-3)-GalNAc relative to GM1, we expect the same outcome.

5.3.4 Siglec-6 binding of ganglioside-enriched liposomes.

As the MD simulations predicted Lys126 and Trp127 in mediating Siglec-6-ganglioside binding when the ganglioside is presented from a lipid bilayer, we created point K126A and W127A mutations and stably expressed these in CHO cells. Results are shown in **Fig. 3**. Additionally, as we had previously demonstrated that Arg122 was not required for Siglec-6 to engage with sialosides presented from a bilayer[14], we posited that another positively charged residue in the V-set may compensate in the absence of Arg122. We analysed the binding pocket of Siglec-6 and found that Lys129 is in relatively close spatial proximity to Arg122 and hypothesized that Lys129 may be able to compensate in the absences of Arg122. Expressions of the Siglec-6 mutants were validated by flow cytometry to be expressed at similar levels as WT Siglec-6, see **Fig. 3.a**. Therefore, we tested these Siglec-6-expressing cells for their ability to engage fluorescent liposomes bearing a neoglycolipid that we previously optimized as a Siglec-6 ligand[14,24], see **Fig. 3.b**. In this assay, two controls were used: untransfected CHO cells and liposomes lacking

the neoglycolipid. We observed a modest 25% reduction, which was statistically significant, in liposome binding to R122A, K126A, and K129A Siglec-6 cells. In line with the results from our previous study, we observed that liposome binding was only modestly decreased (~75% of WT) upon the mutation of the canonical arginine residue (Arg122), a feat unique to Siglec-6. The double mutant of R122A and K129A Siglec-6 showed a greater reduction in binding compared to each single mutant (~50% of WT) but was still above background signal, suggesting that additional interactions contribute to recognition of GM1. However, in the W127A mutant, nearly all binding to the glycolipid liposomes was lost (~5% of WT). Moreover, a near complete loss in binding was observed to the K126A W127A double mutant. These results demonstrate that Trp127 is the major driving force for the recognition of glycolipids presented from a bilayer by Siglec-6.




```
sp|Q6ZMC9|SIG15_HUMAN      PEVSAEAGDAAVLPCTFTHPHRHY---DGPLTAIWRAGEPYAGP-QVFRCAAARGSELCQ   105
sp|P20273|CD22_HUMAN       ETLYAWEGACVWIPCTYRALDGDL-E-----SFILFHNPEYNKNTSKFDGTRLYESTKDG   83
sp|Q9BZZ2|SN_HUMAN         QDVQGVKGSCLLIPCIFSFPADVEVP--DGITAIWYYDYSG--Q----RQVVSHSADPKL   78
sp|P20916|MAG_HUMAN        SSISAFEGTCVSIPCRFDFPDELR---PAVVHGVWYFNSPYPKN----YPPVVFKSRTQV   80
sp|Q96LC7|SIG10_HUMAN      ESVMVPEGLCISVPCSFSYPRQDW-TGSTPAYGYWFKAVTETTK----GAPVATNHQSRE   81
sp|Q96RL6|SIG11_HUMAN      RQVPVPEGLCVIVSCNLSYPRDGW-DESTAAYGYWFKGRTSPKT----GAPVATNNQSRE   94
sp|A6NMB1|SIG16_HUMAN      RQVPVPEGLCVIVSCNLSYPRDGW-DESTAAYGYWFKGRTSPKT----GAPVATNNQSRE   82
sp|O15389|SIGL5_HUMAN      KSVTVQEGLCVLVPCSFSYPWRSW-YSSPPLYVYWFRDGEIPYY----AEVVATNNPDRR   81
sp|Q08ET2|SIG14_HUMAN      KSVTVQEGLCVLVPCSFSYPWRSW-YSSPPLYVYWFRDGEIPYY----AEVVATNNPDRR   81
sp|Q9NYZ4|SIGL8_HUMAN      ELVTVQEGLCVHVPCSFSYPQDGWTDSD-PVHGYWFRAGDRPYQ----DAPVATNNPDRE   87
sp|Q9Y286|SIGL7_HUMAN      SSVTVQEGMCVHVRCSFSYPVDSQTDSD-PVHGYWFRAGNDISW----KAPVATNNPAWA   86
sp|Q9Y336|SIGL9_HUMAN      SSVTVQEGLCVHVPCSFSYPSHGWIYPGPVVHGYWFREGANTDQ----DAPVATNNPARA   82
sp|P20138|CD33_HUMAN       ESVTVQEGLCVLVPCTFFHPIPYY-DKNSPVHGYWFREGAIISR----DSPVATNKLDQE   81
sp|O43699|SIGL6_HUMAN      ESLTVQEGLCVLVPCRLPTTLPAS----YYGYGYWFLEGA--------DVPVATNDPDEE   84
                                 :    * .  : *

sp|Q6ZMC9|SIG15_HUMAN      TALSLHGRFRLLGNPRRNDLSLRVERLALADDRRYFCRVEFAGDVHDRYES-----RHGV   160
sp|P20273|CD22_HUMAN       KVPSEQKRVQFLGDK-NKNCTLSIHPVHLNDSGQLGLRMESKT-EKWME---------RI   132
sp|Q9BZZ2|SN_HUMAN         VEARFRGRTEFMGNPEHRVCNLLLKDLQPEDSGSYNFRFEISEVNRWSDV-------KGT   131
sp|P20916|MAG_HUMAN        VHESFQGRSRLLGDLGLRNCTLLLSNVSPELGGKYYFRGDLGGYNQYTFSE---------   131
sp|Q96LC7|SIG10_HUMAN      VEMSTRGRFQLTGDPAKGNCSLVIRDAQMQDESQYFFRVERGSYVRYNFM------NDGF   135
sp|Q96RL6|SIG11_HUMAN      VEMSTRDRFQLTGDPGKGSCSLVIRDAQREDEAWYFFRVERGSRVRHSFL------SNAF   148
sp|A6NMB1|SIG16_HUMAN      VAMSTRDRFQLTGDPGKGSCSLVIRDAQREDEAWYFFRVERGSRVRHSFL------SNAF   136
sp|O15389|SIGL5_HUMAN      VKPETQGRFRLLGDVQKKNCSLSIGDARMEDTGSYFFRVERGRDVKYSYQ------QNKL   135
sp|Q08ET2|SIG14_HUMAN      VKPETQGRFRLLGDVQKKNCSLSIGDARMEDTGSYFFRVERGRDVKYSYQ------QNKL   135
sp|Q9NYZ4|SIGL8_HUMAN      VQAETQGRFQLLGDIWSNDCSLSIRDARKRDKGSYFFRLERGS-MKWSYKSQLNYKTKQL   146
sp|Q9Y286|SIGL7_HUMAN      VQEETRDRFHLLGDPQTKNCTLSIRDARMSDAGRYFFRMEKGN-IKWNYK------YDQL   139
sp|Q9Y336|SIGL9_HUMAN      VWEETRDRFHLLGDPHTKNCTLSIRDARRSDAGRYFFRMEKGS-IKWNYK------HHRL   135
sp|P20138|CD33_HUMAN       VQEETQGRFRLLGDPSRNNCSLSIVDARRRDNGSYFFRMERGS-TKYSYK------SPQL   134
sp|O43699|SIGL6_HUMAN      VQEETRGRFHLLWDPRRKNCSLSIRDARRRDNAAYFFRLKSKW-MKYGYT------SSKL   137
                                 : * .:  :      .* :             * .    :
```

Figure. 3. a. Expression levels of the Siglec-6 WT and mutants obtained by flow cytometry. **b.** Depiction and results of the cell assay used to assess the ability of Siglec-6 mutants to bind glycolipid liposomes using flow cytometry. **c.** Depiction and results of the ELISA approach used to assess the function of the W127A Siglec-6 Mutant. **d.** 3D structures of the Ig V-set domains of MAG (blue cartoons; PDB 2ZG3), Siglec-6 (cyan cartoons; this work) and Siglec-11 (purple cartoons; PDB AF_AFQ96RL6F1). The residues in the membrane-facing loops that could potentially interact with the bilayer are labelled and highlighted with sticks. **e.** Sequence alignment of all human Siglecs performed with Clustal Omega(*25*) (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) shows that only Siglec-6 has a KW (or similar) combination of residues on the same loop, yet a number of other Siglecs, namely 5, 8, 10, 11, 14 and 16, have a combination of one aromatic and one positively charged residues across the two loops facing the cell membrane when binding gangliosides.

To assess whether W127A Siglec-6 perturbs the structure of Siglec-6, we tested the ability of W127A Siglec-6 to bind to glycolipids outside a bilayer in an ELISA, using conditions that we have previously optimized[26,27], see **Fig. 3.c**. We observed no significant difference in binding between WT and W127A Siglec-6 to a neoglycolipid ligand developed previously[14] to engage Siglec-6. Moreover, none of the full length Siglec-6 mutants showed significantly perturbed expression levels. These results strongly suggest that W127A Siglec-6 is functional, and that the loss of binding for W127A Siglec-6 observed in the cell assay is not due to structural changes in the protein due to the mutation.

5.3.5 Siglec-6 binding to GM1-3 as free oligosaccharides.

The characterisation of the recognition and binding mechanism of gangliosides by Siglec-6 was extended to include native MS (nMS) binding assays with GM1, GM2, and GM3 as free oligosaccharides, i.e. GM1os, GM2os and GM2os. Results from the Concentration Independent (COIN)-Catch-and-Release (CaR)-nMS assay(*28*) are shown in **Fig. 4.a** for Siglec-6 WT and for the W127A mutant, and in **Fig. 4.b** for the R122A mutant. Unlike the case of the membrane-linked gangliosides, Siglec-6 binds the monosialylated gangliosides GM1-3os with similarly low affinity. Binding of the free oligosaccharides is largely unaffected by the mutation of Trp127 (W127A), further suggesting that this mutation does not significantly perturb the structure of Siglec-6 unlike binding to glycolipids in a membrane. Binding is significantly weakened by the loss of the canonical Arg122. Taken together these results suggest that, unlike the case of membrane-linked gangliosides, binding of GM1-3os occurs primarily through the canonical Arg122. As an interesting point, the complex of CD33 with a sialoside analogue (PDB 7AW6), shows an alternative conformational arrangement of the glycan epitope relative to the conformation required for a membrane-bound ganglioside. Indeed, in this complex the reducing end (GM1-equivalent) Gal-β(1-4)-[2-aminoethyl]-Glc moiety is directed towards the C-C' loop of CD33, rather than in the opposite direction, see **Fig. 4.h**.

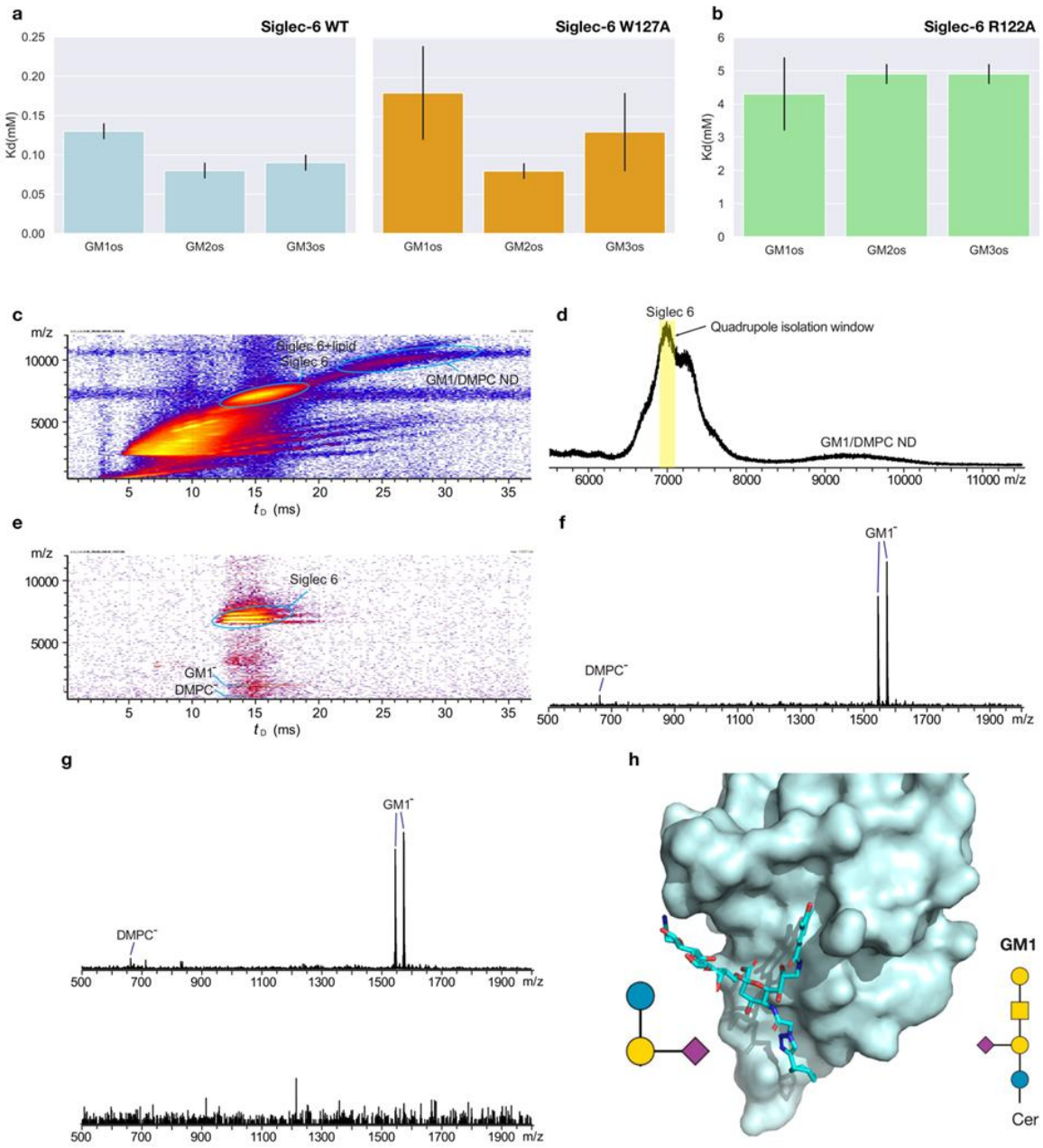


Figure 4. **a.** Binding affinities (*K*d, mM) of the ganglioside oligosaccharides GM1os, GM2os and GM3os for Siglec-6 (Fc) WT (light blue bars) and W127A mutant (orange bars) measured in aqueous ammonium acetate (200 mM, pH 6.8, 25 ºC) by COIN-CaR-nMS. Bar plots with *seaborn* (https://seaborn.pydata.org/). **b.** Binding affinities (*K*d, mM) of the ganglioside oligosaccharides GM1os, GM2os and GM3os for Siglec-6 (Fc) R122A mutant (green bars) measured in aqueous ammonium acetate (200 mM, pH 6.8, 25 ºC) by COIN-nMS. **c – g** IMS-CaR-nMS measurements performed in negative ion mode for 0.8 µM Siglec 6 (Fc) WT with 5 µM 10% GM1/DMPC ND in 200 mM aqueous ammonium acetate solution (pH 7.4). **c.** IMS heat map (m/z versus IMS drift time t_D) and **d.** corresponding full mass spectrum extracted from IMS heat map using Driftscope. **e.** IMS heat map and **f.** corresponding extracted CID mass spectrum showing (glyco)lipid ions released from Siglec-6; ions with m/z 7,000±100 were isolated by quadrupole (isolation window highlighted in yellow in **e**, followed by IMS and CID; a collision energy of 100 V was applied in the Transfer region. **g.** IMS-CaR-nMS measurements for Siglec-6 (Fc) mutants (each at 0.8 µM) and 10% GM1/DMPC ND (5 µM) in 200 mM aqueous ammonium acetate solutions (pH 7.4); CID mass spectra (extracted form IMS heat maps) showing (glyco)lipid ions released from (top) Siglec-6 R122A and (bottom) Siglec-6 W127A. **h.** 3D structure of the V-set domain of CD33 in complex with a sialoside analogue (PDB 7AW6). The protein is rendered in cyan as a surface, while the sialoside is rendered with sticks with C atoms in cyan, O in red and N in blue. SNFG symbols of the sialoside (left) and of the GM1 (right) are shown to indicate the position of the glucose at the reducing end.

5.3.6 Siglec-6 binding to GM1-enriched lipid nanodiscs.

We used Catch-and-Release nMS performed with Ion Mobility Separation (IMS-CaR-nMS) to assess the ability of Siglec-6 to bind phospholipids in addition to gangliosides. To this end, we performed experiments with Siglec-6-FcWT, and R122A and W127A mutants with a GM1 nanodisc (ND) prepared with 10% GM1 and DMPC. Shown in Fig. 4.c and 4.d are: representative IMS heat maps (m/z versus IMS drift time tD) and mass spectrum measured in negative ion mode for 0.8 μM Siglec 6 WT and 5 μM 10% GM1/DMPC ND (corresponding to 100 μM GM1 and ~900 μM DMPC), respectively. Both Siglec-6 and ND appeared as broad peaks and centred at m/z 7,000 (tD = 12.5 to 19.0 ms) and 9,500 (tD = 22 to 32 ms), respectively. Due to the broad m/z distributions of both Siglec-6 and ND, the free and lipid-bound Siglec-6 ions and those of the ND presumably overlapped in the mass spectrum. To address this potential issue, we used quadrupole isolation (m/z 7,000 ± 100) along with IMS to separate the Siglec-6 (and Siglec-6−lipid complexes) ions from the ND ions. Collision-induced dissociation (CID) was performed in the Transfer region (100 V collision energy) to release lipid ions exclusively bound to the protein. The IMS heat map of the IMS-CaR-nMS assay and the CID mass spectrum are shown in Fig. 4.e and 4.f, respectively. The CID shows the release of deprotonated GM1 ions, with deprotonated DMPC ions also detected at lower abundance. Taken together the IMS-CaR-nMS results support the binding of Siglec-6 WT to both GM1 and to phospholipids. We performed analogous binding assays on the Siglec-6 R122A and W127A mutants with GM1/DMPC ND. The CID mass spectra extracted from IMS are shown in **Fig. 4.g**. The results indicate that the Siglec-6 R122A mutant binds phospholipids analogously as the WT, meanwhile Siglec-6 W127A produces no detectable lipid ions, which confirms that the loss of Trp 127 compromises lipid binding.

5.3.7 Siglec-6 binds liposomes and nanodisks devoid of GM1.

As the MD simulations and the mutagenesis studies demonstrated, Trp127 interacts with the bilayer and is critical for the binding of membrane-bound glycolipids by Siglec-6. Therefore, we hypothesized that there is direct binding between the naked liposomes and Siglec-6 and used the cell-based assay to assess if this was the case. Indeed, WT Siglec-6+ cells showed significantly higher binding to naked liposomes compared to Siglec-6-CHO cells, see **Fig. 5.a**. Moreover, the binding of naked liposomes was greatly decreased (~5-fold) to W127A Siglec-6 CHO cells. In line with the results from **Fig. 3.b**, naked

liposome binding was not reduced to background levels with W127A Siglec-6, suggesting that while Trp127 is the major contributor to the bilayer binding, other residues such as K126A also support the interaction. In support of the results obtained on naked liposomes, we also performed IMS-CaR-nMS measurements to test lipid binding of Siglec-6 WT and mutants on 'empty' DMPC ND (5 μM), see **Fig. 5.b**. The CID mass spectra show that Siglec-6 WT binds phospholipids and that the R122A mutant retains lipid-binding. Results obtained for the Siglec-6 W127A mutant demonstrate that the loss of Trp127 compromises the interaction with the bilayer.

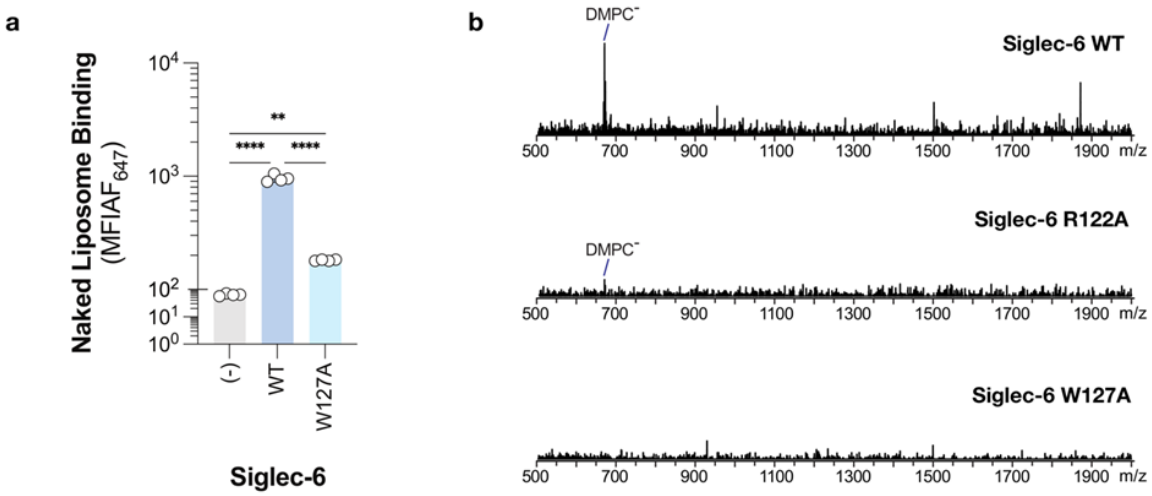


Figure 5 a. Binding of naked liposomes (liposomes that lack a ligand) to WT, W127A and Siglec-6- CHO cells using the cell assay. **b)** IMS-CaR-nMS measurements for Siglec 6 (Fc) WT and mutants (each at 0.8 μM) with DMPC ND (5 μM) in 200 mM aqueous ammonium acetate solutions (pH 7.4); **b.** CID mass spectra (extracted form IMS heat maps) showing lipid ions released from (top) Siglec 6 WT, (middle) Siglec 6 R122A and (bottom) Siglec 6 W127A.

5.4 Discussion

Characterising glycan-binding specificity across human Siglecs is a difficult task, hindered by their weak binding affinity, potential promiscuity, and structural plasticity facilitating both *cis* and *trans* ligand interactions. Identifying the molecular determinants regulating their glycan preference is even more complex, as it requires gathering statistically reproducible, atomistic-level information on the structure and dynamics of the Siglecs, integrated within a sufficiently detailed 3D model of their own biological environment. In this study, we show how all-atom classical MD simulations were used to guide experiments towards the characterisation of the unique binding mechanism and glycan-specificity of human Siglec-6, an inhibitory receptor expressed in mast cells, memory B-cells and, unique to humans, placental syncytiotrophoblasts.

Our results show that Siglec-6 evolved the ability of recognising and binding monosialylated gangliosides not only by engaging the glycan epitope poking out of the bilayer, but also by supplementing its binding affinity through interactions with the surrounding phospholipids. As suggested by MD simulations and supported by liposome and ND binding assays, the interactions between the Siglec-6 and the cell membrane hinge on the insertion of Trp127 into the bilayer, assisted by the adjacent Lys126, which interacts with the phospholipid heads. Different examples of proteins anchoring phospholipid bilayers through exposed aromatic and charged amino acids have been described within various biological contexts, e.g. where such interactions are instrumental for the orientation of transmembrane proteins[30-33] within the bilayer, as well as for the interaction of proteins and peptides with the membrane[34-36] and for the binding of membrane-embedded ligands[35,37-39]. To our knowledge, this is the first instance in which binding to the cell membrane has been characterised as a key step in the ligand-recognition of a carbohydrate-binding protein.

Further to this, we find that the structure of the ganglioside epitope is a determinant for Siglec-6 recognition and binding, with GM1 being the only monosialylated lipid-linked ganglioside recognised within a phospholipid bilayer. As in other Siglecs[1], Siglec-6 binding specificity is controlled by the C-C' loop, which interacts with the terminal Gal in GM1. GM2 and GM3 lack this terminal Gal and Gal-β(1-3)-GalNAc motif, respectively. In a model of the Siglec-6/GM2 complex that we studied by two independent MD simulations, we see that the C-C' loop is conformationally dynamic, enabled by the absence of the terminal Gal, triggers the disruption of the contact with the membrane and ultimately promoting the dissociation of the complex. Mutagenesis and binding assays on GM1-enriched liposomes and ND confirm that bilayer interactions mediated by Siglec-6 requires Trp127, with the support of the adjacent Lys126. As demonstrated through binding assays on 'empty' liposomes and ND, the interaction with the bilayer persists even in the absence of gangliosides. These results inform an interesting hypothesis about the ganglioside recognition strategy, whereby Siglec-6 scans the cell membrane surface through low affinity interactions with lipids, in search for higher affinity and specificity binding epitopes, such as GM1. The MD simulations of isolated GM1 and GM3 gangliosides in a model bilayer suggest that such a strategy may be the most efficient process. Indeed, while we measured that the sialic acid in both GM1 and GM3 remains accessible through the lipids, the size of the exposed epitope is rather small, making it unlikely to be able to engage the Siglec-6 V-set unless in very close

proximity. To this end, the evolutionary advantage of developing a mechanism that hinges on membrane interaction could rest in enhancing Siglec-6 selectivity for GM1, namely turning a sialic acid-binding lectin into a high-specificity molecular precision tool. The study of how such sophisticated scanning mechanism impacts epitope recognition and binding at the molecular level of detail would be very interesting, yet we believe to be beyond the scope of this work, and it may require the use of a coarse-grained MD approach, ideally in combination with super-resolution microscopy[47,48].

Siglec-6 binding has been shown to be independent of the canonical Arg[14], while MAG and Siglec-11 have been also reported to share such feature[15]. In agreement, our MD simulations show that the interaction between Siglec-6 and GM1 is only partially supported by the salt bridge between Arg122 and the Neu5Ac, with binding assays and mutagenesis in this and in earlier work[14] showing a reduction of the binding affinity in the R122A mutant. Sequence alignment shows that the KW motif implicated in the membrane interaction is a unique feature of Siglec-6, see **Fig. 3.e**. Yet, structural analysis supported by the insight we gathered from the MD simulations, show that contacts with the membrane could potentially involve two loops, see **Fig. 3.d.** Other Siglecs, namely 5, 8, 10, 11, 14, and 16, carry at least one aromatic and one positively charged residue (at physiological pH) on either loop. Although we cannot exclude that membrane interactions may occur in other Siglecs through this alternative 'two-loop' strategy, our collaborators showed through MD simulations that Siglec-10, which has a particularly extensive combination of 'ideal' membrane-binding residues across the two loops, namely Arg47 and Trp50 on one, and Arg122 and Tyr125 on the other, see **Fig. 3.e,** does not engage with the lipids in the bilayer[40]. This is not an entirely surprising result, as Siglec-10 does not need to bind the cell membrane because its binding specificity is for polysialylated epitopes in glycoproteins or on secreted glycans[40-43]. Siglec-11 does not need to bind the membrane either, as it does not bind gangliosides[14] but α(2-8) polysialic acids, which have a rigid architecture and terminate the arms of *N*-glycans[44]. Earlier work[14] indicates that MAG binds GM1 only weakly, but not GM2/3, and that it has a higher affinity of polysialylated gangliosides; this interaction appears to be dependent on its canonical Arg118. Ultimately, membrane binding appears, thus far, to be a feature unique to Siglec-6. Yet, the broad accessibility of the binding site in the terminal V-set domain and the conservation of the salt-bridging canonical Arg, indicates that Siglecs are able to bind free sialosides within a varying range of low binding affinities. Indeed, contrary to the case of membrane-bound gangliosides, here we have shown through nMS binding assays

that Siglec-6 binds GM1-3 as free oligosaccharides (GM1-3os) with similarly low binding affinities with a mechanism that is entirely dependent on the canonical Arg122. Based on the crystal structure of the highly homologous Siglec-3 in complex with a sialoside analogue (PDB 7AW6), the relative orientation of the bound glycan epitope could also change to complement different (less constrained) environmental conditions.

5.5 Conclusions

In this work we show through a combination of computational and experimental methods how the recognition and binding of membrane-bound gangliosides by Siglec-6 hinges on its interaction with the cell membrane. The insertion of a Trp127 indole sidechain within the bilayer, assisted by the electrostatic interaction between Lys126 and the phospholipid heads, complements the ligand binding energetics, making the canonical Arg122 dispensable. Furthermore, the interaction with the membrane persists in the absence of the ligands, suggesting its involvement in recognition. These results corroborate a unique and distinctive binding mechanism adopted by Siglec-6, supporting the evolution of many different human Siglecs as molecular precision tools for the recognition of specific sialosides in their distinct biological environment. Further studies in this area are currently in progress.

References

1. Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. Nat. Rev. Immunol. 7, 255–266 (2007).
2. Läubli, H. & Varki, A. Sialic acid-binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. Cell. Mol. Life Sci. 77, 593–605 (2020).
3. Macauley, M. S., Crocker, P. R. & Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. Nat. Rev. Immunol. 14, 653–666 (2014).
4. Pillai, S., Netravali, I. A., Cariappa, A. & Mattoo, H. Siglecs and immune regulation. Annu. Rev. Immunol. 30, 357–392 (2012).
5. Duan, S. & Paulson, J. C. Siglecs as immune cell checkpoints in disease. Annu. Rev. Immunol. 38, 365–395 (2020).
6. Jetani, H. et al. Siglec-6 is a novel target for CAR T-cell therapy in acute myeloid leukemia. Blood 138, 1830–1842 (2021).
7. Fry, T. J. et al. CD22-targeted CAR T cells induce remission in B-ALL that is naive or resistant to CD19-targeted CAR immunotherapy. Nat. Med. 24, 20–28 (2018).
8. Kenderian, S. S. et al. CD33-specific chimeric antigen receptor T cells exhibit potent preclinical activity against human acute myeloid leukemia. Leukemia 29, 1637–1647 (2015).
9. Angata, T. & Varki, A. Discovery, classification, evolution and diversity of Siglecs. Mol. Aspects Med. 90, 101117 (2023).
10. Bornhöfft, K. F., Goldammer, T., Rebl, A. & Galuska, S. P. Siglecs: A journey through the evolution of sialic acid-binding immunoglobulin-type lectins. Dev. Comp. Immunol. 86, 219–231 (2018).
11. 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 alpha 2,8-disialyl and branched alpha 2,6-sialyl residues. A comparison with Siglec-9. J. Biol. Chem. 277, 6324–6332 (2002).
12. Schnaar, R. L. Gangliosides as Siglec ligands. Glycoconj. J. 40, 159–167 (2023).
13. Gonzalez-Gil, A. & Schnaar, R. L. Siglec ligands. Cells 10, 1260 (2021).
14. Schmidt, E. N. et al. Siglec-6 mediates the uptake of extracellular vesicles through a noncanonical glycolipid binding pocket. Nat. Commun. 14, 2327 (2023).
15. Varki, A. & Angata, T. Siglecs--the major subfamily of I-type lectins. Glycobiology 16, 1R–27R (2006).
16. Vinson, M. et al. Myelin-associated glycoprotein interacts with ganglioside GT1b. A mechanism for neurite outgrowth inhibition. J. Biol. Chem. 276, 20280–20285 (2001).
17. Angata, T. et al. Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. J. Biol. Chem. 277, 24466–24474 (2002).
18. Tang, S. et al. Myelin-associated glycoprotein interacts with neurons via a sialic acid binding site at ARG118 and a distinct neurite inhibition site. J. Cell Biol. 138, 1355–1366 (1997).

19. Gu, R.-X., Ingólfsson, H. I., de Vries, A. H., Marrink, S. J. & Tieleman, D. P. Ganglioside-lipid and ganglioside-protein interactions revealed by coarse-grained and atomistic molecular dynamics simulations. J. Phys. Chem. B 121, 3262–3275 (2017).
20. Manna, M., Róg, T. & Vattulainen, I. The challenges of understanding glycolipid functions: An open outlook based on molecular simulations. Biochim. Biophys. Acta 1841, 1130–1145 (2014).
21. Shorthouse, D., Hedger, G., Koldsø, H. & Sansom, M. S. P. Molecular simulations of glycolipids: Towards mammalian cell membrane models. Biochimie 120, 105–109 (2016).
22. Enkavi, G., Javanainen, M., Kulig, W., Róg, T. & Vattulainen, I. Multiscale simulations of biological membranes: The challenge to understand biological phenomena in a living substance. Chem. Rev. 119, 5607–5774 (2019).
23. Noid, W. G., Szukalo, R. J., Kidder, K. M. & Lesniewski, M. C. Rigorous progress in coarse-graining. Annu. Rev. Phys. Chem. 75, 21–45 (2024).
24. Majumder, A., Sahrmann, P. G. & Voth, G. A. Bottom-up coarse-grained models of asymmetric membranes. J. Phys. Chem. B 129, 10333–10342 (2025).
25. Cino, E. A., Ramirez-Echemendia, D. P., Hu, S. & Tieleman, D. P. Analyzing lipid distributions and curvature in molecular dynamics simulations of complex membranes. Methods Enzymol. 701, 579–601 (2024).
26. Bui, D. T. et al. Absolute affinities from quantitative shotgun glycomics using Concentration-Independent (COIN) native mass spectrometry. ACS Cent. Sci. 9, 1374–1387 (2023).
27. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–8, 27–8 (1996).
28. Patel, D. S. et al. Influence of ganglioside GM1 concentration on lipid clustering and membrane properties and curvature. Biophys. J. 111, 1987–1999 (2016).
29. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature (2021) doi:10.1038/s41586-021-03819-2.
30. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2022).
31. Awoyemi, T. et al. Glycosylated Siglec-6 expression in syncytiotrophoblast-derived extracellular vesicles from preeclampsia placentas. Biochem. Biophys. Res. Commun. 533, 838–844 (2020).
32. Schmidt, E. N. et al. Dissecting the abilities of murine Siglecs to interact with gangliosides. J. Biol. Chem. 300, 107482 (2024).
33. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).
34. Rodrigues, E. et al. A versatile soluble siglec scaffold for sensitive and quantitative detection of glycan ligands. Nat. Commun. 11, 5091 (2020).
35. Schmidt, E. N., Jung, J. & Macauley, M. S. Flow cytometry-based detection of Siglec ligands. Methods Mol. Biol. 2657, 181–193 (2023).
36. Brinkman-Van der Linden, E. C. M. et al. Human-specific expression of Siglec-6 in the placenta. Glycobiology 17, 922–931 (2007).
37. Vostrikov, V. V., Daily, A. E., Greathouse, D. V. & Koeppe, R. E., 2nd. Charged or aromatic anchor residue dependence of transmembrane peptide tilt. J. Biol. Chem. 285, 31723–31730 (2010).
38. Sustich, S. J., Afrose, F., Greathouse, D. V. & Koeppe, R. E., 2nd. Influence of interfacial tryptophan residues on an arginine-flanked transmembrane helix. Biochim. Biophys. Acta Biomembr. 1862, 183134 (2020).
39. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale

for proteins at membrane interfaces. Nat. Struct. Biol. 3, 842–848 (1996).
40. Frazier, A. A. et al. Membrane orientation and position of the C2 domain from cPLA2 by site-directed spin labeling. Biochemistry 41, 6282–6292 (2002).
41. Christiaens, B. et al. Tryptophan fluorescence study of the interaction of penetratin peptides with model membranes: Penetratin: interaction with model membranes. Eur. J. Biochem. 269, 2918–2926 (2002).
42. Barkdull, A. P., Holcomb, M. & Forli, S. A quantitative analysis of ligand binding at the protein-lipid bilayer interface. Commun. Chem. 8, 89 (2025).
43. Suárez, T., Gallaher, W. R., Agirre, A., Goñi, F. M. & Nieva, J. L. Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion. J. Virol. 74, 8038–8047 (2000).
44. Zhang, W., Crocker, E., McLaughlin, S. & Smith, S. O. Binding of peptides with basic and aromatic residues to bilayer membranes: phenylalanine in the myristoylated alanine-rich C kinase substrate effector domain penetrates into the hydrophobic core of the bilayer. J. Biol. Chem. 278, 21459–21466 (2003).
45. Stahelin, R. V. et al. Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs. J. Biol. Chem. 277, 26379–26388 (2002).
46. Bazzi, M. D. & Nelsestuen, G. L. Association of protein kinase C with phospholipid monolayers: two-stage irreversible binding. Biochemistry 27, 6776–6783 (1988).
47. Riera, R. et al. Single-molecule imaging of glycan-lectin interactions on cells with Glyco-PAINT. Nat. Chem. Biol. 17, 1281–1288 (2021).
48. Tholen, M. M. E., Riera, R., Izquierdo-Lozano, C. & Albertazzi, L. Multiplexed Lectin-PAINT super-resolution microscopy enables cell glycotyping. Commun. Biol. 8, 267 (2025).
49. Sobczak, K. et al. The unique molecular recognition features of Siglec-10: structural insights into sialoglycan and antibody interactions. bioRxiv (2025) doi:10.1101/2025.06.10.658867.
50. Boelaars, K. & van Kooyk, Y. Targeting myeloid cells for cancer immunotherapy: Siglec-7/9/10/15 and their ligands. Trends Cancer 10, 230–241 (2024).
51. Bandala-Sanchez, E. et al. T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10. Nat. Immunol. 14, 741–748 (2013).
52. DeBono, N. J. et al. The molecular basis of immunosuppression by soluble CD52 is defined by interactions of N-linked and O-linked glycans with HMGB1 box B. J. Biol. Chem. 301, 108350 (2025).
53. 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).
54. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. 29, 1859–1865 (2008).
55. Lee, J. et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J. Chem. Theory Comput. 12, 405–413 (2016).
56. Case, D. A. et al. AmberTools. J. Chem. Inf. Model. 63, 6183–6191 (2023).
57. Li, J. et al. Screening glycolipids against proteins in vitro using picodiscs and catch-and-release electrospray ionization-mass spectrometry. Anal. Chem. 88, 4742–4750 (2016).

Chapter 6. Effect of bisection on the structure, dynamics, maturation and recognition of *N*-glycans

S. D'Andrea*, A Satheesan*, E. Patterson, CA. Fogarty, A. Maggioni, M. von Itzstein, D. Kolarich and E.Fadda. *Effect of bisection on the 3D structure and recognition of N-glycans during maturation: Does MGAT3 introduce a stop codon?* (in preparation)

* These authors contributed equally to the work

6.1 Introduction

N-glycosylation is a co- and post-translational modification that takes place as nascent polypeptides enter and progress through the secretory pathway. In the endoplasmic reticulum (ER), a pre-assembled $(Glc)_3Man_9GlcNAc_2$ oligosaccharide is transferred from a dolichol pyrophosphate (Dol-PP) lipid donor onto selected asparagine residues within Asn-X(any aa but Pro)-Ser/Thr sequons, and more rarely within non-canonical Asn-X-C/V motifs[1–3]. These immature N-glycans participate in protein folding, quality control, and complex assembly in the ER, while subsequent processing in the Golgi apparatus generates mature glycoforms that support appropriate trafficking, surface expression, and secretion[4,5]. The remodeling of N-glycans is driven by the combined action of glycosidases and glycosyltransferases that trim oligomannose structures and then elaborate the conserved pentasaccharide core into hybrid and complex architectures.

Because N-glycosylation is not directly encoded by a nucleic-acid template, the N-glycan repertoire on any given glycoprotein is inherently heterogeneous. The distribution of structures depends on cell type and tissue[6], on the expression and localisation of branching and capping enzymes[78], and on the local accessibility of each glycosylation site on the protein surface. As glycoproteins move along the secretory pathway, N-glycans at solvent-exposed sites are more likely to undergo extensive processing, whereas sterically shielded sites tend to retain oligomannose or less elaborated forms. The resulting microheterogeneity is therefore both regulated and stochastic[9], and encodes functionally important information in a cell-type-specific manner[8,10].

Complex N-glycans are the most elaborated products of this pathway. In mammalian cells they arise from a common $Man_5GlcNAc_2$ intermediate, in which N-acetylglucosaminyltransferase I (GlcNAc-T1, MGAT1) introduces a β2-linked GlcNAc onto the α3-linked mannose, committing the glycan to the hybrid/complex route[11] (**Figure 6.1**). Trimming by α-mannosidases on the α6 arm, followed by the addition of a second β2-GlcNAc by N-acetylglucosaminyltransferase II (GlcNAc-T2, MGAT2), yields the biantennary (F)A2 scaffold. Because GlcNAc-T1 is the entry enzyme for this branch of the pathway, its inhibition suppresses further elaboration and favours oligomannose-type profiles[9]. Higher levels of branching are introduced when N-acetylglucosaminyltransferase IV (GlcNAc-T4, MGAT4) and N-acetylglucosaminyltransferase V (GlcNAc-T5, MGAT5) add additional GlcNAc residues to the α3 and α6 arms, giving rise to tri- and tetra-antennary structures[12]. These outer arms are then extended and capped[11] by β4-galactosyltransferase (β4Gal-T1)[13], linkage- and substrate-specific sialyltransferases such as ST3/6GalTs and ST3/6GalNAcTs[14], and by fucosyltransferases including FUT8, which introduces core α6-fucose, as well as α3-fucose on antennae to form Lewis X-type termini[14,15]. Together, this set of reactions generates families of multi-antennary glycans with distinct patterns of galactosylation, sialylation, and fucosylation.

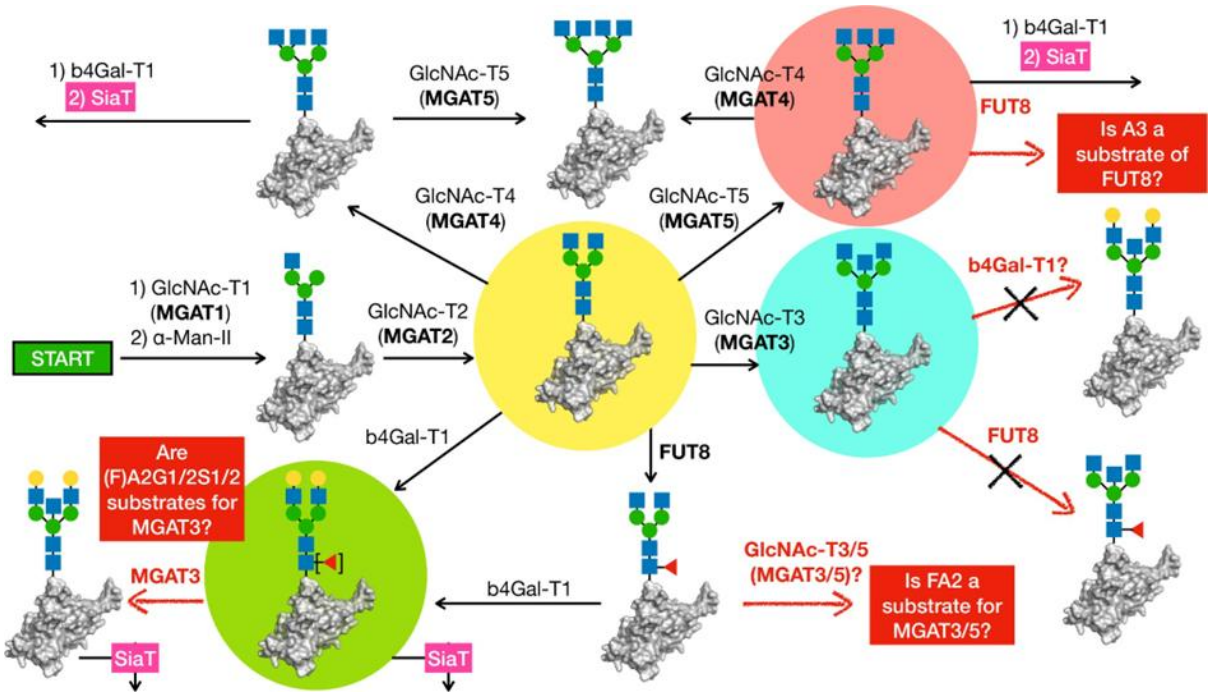


Figure 6.1 Schematic representation of the possible pathways leading to the maturation of a complex N-glycan. The biantennary A2 N-glycan is highlighted in yellow as a central node in the maturation. The bisected A2B system obtained as a product of MGAT3 on A2 is highlighted in cyan and tested in this work as substrates for galactosylation via b4GalT1 and for core fucosylation via FUT8. The triantennary A3 system with two antennae on the α6 arm is

highlighted in pink and tested for core fucosylation by FUT8. Galactosylated (F)A2G1/2 N-glycans as potential substrates of MGAT3 are highlighted in green. Questions that remain to be addressed are shown in red boxes, and the corresponding pathways we are attempting to clarify are indicated by red arrows. Known pathways are shown with black arrows. Information used to build this scheme is based on the following references(Stanley et al. 201; Schjoldager et al. 2020; Nakano et al. 2019).

The ensemble of N-glycans displayed by a cell is therefore best viewed as part of a broader "glycocode"[16]: a collection of structural motifs that modulate cell–cell and cell–environment communication. Within this framework, N-acetylglucosaminyltransferase III (GnT-III or MGAT3) plays a special role. MGAT3 catalyses the transfer of a β4-linked GlcNAc onto the central mannose of the conserved core, generating the so-called bisecting GlcNAc (**Figure 6.1**). Although bisected N-glycans are relatively rare at the whole-organism level, they can be strongly enriched in specific tissues[16,17] and have been linked to diverse pathological conditions[18], including cancer progression[19]. This restricted distribution, together with their association with disease[20], has led to the idea that bisection may act as a regulatory stop codon within the glycocode, dampening signalling processes that typically rely on highly branched, heavily functionalised multi-antennary glycans[6]. Biochemical and genetic studies suggest that MGAT3 acts after GlcNAc-T1 on the α3 arm but before the main branching enzymes GlcNAc-T2, -T4, and -T5[17]. As a consequence, hybrid and biantennary substrates are thought to be the primary acceptors for bisection (**Figure 6.1**). In some contexts, such as IgG or tumour-associated glycans, bisected structures can still carry extended and sialylated α3 and α6 antennae, but more commonly bisection correlates with reduced arm elongation and branching[21] and with diminished core fucosylation, likely via interference with FUT8 binding and catalysis[22]. A decrease in galectin binding at the cell surface[32] further supports the view that bisection limits the formation or persistence of outer-arm motifs that participate in lectin-mediated signalling networks. Despite this, the detailed structural and biosynthetic consequences of introducing a bisecting GlcNAc, and the precise substrate preferences of MGAT3, remain only partially understood. One major difficulty is analytical: bisected and triantennary N-glycans are isobaric, so distinguishing between them often requires lectin enrichment, GC–MS, or multi-stage mass spectrometry.

In this chapter, I use extensive all-atom molecular dynamics simulations, combined with multiple independent replicas, to clarify how the bisecting GlcNAc reshapes N-glycan architecture and influences recognition by human β4Gal-T1 and FUT8. I compare bisected N-glycans with fully functionalised triantennary glycans, in which two antennae

occupy the α6 arm, as well as with biantennary reference structures characterised previously. This allows me to pinpoint the steps in the maturation pathway at which bisection is most likely to interfere with antenna elongation and capping. In collaboration with Prof. Daniel Kolarich and Dr. Andrea Maggioni in Prof. Mark von Itzstein's group at the Institute for Glycomics, Griffith University, the binding of a soluble MGAT3 construct to mature complex N-glycans is also being characterised experimentally. Together, these computational and experimental approaches aim to define when and how bisection is introduced into fucosylated, galactosylated, and sialylated N-glycans, features that are common in IgG, and to elucidate the biosynthetic routes that give rise to these more elaborated, yet comparatively less abundant, multi-antennary structures.

6.2 Computational Methods

The starting structures of the bisected and triantennary *N*-glycans were generated using the carbohydrate builder tool on GLYCAM-WEB[23](https://glycam.org/). For each *N*-glycan, 9 structures were obtained using different combinations of α(1–6) torsion angles. The topology files for each structure were generated using the *tleap* program included in the AMBER18[24] software package. The carbohydrate atoms were represented using the GLYCAM06j-1 version of the GLYCAM06 force field[25], while the counterions in the simulations were represented using the AMBER ff14SB[26]. Water molecules were represented using the TIP3P water model. The MD simulations of the *N*-glycan structures were run at 200 mM NaCl. All simulations were run following an identical protocol, starting with an energy minimisation stage through 500,000 steps of steepest descent. During minimisation, the positions of all heavy atoms were restrained with a potential weight of 5 kcal $mol^{-1}Å^{-2}$ except for water molecules and counterions. Following minimisation, the system was brought up to standard temperature through a heating phase in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During the first stage, the temperature was raised from 0 to 100 K, with the volume held constant; during the second stage, from 100 to 300 K, with the pressure held constant. The system was then equilibrated for 1 ns in the NPT ensemble at a 1 atm pressure with the Berendsen barostat. The same restraints used for the minimisation stage were kept for these equilibrations. Conformational equilibration was carried out for 10 ns without restraints. The production phase followed, and analysis was carried out on the 500 ns trajectories for each starting structure. The torsion angles and

RMSD values were calculated using VMD[27]. A density-based clustering method, DBSCAN, was used to identify the conformations occupied by each *N*-glycan, based on torsion angles calculated from all trajectories for that *N*-glycan. The most populated structures of A2B and A3 were structurally aligned with the β4GalT1 structure co-crystallised with the tetrasaccharide from lacto-N-neohexose, PDB 4EE4[28]. To generate the fit, the terminal β2-GlcNAc of arms A2B and A3 were superimposed onto the terminal β6-GlcNAc of the co-crystallised Lacto-N-neohexose. To assess the potential recognition of the bisected and triantennary N-glycans by the FUT8 using as a template the protein co-crystallised with a biantennary structure PDB 6VLD[29]. In this case, the A2B and A3 glycans structures were superimposed onto the biantennary structure, using the shared chitobiose core for alignment. All structural alignments were performed with PyMOL (www.pymol.org). All non-clashing complexes obtained by aligning A2B and A3 with b4GalT1 and FUT8 were tested for stability using MD, with simulations run using the same protocol described above for the free glycans. We used the AMBER ff14SB force field to represent protein atoms. While I ran the simulations of the FUT8 system, my colleague, Dr Akash Satheesan, worked on the simulations of the β4GalT1 systems. Production runs for each N-glycan-protein complex were run up to 1 μs.

6.3 Results

6.3.1 Biantennary *N*-glycans

Previous work from our group30 used extensive molecular dynamics simulations to show that biantennary N-glycans do not share a single preferred conformation, but instead display sequence-dependent three-dimensional preferences. In that study, the biantennary (F)A2 glycan, in which both antennae terminate in β2-linked GlcNAc and which I treat here as a key branching scaffold for complex N-glycan maturation, was found to be comparatively rigid. Most of its residual flexibility is concentrated in the α6 arm, which interconverts between two main conformational states30. These states were described as open and closed, with the open state characterised by ($\varphi \approx 70°$, $\psi \approx 180°$) and the closed state by ($\varphi \approx 70°$, $\psi \approx 80°$); in both cases the ω torsion is sampled predominantly around 50° (Harbison et al., 2019). In the open state, the α6 arm extends away from the chitobiose core and is readily accessible to glycosyltransferases for further elaboration, whereas in the closed state it folds back towards the core, stacking against the chitobiose disaccharide and, when present, contacting the core fucose. For the (F)A2 scaffold, the two states are

populated to a similar extent, with the open conformation only slightly favoured over the closed one (about 55% versus 45% over the simulation time). Adding galactose to the α6 arm, to form (F)A2G or (F)A2G$_2$, shifts this equilibrium so that the closed state becomes dominant (~75% population), and terminal capping with sialic acid further enhances the preference for the closed conformer[30].

6.3.2 Bisected and triantennary *N*-glycans

The trajectories of the bisected (A2B) and triantennary (A3) *N*-glycans from MD simulations were analysed using VMD, and the torsion angles associated with their linkages were clustered using DBSCAN. The populations of the different torsion angles for the α6 arm in the A2B system, as well as the torsional values for both the α6 arm and the β6 torsion angles of the third branch in the A3 system, are reported in **Table 4.1**.

A2B	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
GlcNAc-β2-Man-	71.8(9.4) /167.9(9.3)/ 54(9.4)/**72**	74.2(13.1)/82.6(15.2)/ 44(6.8)/**28**		
Gal-β4-GlcNAc-β2-Man-	80.3(14.7)/89.2(14.2)/ 46(7.1)/**50**	71.9(8.6)/167.1(9.2)/ 54.6(8.9)/**41**	73.7(6.2)/65.9(7.0)/ 166.2(5.6)/**9** .	
Sia-α3-Gal-β4-GlcNAc-β2-Man-	77.6(15.1)/85.6(15.0)/ 45.2(7.6)/**70**	71.6(8.3)/167.0(8.7)/ 54.4(8.6)/**30**		
A3				
GlcNAc-β2-[GlcNAc-β6]-Man-α6-Man	70.0(9.0)/-172.0(12.9) 55.2(9.5)/**49**	72.3(11.0)/89.3(13.9)/ 48.5(9.1)/**44**	77.7(4.9)/-73.9(6.0)/ -139.0(6.1)/**4**	62.4(5.2)/-177.0(5.9) 179.0(5.1)/**3**
Sia-α6-Gal-β4-GlcNAc-β2-[Sia-α6-Gal-β4-GlcNAc-β6]-Man-α6-Man	73.3(11.9)/89.3(16.9)/ 48.3(11.0)/**47**	68.3(9.51)/-175(13.8)/ 53.3(10.2)/**36**	68.3(11.6)/61.6(11.5)/ 170(8.28)/**12**	90(6.26)/-48.1(7.68)/- 145(6.93)/**3**
Sia-α3-Gal-β4-GlcNAc-β2-[Sia-α3-Gal-β4-GlcNAc-β6]- Man-α6-Man	71.9(12.1)/87.2(16.6)/ 48.3(9.5)/**46**	69.4(8.2)/-173(12.2)/ 54.5(8.6)/**23**	67.8(6.32)/-166(10.7)/ -65.2(8.7)/**15**	78(12.1)/-67.9(8.5)/ -143.0(8.4)/**15**
GlcNAc-β6-Man-	-78.7(12.6)/- 178.8(18.7)/**52**	-75.0(10.3)/-170.6(15 .9)/-168.0(11.0)/**42**	-81.0(7.9)/-89.0(7.9)/- 160.5(9.3)/**6**	

Gal-β4-GlcNAc-β6-Man-	-81.2(11.6)/171.4(14.4)/53.1(8.9)/**56**	-74.68(9.35)/-173.15(15.08)/-168.72(9.82)/**32**	-72.23(8.21)/-85.42(9.77)/-157.94(9.14)/**12**	
Sia-α6-Gal-β4-GlcNAc-β6-Man-	-71.0(9.4)/-163.8(10.8)/53.1(9.9)/**68**	-66.9(6.5)/-145.5(9.6)/-161.3(7.3)/**17**	-68.8(6.5)/175.3(8.6)/-169.2(6.9)/**15**	
Sia-α3-Gal-β4-GlcNAc-β6-Man-	-74.0(10.2)/-174.7(17.0)/-166.8(12.4)/**49**	-70(8.4)/-87.6(10.6)/-161.9(10.5)/**24**	-76.3(9.3)/-179.2(18.6)/54.4(8.8)/**19**	-87.9(7.1)/-84.0(7.26)/-88.1(8.0)/**5**

Table 4.1. Conformations of the α6 arm and relative populations obtained from clustering analysis of the MD results in function of the *N*-glycan architecture, namely bisected (A2B) or triantennary (A3), and sequence. The values of the β6 torsion angle in the A3 system, regulating the orientation of the third branch, are also shown. For each cluster, the values (in degrees) of the torsion angles φ (O1C1OxCx)(sd)/ψ (C1OxCxCx+1)/ω (O6C6C5C4) correspond to the median calculated for each cluster with standard deviations shown in parenthesis and relative populations (%) in red. The α6 arm sequences are represented according to the SNFG convention (Neelamegham *et al*. 2019).

6.3.3 Effect of bisection on the *N*-glycan architecture

Introducing a bisecting GlcNAc leaves the geometry of the conserved N-glycan core and the conformational preferences of the α3 arm essentially unchanged when compared with the corresponding biantennary reference. Its main effect is instead on the behaviour of the α6 arm (Table 4.1). For the A2B glycan, in which both antennae terminate in β2-linked GlcNAc, the α6 arm samples two dominant conformational states, hereafter referred to as open and closed. The open basin (φ ≈ 71.8°, ψ ≈ 167.9°, ω ≈ 54.5°) corresponds to an extended α6 arm, whereas the closed basin (φ ≈ 74.2°, ψ ≈ 82.6°, ω ≈ 44.0°) brings the arm back towards the core (Figure 4.2). Compared with the (F)A2 scaffold, adding the bisecting GlcNAc redistributes the α6 conformational ensemble, increasing the population of the open state to ~72% over the course of the MD simulations. When both antennae are further elongated by introducing terminal galactose, the balance between the two states becomes more even, with the open and closed conformations accounting for 41% and 50% of the ensemble, respectively. Subsequent capping with sialic acid again biases the system, this time towards the closed conformation, whose population rises to ~70%.

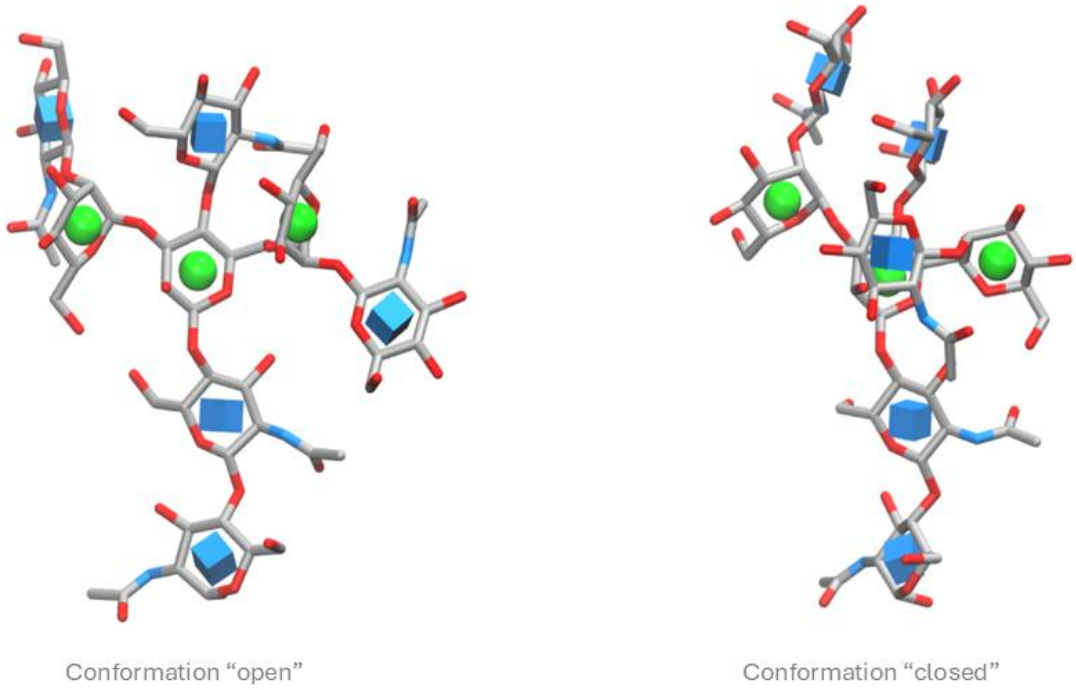


Figure 6.2. Representative structure of the “open” (left) and “closed” (right) conformation of the α6 arm in the bisected A2B *N*-glycan. Graphical rendering with VMD (Humphrey *et al*. 1996) (https://www.ks.uiuc.edu/Research/vmd/).

6.3.4 Effect of the third antenna on the *N*-glycan architecture

Compared with the corresponding biantennary and bisected glycans, triantennary structures display greater mobility of the α(1–6) arm and therefore explore a broader conformational space. For triantennary N-glycans in which all three antennae terminate in β(1–2)-linked GlcNAc, the α(1–6) arm populates four main conformational basins (**Table 4.1**). The open basin, centred around (φ = 70.0°, ψ = –172.0°, ω = 55.2°), is the most populated, accounting for 49% of the ensemble, followed closely by the front-fold basin at 44% (φ = 72.3°, ψ = 89.3°, ω = 48.5°). Two additional states are sampled less frequently: a back-fold conformation (φ = 77.7°, ψ = –73.9°, ω = –139.0°) with a population of 4%, and a second, more extended open-like basin (φ = 64.5°, ψ = –177.0°, ω = 179.0°) that represents 3% of the conformational ensemble. Elongation of the α(1–6) arm with terminal galactose only modestly perturbs this distribution. The open basin remains the most populated (46%), and the front-fold conformation remains the second most frequent (37%), while the back-fold and open-like basins increase slightly to 10% and 6%, respectively. A more pronounced redistribution occurs upon capping with sialic acid. For α(2–6)-linked and α(2–3)-linked sialylation, the front-fold conformation

becomes the dominant state (47% and 46%, respectively), whereas the population of the open basin decreases to 36% for α(2–6) and 23% for α(2–3) sialylation. The remaining conformational states persist but remain comparatively low in population. The β(1–6)-linked GlcNAc branching from the α(1–6) arm also exhibits several recurring conformational preferences. One of these, which we describe as a "claw-like" arrangement, is characterised by the β(1–6)-linked GlcNAc facing the β(1–2)-linked GlcNAc on the same arm. In the triantennary system with β(1–2)-GlcNAc-terminated antennae, this claw-like state dominates the ensemble, with an occupancy of 52% (φ = –78.7°, ψ = –178.8°, ω = 52.7°). A second, nearly as frequent conformation (42%; φ = –75.0°, ψ = –170.6°, ω = –168.0°) reorients the β(1–6)-linked GlcNAc towards the α(1–3) arm, while a third, less populated basin (6%; φ = –81.0°, ψ = –89.0°, ω = –160.5°) positions the β(1–6) arm behind the α(1–2)-linked GlcNAc. Adding β(1–4)-linked galactose to this branch does not radically reorganise the conformational preferences. The claw-like state remains the most populated (56%; φ = –81.2°, ψ = 171.4°, ω = 53.1°), followed by the α(1–3)-oriented conformation at 32% (φ = –74.7°, ψ = –173.2°, ω = –168.7°), while the third basin, in which the β(1–6)-linked GlcNAc lies behind the α(1–2)-linked GlcNAc, accounts for 12% (φ = –72.2°, ψ = –85.4°, ω = –157.9°). Terminal α(2–6) sialylation slightly enhances the preference for the claw-like arrangement, which increases to 68% (φ = –71.0°, ψ = –163.8°, ω = 53.1°). In contrast, α(2–3) sialylation shifts the equilibrium towards conformations in which the β(1–6)-linked GlcNAc points towards or behind the α(1–3) arm. Under these conditions, the α(1–3)-oriented state becomes the most populated (49%; φ = –74.0°, ψ = –174.7°, ω = –166.8°), the conformation with the β(1–6)-linked GlcNAc positioned behind the α(1–3)-linked GlcNAc increases to 24% (φ = –70.0°, ψ = –87.6°, ω = –161.9°), and the claw-like state drops to 19% (φ = –76.3°, ψ = –179.2°, ω = 54.4°).

6.3.5. Assessing recognition of A2B and A3 by FUT8

FUT8 (EC 2.4.1.68) is the Golgi α1,6-fucosyltransferase that catalyses the transfer of fucose from GDP-β-L-fucose to the innermost N-acetylglucosamine of complex and hybrid N-glycans, a reaction commonly referred to as core fucosylation. The enzyme operates as a homodimer, stabilised by a four-helix bundle formed by its N-terminal coiled-coil domains. This dimeric arrangement is essential to maintain the correct architecture of the SH3 domain, which engages the acceptor glycan and contributes to specific substrate recognition[31]. To investigate how a bisecting GlcNAc influences FUT8

recognition, I used the crystal structure PDB 6LVD, which provides the FUT8 dimer at 2.28 Å resolution in complex with GDP and an asialo-agalacto biantennary glycopeptide (A2SGP) acceptor substrate. Starting from this complex, I modelled FUT8 bound to the most populated A2B and A3 conformers identified in the MD simulations. The co-crystallised A2SGP glycopeptide served as a template: A2B and A3 were aligned to the FUT8 active site by superimposing their shared chitobiose core, ensuring a consistent placement of the N-glycan portion within the binding pocket. For A2B, the resulting structural overlays show that the bisecting GlcNAc creates steric clashes that prevent full accommodation of the glycan in the active site and thus are incompatible with productive recognition by FUT8 (**Figure 6.3**).

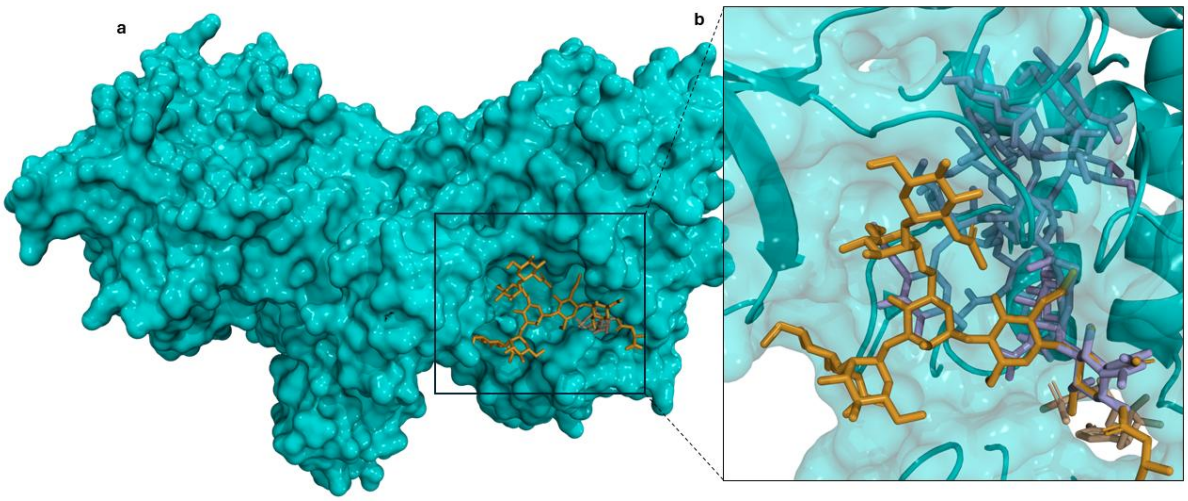


Figure 6.3.a. Crystal structure of FUT8 co-crystallized with A2SGP. FUT8 is rendered as surface (cyan) and the co-crystallized A2SGP is shown as sticks in orange). **b**. Close-up of the acceptor site: the crystallographic A2SGP (orange sticks) is overlaid with the bisecting N-glycan conformations sampled in MD (light-blue sticks), highlighting their steric incompatibility within FUT8. Renderings generated using PyMOL (www.pymol.org)

When the open A3 conformer was superposed onto the co-crystallised A2SGP in FUT8, the resulting model produced a plausible enzyme–substrate complex. I then subjected this complex to 1 µs of all-atom MD simulations to assess its stability (**Figure 6.4**). Throughout the trajectory, the A3 glycan remained stably accommodated in the FUT8 binding pocket, supported by persistent contacts with residues H505, R545, Y498, and D495 (**Figure 6.4**). The glycan backbone fluctuated around a mean RMSD of 4.57 Å with a standard deviation of 0.75 Å, indicating a well-maintained binding mode over the course of the simulation.

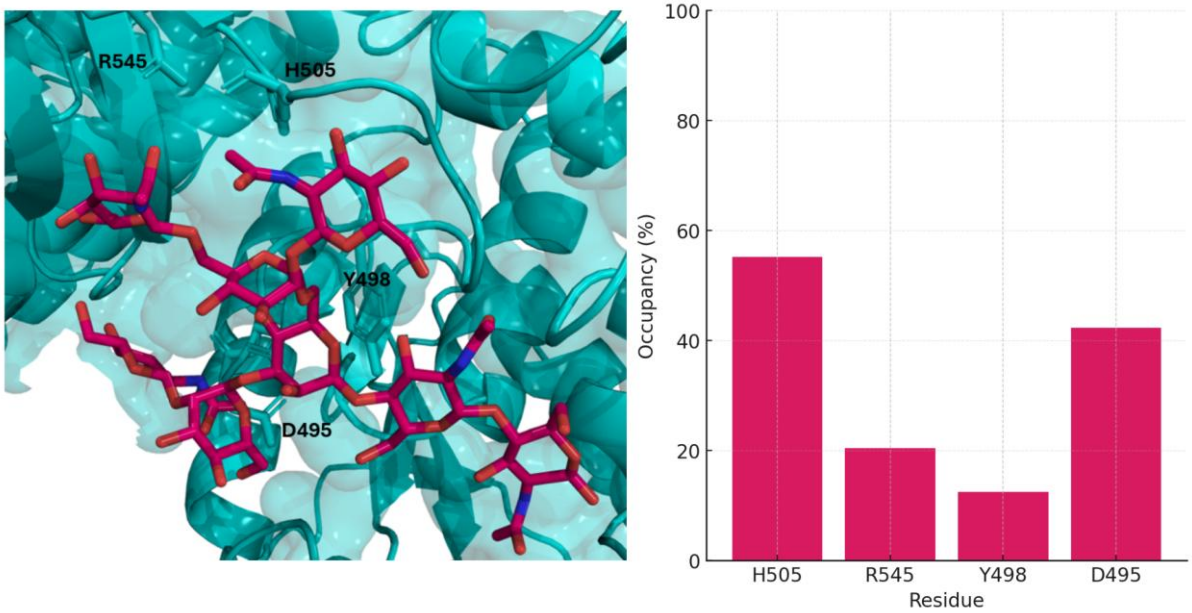


Figure 6.4. *Left.* Structural view of the FUT8 pocket with A3 shown as sticks in green. The interacting residues H505, R545, Y498, and D495 are labeled; the protein is shown as cartoon with a semi-transparent surface (cyan). *Right*. Bar chart of interaction occupancy (%) for these residues over the MD trajectory.

Additional analyses performed by my colleague Dr. Akash Sateesan, in which he assessed the suitability of bisected biantennary (A2B) and triantennary (A3) N-glycans as substrates for β4GalT1, are presented in the Appendices.

6.4 Discussion

Structural alignments and MD simulations reveal how different N-glycan architectures (biantennary, triantennary, and bisected) modulate recognition and maturation. Previous studies on biantennary structures showed that the α6 arm adopts a more closed conformation with the addition of terminal galactose and sialic acid. Here, I demonstrate that *N*-glycan bisection stabilises the open conformation of A2B structures, likely due to steric hindrance from the bisecting GlcNAc, which limits the α6 arm's rotational freedom and hinders its transition to a closed state. Adding terminal Gal and Neu5Ac shifts the equilibrium toward the closed state, indicating that arm elongation supports this conformation. Introducing a third antenna on the α6 arm to form the triantennary A3 structure increases conformational complexity. The α6 arm in A3 can adopt a wider range of states, including the 'front-fold' and 'back-fold' conformations seen in oligomannose structures[28,32]. The flexible β6 linkage of the third antenna samples a wide range of torsion angles, increasing the conformational diversity of the glycan; this diversity is further

enhanced when the antennae are elongated. Building on the detailed conformational analysis of bisected and triantennary glycans in solution, I next examined whether these structures can be recognised by the enzymes responsible for core and arm functionalisation, FUT8 and β4GalT1. Combining my data with that of my colleague, Akash Satheseen, we found that A2B does not act as a substrate for either FUT8 or β4GalT1: structural alignment of the most populated MD conformers onto the co-crystallised ligands leads to steric clashes that cannot be resolved without distorting the protein–glycan interface. This outcome is consistent with previous work reporting that A2B is incompatible with FUT8 and can inhibit FUT8 activity, thereby blocking core fucosylation[33]. In line with these observations, bisection of biantennary N-glycans has been shown to generally suppress arm elongation, which agrees with our results and prevents further galactosylation.

Although these findings explain why bisection hampers recognition of A2B by FUT8 and β4GalT1, bisected biantennary N-glycans bearing core fucose and extensively elaborated antennae are nevertheless detected in IgGs and tumour cells[34]. The bisecting GlcNAc is also present in core-fucosylated tetraantennary glycans carrying terminal SDa (Neu5Acα2-3[GalNAcβ1-4]Galβ1-4GlcNAc) epitopes, as described for pregnancy-associated glycoproteins in bovines. Together, these observations point to an alternative route for the generation of mature bisected N-glycans, potentially involving the action of MGAT3 on already functionalised scaffolds (see **Figure 6.1**). At present there are no PDB structures available for MGAT3, and the AlphaFold model (UniProt Q09327) does not provide sufficient structural detail for docking-based analyses. To tackle this limitation, we have established a collaboration with Dr Maggioni and Prof Kolarich at the Institute for Biomedicine and Glycomics, Griffith University, who have expressed a soluble human MGAT3 construct and are currently probing its activity by mass-spectrometry-based glycomic profiling. Our MD-based structural alignment further indicates that triantennary A3 structures are compatible with binding to both FUT8 and β4GalT1, suggesting that they can act as substrates for these enzymes. Previous studies have shown that knocking out β4GalT1 activity decreases galactosylation of triantennary *N*-glycans[35,36,37], supporting the possibility that triantennary structures could serve as substrates for β4GalT1. In the case of FUT8, earlier studies indicate higher activity toward triantennary structures synthesised by MGAT4, in which a β4-linked GlcNAc is added to the α3 mannose arm. However, triantennary structures synthesised by MGAT5, which have a β6-linked GlcNAc on the α6 mannose arm (the triantennary structure we tested), showed lower activity toward FUT8 than both biantennary and MGAT4-synthesised triantennary structures[38].

6.5 Conclusions

In this work, I used MD simulations to compare the architecture and dynamics of bisected and triantennary N-glycans with the corresponding biantennary scaffold. My results show that introducing a bisecting GlcNAc shifts the conformational equilibrium of the glycan, whereas adding a third antenna increases the overall conformational diversity. I then examined how these changes influence N-glycan maturation and substrate compatibility by using the equilibrium A2B and A3 structures in solution as putative recognition substrates for FUT8 and β4GalT1. By combining structural alignment of the equilibrium N-glycan structures with additional MD simulations of the viable enzyme–substrate complexes, I find that bisection prevents A2B from being recognised by either FUT8 or β4GalT1, while the triantennary A3 glycan remains a suitable substrate for both enzymes. Together, these findings provide a structural framework that supports and extends earlier experimental observations, and highlight N-glycan bisection as an effective suppressor of further N-glycan maturation.

References

1. Khaleque, M., Tropea, B., Singh, O., Elango, D., Liu, S., Allen, J. D., ... & Schulz, B. L. (2025). Regulation of N-glycosylation efficiency by eukaryotic oligosaccharyltransferase. *bioRxiv*, 2025-09.
2. Mohorko E, Glockshuber R, Aebi M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. J Inherit Metab Dis. 2011 Aug;34(4):869–78.
3. Wild R, Kowal J, Eyring J, Ngwa EM, Aebi M, Locher KP. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. Science. 2018 Feb 2;359(6375):545–50.
4. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. Nat Rev Mol Cell Biol. 2020 Dec;21(12):729–49.
5. Stanley P, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
6. Taniguchi N, Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. Adv Cancer Res. 2015 Feb 7;126:11–51.
7. Varki A. Essentials of Glycobiology. 1999. 680 p.
8. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol. 2019 Jun;15(6):346–66.
9. Hossler P, Mulukutla BC, Hu WS. Systems analysis of N-glycan processing in mammalian cells. PLoS One. 2007 Aug 8;2(8):e713.
10. Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology. 2012 Nov;22(11):1440–52.
11. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. Essentials of Glycobiology, Fourth Edition. 2022. 892 p.
12. Taniguchi N, Honke K, Fukuda M. Handbook of Glycosyltransferases and Related Genes. Springer Science & Business Media; 2011. 668 p.
13. Qasba PK, Ramakrishnan B, Boeggeman E. Structure and function of beta -1,4-galactosyltransferase. Curr Drug Targets. 2008 Apr;9(4):292–309.
14. Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P. The human sialyltransferase family. Biochimie. 2001 Aug;83(8):727–37.
15. Mondal N, Dykstra B, Lee J, Ashline DJ, Reinhold VN, Rossi DJ, et al. Distinct human α(1,3)-fucosyltransferases drive Lewis-X/sialyl Lewis-X assembly in human cells. J Biol Chem. 2018 May 11;293(19):7300–14.
16. Deciphering the glycocode: the complexity and analytical challenge of glycomics. Current Opinion in Chemical Biology. 2007 Jun 1;11(3):300–5.
17. Miwa HE, Song Y, Alvarez R, Cummings RD, Stanley P. The bisecting GlcNAc in cell growth control and tumor progression. Glycoconj J. 2012 Dec;29(8-9):609–18.
18. Chen Q, Tan Z, Guan F, Ren Y. The Essential Functions and Detection of Bisecting GlcNAc in Cell Biology. Front Chem. 2020 Jul 3;8:511.
19. Link-Lenczowski P, Bubka M, Balog CIA, Koeleman CAM, Butters TD, Wuhrer M, et al. The glycomic effect of N-acetylglucosaminyltransferase III overexpression in metastatic melanoma cells. GnT-III modifies highly branched N-glycans. Glycoconj J. 2018 Apr;35(2):217–31.
20. Nakano M, Mishra SK, Tokoro Y, Sato K, Nakajima K, Yamaguchi Y, et

al. Bisecting GlcNAc Is a General Suppressor of Terminal Modification of -glycan. Mol Cell Proteomics. 2019 Oct;18(10):2044–57.
21. Kizuka Y, Taniguchi N. Enzymes for N-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer. Biomolecules [Internet]. 2016 Apr 28;6(2). Available from: http://dx.doi.org/10.3390/biom6020025
22. Umaña P, Jean–Mairet J, Moudry R, Amstutz H, Bailey JE. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. Nature Biotechnology. 1999 Feb;17(2):176–80.
23. Website [Internet]. Available from: https://doi.org/10.1101/2025.05.08.652828
24. Lee TS, Cerutti DS, Mermelstein D, Lin C, LeGrand S, Giese TJ, et al. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. J Chem Inf Model. 2018 Oct 22;58(10):2043–50.
25. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008 Mar;29(4):622–55.
26. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.
27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14(1):33–8, 27–8.
28. Ramakrishnan B, Boeggeman E, Qasba PK. Binding of N-acetylglucosamine (GlcNAc) β1-6-branched oligosaccharide acceptors to β4-galactosyltransferase I reveals a new ligand binding mode. J Biol Chem. 2012 Aug 17;287(34):28666–74.
29. Järvå MA, Lingford JP, John A, Soler NM, Scott NE, Goddard-Borger ED. Trefoil factors share a lectin activity that defines their role in mucus. Nat Commun. 2020 May 13;11(1):2265.
30. Harbison AM, Brosnan LP, Fenlon K, Fadda E. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. Glycobiology. 2019 Jan 1;29(1):94–103.
31. Järvå MA, Dramicanin M, Lingford JP, Mao R, John A, Jarman KE, et al. Structural basis of substrate recognition and catalysis by fucosyltransferase 8. J Biol Chem. 2020 May 8;295(19):6677–88.
32. Fogarty CA, Fadda E. Oligomannose -Glycans 3D Architecture and Its Response to the FcγRIIIa Structural Landscape. J Phys Chem B. 2021 Mar 18;125(10):2607–16.
33. Bosman GP, Stoof ID, Bastiaansen HP, Quarles van Ufford L, Dobruchowska JM, Langenbach JWH, et al. Bisecting -Acetylglucosamine of the -Glycan of Immunoglobulin G Does Not Affect Binding to Fc Gamma Receptors. ACS Chem Biol. 2025 Mar 21;20(3):680–9.
34. Hodoniczky J, Zheng YZ, James DC. Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling in vitro. Biotechnol Prog. 2005 Nov-Dec;21(6):1644–52.
35. Bondt A, Rombouts Y, Selman MHJ, Hensbergen PJ, Reiding KR, Hazes JMW, et al. Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. Mol Cell Proteomics. 2014 Nov;13(11):3029–39.
36. Bydlinski N, Maresch D, Schmieder V, Klanert G, Strasser R, Borth N. The contributions of individual galactosyltransferases to protein specific N-glycan

processing in Chinese Hamster Ovary cells. J Biotechnol. 2018 Sep 20;282:101–10.
37. Zhao Y, Nayak S, Raidas S, Guo L, Della Gatta G, Koppolu S, et al. In-Depth Mass Spectrometry Analysis Reveals the Plasma Proteomic and N-Glycoproteomic Impact of an Amish-Enriched Cardioprotective Variant in B4GALT1. Mol Cell Proteomics. 2023 Aug;22(8):100595.
38. Tseng TH, Lin TW, Chen CY, Chen CH, Lin JL, Hsu TL, et al. Substrate Preference and Interplay of Fucosyltransferase 8 and N-Acetylglucosaminyltransferases. J Am Chem Soc. 2017 Jul 19;139(28):9431–4.

Chapter 7. My contribution to the GlycoShape Glycan 3D Structures Database

The work in this chapter is part of the following paper:

Ives, C. M., Singh, O., **D'Andrea**, S., Fogarty, C. A., Harbison, A. M., Satheesan, A., Tropea, B., Fadda, E. (2024). Restoring protein glycosylation with GlycoShape. *Nature Methods*, *21*(11), 2117-2127.

7.1 Introduction

The native fold of a protein determines its biological function, regulating mechanisms driving protein–protein and protein–ligand recognition, as well as binding and unbinding events in operational or physiological thermodynamic conditions[1]. Although the precise sequence and chemical nature of amino acids are the key determinants of correct folding, regulating structural stability and protein dynamics, this paradigm, guided by the genetic template, is complemented by post-translational modifications (PTMs)[2]. PTMs label proteins by covalently adding functional groups or more complex molecular entities, guiding the system toward specific functions or biological pathways[3].

Among PTMs, glycosylation is one of the most common, abundant, and mysterious. It refers to the enzymatically controlled functionalization of biomolecules with complex carbohydrates, or glycans. Glycosylation is an extremely flexible biological strategy, allowing structural and sequence variations "on the fly" that reflect environmental conditions and health or disease states in which proteins must operate[4]. The importance and ubiquity of glycosylation are underscored by the estimate that approximately 3–4% of the human genome is dedicated solely to encoding the mechanisms that regulate it. Glycosylation can occur co- or post-translationally. The two most common types in eukaryotes are: N-glycosylation, which involves the covalent addition of a glycan to the amide nitrogen (N) atom of specific Asn residues[4,5], and O-glycosylation, where the covalent bond occurs on the oxygen (O) atom of the hydroxyl group of Ser or Thr residues. Less common but highly conserved modifications include C-mannosylation, which involves C2 functionalization of the Trp indole side chain with mannose[6,7]. The structural complexity of glycans and their glycosylation patterns is unparalleled in nature,

with sequences ranging from a single monosaccharide to hundreds of units, arranged in linear and branched structures[8,9].

Glycosylation performs multiple biological functions across different systems and organisms, including protein folding, stability, cell adhesion, and signaling. Alterations in glycosylation patterns can be a useful diagnostic tool for various diseases[10]. For instance, increased sialylation of glycoconjugates, resulting from the upregulation of sialyltransferases, has been linked to cancer. Increased expression levels of fucosyltransferases, like FUT8 (which catalyzes the transfer of a fucose to the core of an N-glycan), and specific GlcNAc transferases, like MGAT3 (which transfers GlcNAc to the central mannose of the N-glycan pentasaccharide scaffold), as is shown in **Chapter 6**, have been linked to cancer progression.

Unlike proteins, glycan structures are not directly encoded in the genome. Glycan biosynthesis is highly varied and dynamic, resulting in the typical heterogeneity of glycoconjugates: microheterogeneity refers to the variety of glycan structures present at the same glycosylation sites, while macro-heterogeneity refers to the different degrees of site occupancy[11–14]. This diversity is further increased because glycosylation is sensitive to environmental conditions. Structurally, glycosidic linkages give glycans significant flexibility, making experimental characterization of their 3D structures challenging, even under cryogenic conditions. Glycans are often removed to facilitate protein crystallization[15]. The difficulty in resolving glycan 3D structures significantly hinders understanding their functions in various environments. Limited information on the three-dimensional (3D) structure, occupancy, and identity of glycans at different sites greatly contributes to our limited understanding of the multiple functions glycans perform. To address this challenge and bridge the gap left by traditional experimental techniques, the integration of advances in glycobioinformatics tools and databases with high-performance computing (HPC) molecular simulations is essential[16,17,18,19]. Such tools are extraordinary, especially in the context where information from multiple sources is required to decipher the "GlycoCode"[20]. To support this effort, our lab has developed GlycoShape (https://glycoshape.org), an open-access glycan 3D structure database. GlycoShape integrates missing 3D structural information for glycans and glycoproteins, leveraging over 1 millisecond of cumulative sampling from molecular dynamic simulations (MD) in dilute solution under standard conditions. In general, GlycoShape can be used to improve and, where applicable, correct underlying protein structural information.

The GlycoShape Glycan Database (GDB) contains structural information collected by our research group through MD simulations. The database currently hosts data for over 840 glycan structures and it is continually growing. These structures also include selected examples from non-human species (mammals, invertebrates, plants, fungi, and bacteria). Each glycan in the GDB is represented by distinct 3D conformers. This 3D information can be used by a custom-made tool called Re-Glyco (https://glycoshape.org/reglyco) to rebuild glycoproteins in their functional and native states. Re-Glyco allows users to retrieve protein structures from public repositories (such as the RCSB PDB or the AlphaFold Protein Structure Database[20–22]) or upload their own PDB file. The algorithm employs a highly efficient genetic algorithm (GA) that minimizes a loss function accounting for steric hindrance, and is capable of restoring all types of glycosylation (C-, N-, and O-).

This chapter focuses on the 185 structures I built and analyzed, which cover multiple glycan classes most notably O-glycans and ganglioside-like oligosaccharides.

7.2 Computational Methods

7.2.1 Molecular Dynamics

The 3D starting structures of all glycans were built with the carbohydrate builder tool on GLYCAM-WEB (https://glycam.org/). Conformational sampling requires at least 3 replicas per glycan structure. For glycan structures that theoretically correspond to two energetically available conformers, I used two replicas per conformer. For structures that theoretically corresponded to three or more energetically available conformers, we used one replica for each conformer. Topology files were generated with the *tleap* program included in the AMBER 18[23] software package. The carbohydrate atoms were represented using the GLYCAM06j version of the GLYCAM06 force field[23,24], while the counterions were represented using the AMBER ff14SB[25]. Water molecules were represented using the TIP3P water model[26]. All MD simulations were run at 200 mM NaCl concentration using AMBER18 or GROMACS 2022.4 (GROMACS)[27], depending on the HPC infrastructure available. All AMBER18 simulations began with an energy minimisation phase of 500,000 steepest-descent steps. During minimisation, only hydrogen atoms, water molecules, and counterions were left unrestrained, while the position of all other atoms was kept restrained with a force constant of 5 kcal/mol·Å-2. Following minimisation, the system was brought up to standard temperature through a heating phase

in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During phase one, the temperature was raised from 0 to 100 K, with the volume held constant, and from 100 to 300 K during phase two, with the pressure held constant. The system was then equilibrated for 1 ns in the NPT ensemble at a 1 atm pressure with the Berendsen barostat. During these equilibration stages, we used the same position restraints we used during the energy minimisation. A conformational equilibration was performed for 10 ns with all positional restraints removed. The production phase followed 500 ns for each replica of the glycan structure with all positional restraints removed. For the GMX simulations, the Amber topology (.prm7) and structure (.rst7) files were converted to the corresponding GMX topology (.top) and structure (.gro) files using ACPYPE[28]. The GMX simulation protocol is analogous to the one we used to set up the AMBER18 simulations, starting with an energy minimisation stage followed by 500,000 steps of steepest descent. Following minimisation, the system was equilibrated in the NVT ensemble using a Langevin thermostat to reach 300 K, followed by equilibration in the NPT ensemble using a Parrinello–Rahman barostat to reach 1 bar. Production runs followed, during which we ran a minimum of 500 ns of unrestrained MD trajectories for each structure replica.

7.2.2 Data Processing and Clustering

Upon completion of the production phases, the data are processed and clustered using the Glycan Analysis Pipeline (GAP) as described in the reference[29]. In this section, I will briefly summarise the GAP workflow below for completeness. The production trajectories for each glycan are merged into one dataset. Each frame from the dataset is then transformed into a graph (distance) matrix. The matrices are then flattened to a one-dimensional array by discarding their lower halves. The dimensionality of the array is then reduced using principal component analysis (PCA) from the *sklearn* library. The data is then clustered using a Gaussian Mixture Model (GMM) from the *sklearn* library, and the optimal number of clusters is determined based on the silhouette score. The 3D structures associated with these clusters correspond to the corresponding Kernel Density Estimate (KDE) max values and their corresponding torsion angles (ϕ, ψ, ω) can be sourced on GlycoShape (https://glycoshape.org).

7.3 Results

7.3.1 O-glycans structures

Protein glycosylation in mammals occurs along the secretory pathway, specifically within the endoplasmic reticulum and Golgi apparatus. Unlike N-glycosylation, which attaches a pre-assembled oligosaccharide to asparagine within N-X-S/T sequons (where X is not proline) during protein folding and is subsequently remodelled in the Golgi, O-glycosylation is predominantly initiated after protein folding in the Golgi[30]. This process proceeds one monosaccharide at a time through the action of a diverse set of glycosyltransferases. This Chapter focuses on O-glycans initiated by O-linked α-GalNAc on serine or threonine residues[31]. The human genome encodes 20 distinct isoforms of UDP-GalNAc:polypeptide N-acetyl-galactosaminyltransferases (GalNAc-Ts; genes *GALNTx*). Differential expression and localization of GalNAc-Ts regulate the tissue-specific composition of the O-glycoproteome and, as a result, influence protein function. Elaboration of the initial α-GalNAc (Tn antigen) generates mucin-type O-glycans, which are commonly classified according to their core structures, see **Figure 1.1**. Four major core types are recognized: core 1 (Galβ1-3GalNAcα; the T antigen), core 2 (GlcNAcβ1-6 linked to core 1), core 3 (GlcNAcβ1-3GalNAcα), and core 4 (GlcNAcβ1-6 linked to core 3). Additional minor cores are present in specific biosynthetic contexts. Most mucin-like O-glycans on non-mucin proteins belong to the core 1 and core 2 families, whereas cores 3 and 4 are primarily restricted to mucins, with some documented exceptions[31,32].

In this work, I constructed and curated O-glycan entries representing the various core types and their elaborated derivatives. For each glycan, I performed solution MD simulations to generate conformational ensembles. The most populated cluster representatives were deposited into GlycoShape as three-dimensional conformers, see **Figure 7.2**. Where necessary, multiple conformers were retained to capture biologically relevant microstates and facilitate downstream use by Re-Glyco.

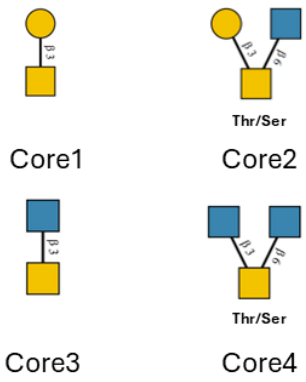
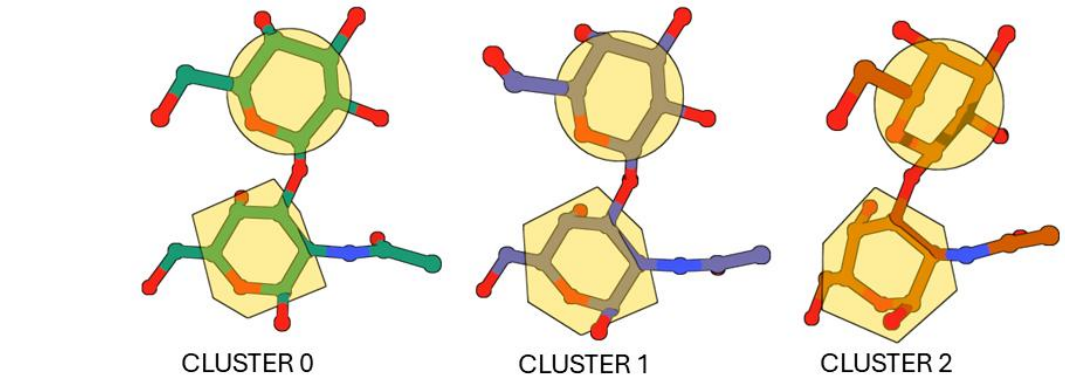


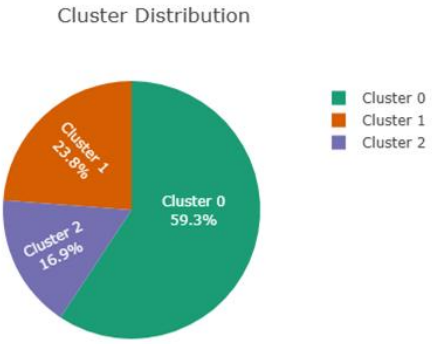
Figure 7.1. Schematic representation of the four major mucin O-glycan cores using SNFG notation: core 1 (Galβ1-3GalNAcα-Ser/Thr), core 2 (GlcNAcβ1-6[Galβ1-3]GalNAcα-Ser/Thr), core 3 (GlcNAcβ1-3GalNAcα-Ser/Thr), and core 4 (GlcNAcβ1-6[GlcNAcβ1-3]GalNAcα-Ser/Thr). Symbols and colors follow SNFG (Gal: yellow circle; GalNAc: yellow square; GlcNAc: blue square).

After assembling the panel of mucin O-glycans spanning cores 1–4 and their common decorations (± sialylation, ± fucosylation), I ran solution MD for each structure. For every trajectory we extracted the ϕ/ψ torsions of the defining glycosidic linkage and clustered the conformations using the Glycan Analysis Pipeline (GAP)[29]: MD frames are converted to distance-matrix vectors, reduced by PCA, then clustered with a Gaussian Mixture Model (GMM); the optimal number of clusters is chosen by the silhouette score, and representative conformers are the KDE-max structures within each cluster, as shown in **Figure 7.2.**



phi_angle

Cluster	Mean	Standard Deviation
0	-73.66	11.01
1	-71.83	10.72
2	-84.25	25.32



psi_angle

Cluster	Mean	Standard Deviation
0	107.69	24.20
1	110.08	26.03
2	104.65	24.00

Figure 7.2. Example conformational clustering for a core-1 O-glycan (SNFG renderings). Illustrative conformers and populations for Galβ1-3GalNAcα-Ser/Thr (core 1), obtained with the GAP workflow. Three clusters capture the dominant β1→3 torsion: Cluster 0 (59.3%) $\phi = -73.66° \pm 11.01°$, $\psi = 107.69° \pm 24.20°$; Cluster 1 (23.8%) $\phi = -71.83° \pm 10.72°$, $\psi = 110.08° \pm 26.03°$; Cluster 2 (16.9%) $\phi = -84.25° \pm 25.32°$, $\psi = 104.65° \pm 24.00°$. The same analysis was applied to all derivatives across cores 1–4 (sialylated/non-sialylated; fucosylated/non-fucosylated).

7.3.2 Ganglioside oligosaccharide structures

My contribution to the 185 structures also includes ganglioside-like oligosaccharides[33]. Gangliosides are glycosphingolipids enriched in vertebrate plasma membranes, composed of a ceramide lipid anchor and an oligosaccharide headgroup. Here we focus on the headgroups treated as soluble oligosaccharides for conformational sampling and database deposition. The standard nomenclature reflects sialic-acid content[34,35] GM (mono-), GD (di-), GT (tri-) with representative examples GM3, GM2, GM1, GD1a, GD1b, GT1b. Patterns and positions of sialylation modulate charge and receptor recognition (e.g., Siglec-6/10), which motivated our coverage: the selection of structures was guided by the projects described earlier (Siglec-6 with GM1/GM3; Siglec-10 with GT1b).

For each ganglioside headgroup in **Figure 7.3**, we applied the same protocol used for mucin O-glycans. The resulting most-populated 3D conformers were curated and deposited in the GlycoShape database.

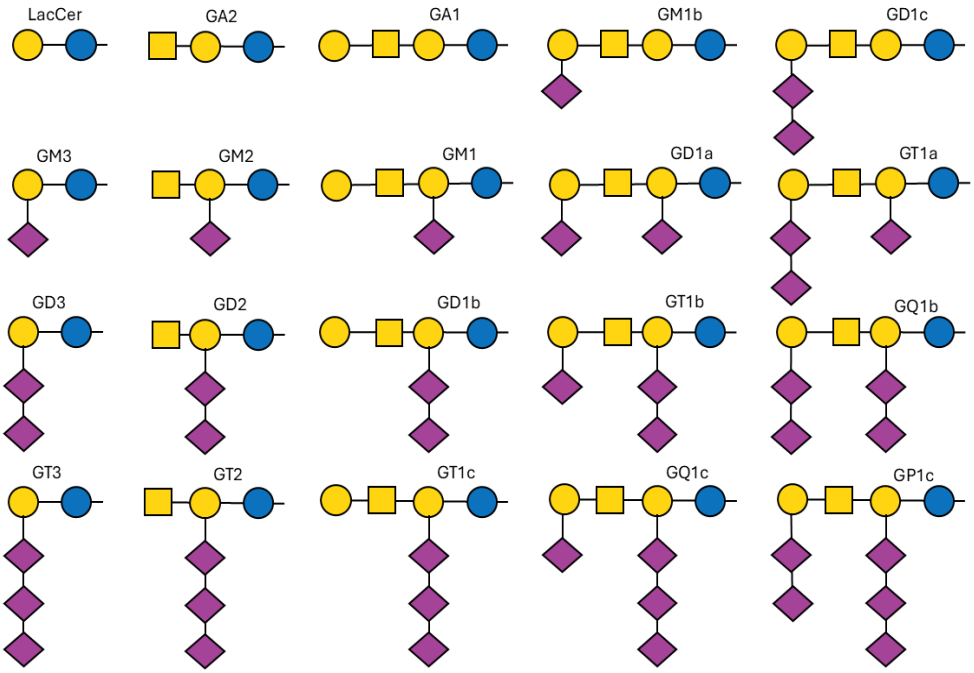


Figure 7.3 Asialo precursors: LacCer, GA2, GA1. Monosialo: GM3, GM2, GM1, GM1b. Disialo: GD3, GD2, GD1a, GD1b, GD1c. Trisialo: GT3, GT2, GT1a, GT1b, GT1c. Tetrasialo: GQ1b, GQ1c. Pentasialo: GP1c. Symbols/colours follow SNFG (Gal: yellow circle; GalNAc: yellow square; Glc: blue circle; Neu5Ac: purple diamond); series labels a/b/c denote alternative sialylation topologies within each class.

7.4 Discussion

During my PhD, I contributed 185 curated 3D structures to GlycoShape. Glycan selection was closely aligned with the projects discussed in previous chapters, ensuring the database directly supported the research questions of this thesis.

O-glycans were prioritized due to their central role in Chapter 3, where we analyzed core-1 O-glycosylation in both sialylated and non-sialylated forms. To support these studies, I generated conformational ensembles for the four major mucin cores and deposited their derivatives, explicitly encoding the dominant φ/ψ angles of key linkages for use in modeling and rebuilding tasks.

A second focus was the family of ganglioside-like oligosaccharides, selected because of their relevance to Siglecs recognition in this thesis. In particular, the GM1/GM3 series is used in the Siglec-6 work (Chapter 5), while GT1b is key for Siglec-10 (Chapter 4). Treating these as soluble headgroups for MD allows us to supply transferable 3D priors

that can later be embedded in membrane systems; the same clustering pipeline yields a compact set of conformers that capture the most populated states. Beyond glycan ensembles, Re-Glyco was essential in several chapters. In Chapter 4 (CD52), it enabled N-glycosylation of peptide and protein models for simulation and analysis. In Chapter 5 (Siglec-6), GlcNAc Scan identified potential N-glycosylation sites, and Re-Glyco was used to rebuild glycosylated receptor models for subsequent molecular dynamics. This integrated workflow, GAP ensembles, Re-Glyco rebuilding, and system-specific simulations, recurs throughout the thesis and supports the mechanistic findings in earlier chapters. In summary, the database additions in this chapter were designed to support the thesis narrative: O-glycans enabled analysis of core-1 (with and without sialic acid) effects from Chapter 3, while ganglioside headgroups supported Siglec-6 and Siglec-10 recognition studies.

7.5 Conclusions

This chapter details my direct contributions to the GlycoShape database, which addresses gaps in 3D structural data for glycans and glycoproteins using high-performance MD simulations. During my PhD, I curated and analyzed 185 3D structures across multiple glycan classes, with a focus on O-glycans and ganglioside-like oligosaccharides. Structure selection was guided by the projects discussed in earlier chapters, and prioritized these two glycan families. For O-glycans, which were central to Chapter 3, I generated conformational ensembles for the four major mucin cores and their derivatives, including sialylated and fucosylated forms. The most representative conformers are deposited in GlycoShape, with the dominant ϕ/ψ torsion angles explicitly encoded for modeling and rebuilding. Ganglioside-like oligosaccharides were selected for their relevance to Siglecs recognition. I sampled the headgroups of the GM1/GM3 series for Siglec-6 in Chapter 5 and the GT1b headgroup for Siglec-10 in Chapter 4, providing transferable 3D priors for future membrane system studies. The most populated conformers were curated and added to GlycoShape, supporting Siglec-6 and Siglec-10 recognition studies. The glycan conformational ensembles I generated with the Glycan Analysis Pipeline (GAP) and deposited in GlycoShape enabled the use of the tool Re-Glyco. Re-Glyco was essential for N-glycosylation of peptide and protein models in Chapter 4 and for rebuilding glycosylated receptor models for MD in Chapter 5. In summary, the contributions to the database described here are strategically designed to

support my projects, providing a 3D structural foundation for the molecular simulations and mechanistic insights presented in previous chapters.

References

1. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. Essentials of Glycobiology, Fourth Edition. 2022. 892 p.
2. He M, Zhou X, Wang X. Glycosylation: mechanisms, biological functions and clinical implications. Signal Transduct Target Ther. 2024 Aug 5;9(1):194.
3. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. Nat Rev Mol Cell Biol. 2020 Dec;21(12):729–49.
4. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012 June 22;13(7):448–62.
5. Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
6. Bloch JS, John A, Mao R, Mukherjee S, Boilevin J, Irobalieva RN, et al. Structure, sequon recognition and mechanism of tryptophan C-mannosyltransferase. Nat Chem Biol. 2023 May;19(5):575–84.
7. Hofsteenge J, Müller DR, de Beer T, Löffler A, Richter WJ, Vliegenthart JF. New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. Biochemistry. 1994 Nov 22;33(46):13524–30.
8. Fadda E. Molecular simulations of complex carbohydrates and glycoconjugates. Curr Opin Chem Biol. 2022 Aug;69:102175.
9. Woods RJ. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. Chem Rev. 2018 Sept 12;118(17):8005–24.
10. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol. 2019 June;15(6):346–66.
11. Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology. 2012 Nov;22(11):1440–52.
12. Zacchi LF, Schulz BL. N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. Glycoconj J. 2016 June;33(3):359–76.
13. Čaval T, Heck AJR, Reiding KR. Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation Between Sites on the Same Glycoprotein. Mol Cell Proteomics. 2021;20:100010.
14. Struwe WB, Robinson CV. Relating glycoprotein structural heterogeneity to function - insights from native mass spectrometry. Curr Opin Struct Biol. 2019 Oct;58:241–8.
15. Agirre J, Davies G, Wilson K, Cowtan K. Carbohydrate anomalies in the PDB. Nat Chem Biol. 2015 May;11(5):303.
16. Cioce A, Calle B, Rizou T, Lowery SC, Bridgeman VL, Mahoney KE, et al. Cell-specific bioorthogonal tagging of glycoproteins. Nat Commun. 2022 Oct 25;13(1):6237.
17. Anggara K, Sršan L, Jaroentomeechai T, Wu X, Rauschenbach S, Narimatsu Y, et al. Direct observation of glycans bonded to proteins and lipids at the single-molecule level. Science. 2023 Oct 13;382(6667):219–23.
18. Wu X, Delbianco M, Anggara K, Michnowicz T, Pardo-Vargas A, Bharate P, et al. Publisher Correction: Imaging single glycans. Nature. 2020 July;583(7818):E32.

19. Feizi T. Carbohydrate recognition in the immune system: contributions of neoglycolipid-based microarrays to carbohydrate ligand discovery. Ann N Y Acad Sci. 2013 July;1292(1):33–44.
20. Dance A. Refining the toolkit for sugar analysis. Nature. 2021 Nov;599(7883):168–9.
21. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan 7;50(D1):D439–44.
22. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Applying and improving AlphaFold at CASP14. Proteins. 2021 Dec;89(12):1711–21.
23. Lee TS, Cerutti DS, Mermelstein D, Lin C, LeGrand S, Giese TJ, et al. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. J Chem Inf Model. 2018 Oct 22;58(10):2043–50.
24. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008 Mar;29(4):622–55.
25. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.
26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983 July 15;79(2):926–35.
27. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015 Sept 1;1-2:19–25.
28. ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. SoftwareX. 2019 July 1;10:100241.
29. Ives CM, Singh O, D'Andrea S, Fogarty CA, Harbison AM, Satheesan A, et al. Restoring protein glycosylation with GlycoShape. Nat Methods. 2024 Nov;21(11):2117–27.
30. Varki A, Chrispeels MJ. Essentials of Glycobiology. CSHL Press; 1999. 702 p.
31. Mitoma J, Petryniak B, Hiraoka N, Yeh JC, Lowe JB, Fukuda M. Extended core 1 and core 2 branched O-glycans differentially modulate sialyl Lewis X-type L-selectin ligand activity. J Biol Chem. 2003 Mar 14;278(11):9953–61.
32. Holmén JM, Karlsson NG, Abdullah LH, Randell SH, Sheehan JK, Hansson GC, et al. Mucins and their O-Glycans from human bronchial epithelial cell cultures. Am J Physiol Lung Cell Mol Physiol. 2004 Oct;287(4):L824–34.
33. Schnaar RL. Gangliosides as Siglec ligands. Glycoconj J. 2023 Apr;40(2):159–67.
34. Schnaar RL, Gerardy-Schahn R, Hildebrandt H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. Physiol Rev. 2014 Apr;94(2):461–518.
35. Kim CH. Ganglioside Biochemistry. Springer Nature; 2020. 223 p.

Chapter 8. Conclusions

The aim of my PhD research was to explore at the atomistic level of detail, how sialylation impacts a variety of molecular systems relevant to human health and disease, and to clarify how this modification regulates the conformation and interactions of the glycoconjugates involved. To achieve this, I used all-atom MD simulations as the primary research tool, a powerful computational approach that enabled me to characterise the structure and dynamics underlying molecular recognition and function.

The results of my work supports the rationale that the role of sialylation is highly diverse and complex, operating within a context-dependent. These effects change with sialic acid position, whether on receptors, peptide ligands, or glycolipids, and with its presentation, such as monovalent or multivalent, and in soluble or membrane-anchored constructs. Sialic acids do not behave as passive terminal caps: their net negative charge and conformational flexibility allow them to remodel local electrostatics, steric accessibility, and interaction networks in ways that can either promote or hinder recognition.

Across the systems examined in this thesis, I observed that sialic acids can act as local inhibitors, structural switches, or multivalent recognition elements. For example, in the endocytic receptor LRP1 (Low-Density Lipoprotein Receptor-Related Protein 1), sialylation can introduce steric hindrance that reshapes ligand preference. In contrast, within the CD52–HMGB1–Siglec-10 immune signalling pathway, hypersialylation functions as a structural switch that promotes and stabilises high-avidity interactions. For membrane-anchored gangliosides recognised by Siglec-6, binding depends on the combined effects of the sialylated epitope and its orientation relative to the lipid membrane, highlighting the importance of membrane context. Moreover, the final distribution of sialic acids is further controlled by upstream modifications, such as the bisecting GlcNAc motif, which can halt glycan maturation and limit the formation of terminal sialylation sites. In parallel, I actively contributed to the creation and refinement of the GlycoShape 3D database (https://glycoshape.org), depositing 185 curated 3D glycan structures. In particular, I focused on O-glycans and ganglioside oligosaccharides,

which were essential for building, visualising, and analysing the glycan structures studied throughout this thesis. This resource not only underpinned the systems investigated here but also provides a reusable structural platform for future work in computational glycobiology.

In summary, by providing an atomistic view across diverse molecular contexts, this thesis shows that sialic acids emerge as active determinants of molecular recognition, far beyond the role of passive terminal decorations. The insights gained here provide a robust structural framework for understanding how sialylation tunes receptor–ligand interactions and offer guiding principles for the design of future glycan-focused therapeutic strategies.

Appendices

Appendix I: Supplementary material from the paper: **D'Andrea S**, Schmidt EN, Bui D, Singh O, Han L, Mahal L, Klassen JS, Macauley MS and Fadda E, in press Communication biology (2025)

Appendix I

Structural confidence and secondary structure analysis of the Siglec-6 V-set domain.

To evaluate the reliability of the AlphaFold-predicted Siglec-6 model and support our structural interpretations from MD simulations, we analyzed per-residue pLDDT confidence values and the secondary-structure stability of the V-set domain. The AlphaFold model shows high to very high confidence (pLDDT > 70) across the V-set domain (residues 32–144), including the flexible C–C′ loop (residues 70–75) and the KW segment (Lys126, Trp127). The loop and the KW segment display blue and light cyan colors, respectively, supporting the reliability of the geometry used in our analysis, see **Figure S.1.a,b.**

We quantified secondary-structure evolution using the DSSP algorithm in *cpptraj* along the MD trajectories of Siglec-6 bound to GM1 (2.5 μs) and GM2 (2.0 μs, merged replicas). For residues 70–75 **(Figure S.1.c)**, the C–C′ loop remains predominantly disordered (coil) with minor bend contributions in both complexes. These results confirm that differences between GM1 and GM2 mainly reflect changes in loop orientation, not secondary structure.

Finally, we overlaid representative snapshots from the GM1 (cyan) and GM2 (magenta) simulations, see **Figure S.1.d**. The V-set fold remains stable in both systems, but the C–C′ loop adopts distinct positions relative to the membrane surface, consistent with Trp127 inserting into or withdrawing from the bilayer.

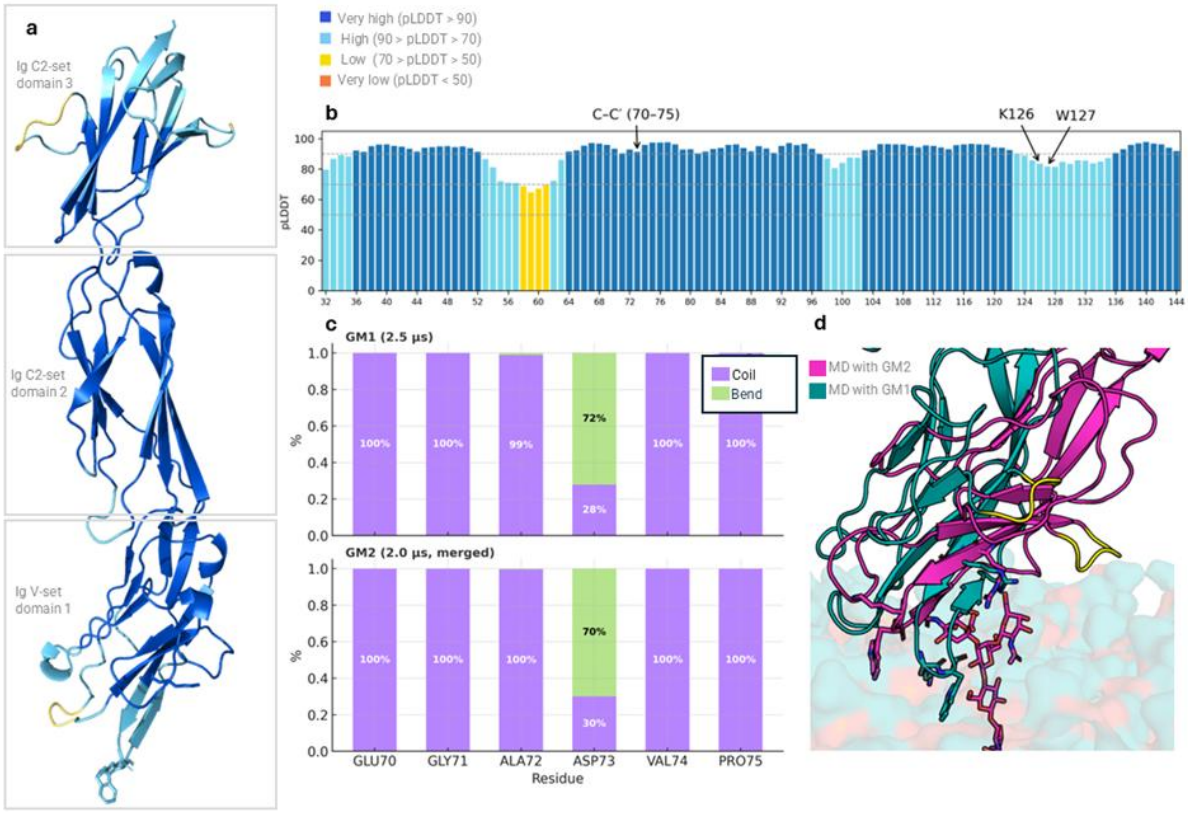


Figure S.1. a. AlphaFold-predicted structure of the Siglec-6 , composed of one Ig V-set and two Ig C2-set domains, colored by pLDDT (B-factor), pLLDT < 50 very low (orange), 50–70 low (yellow), 70–90 high (light cyan), > 90 very high (blue). b. pLDDT profile of the AlphaFold model 32–144(Ig V set domain), the C–C′ loop (70–75) and the KW segment (Lys126, Trp127) indicated by arrows. c) Secondary structure was computed using the DSSP algorithm implemented in *cpptraj* on the MD trajectories of Siglec-6 bound to GM1 (2.5 µs) and GM2 (2.0 µs, merged replicas). The percentage of coil (purple) and bend (green) conformations is reported for each residue. d. Structural overlay of representative snapshots from the GM1 system (cyan) and GM2 system (magenta) trajectories showing the conformational stability of the V-set domain on the membrane and the localization and different orientation of the C–C′ highlight in yellow.

MD simulations of the fully *N*-glycosylated Siglec-6 in complex with membrane bound GM1. We ran two independent MD simulations of 1 µs each on a 3D model of the fully glycosylated Siglec-6 in complex with an isolated GM1 ganglioside embedded in the phospholipid bilayer. The molecular composition of the 3D model is described in the Methods section and, aside from the N-glycans, is analogous as the one presented in the main text. The Siglec-6 sequence (453 aa) carries seven *N*-glycosylation sequons, with only one in the V-set domain at Asn103, see **Figure S.2**. All sites are occupied with a biantennary monogalactosylated complex *N*-glycan (GlyTouCan ID G99129GB; 1478.5 Da). As represented in the snapshot from the MD simulations shown in **Figure S.2**, only the *N*-glycans at N104 and N149 interact with the protein, and more specifically engage

in contacts that may affect the relative orientation and dynamics of the V-set domain (d1) and the adjacent C2 domain (d2). The RMSD values calculated over the entire extracellular region of the glycosylated system (i.e. over the backbone atoms of d1, d2, and d3) are higher than the non-glycosylated one, see **Figure S.2.d**. This increase is not due to destabilization by glycans, but rather reflects the higher conformational mobility of the distal d3 domain, where glycans such as those at N258 and N295 are not involved in protein contacts and remain solvent exposed and dynamic. It is important to note that d3 is connected to the transmembrane domain, and in a complete cellular context, its dynamics may be constrained by interactions with the membrane. Instead, our data support a stabilizing function of the glycans at N103 and N149, which maintain the relative positioning and rigidity of domains d1 and d2. When compared to the non-glycosylated Siglec-6/GM1 complex, the fully glycosylated Siglec-6 shows a slightly lower average RMSD value relative to the AF (AF-O43699-F1) backbone, suggesting a reduced interdomain flexibility and higher structural stability of the d1–d2 interface, see **Figure S.2.c**. However, the relative flexibility of the system demonstrated by the spread around the median RMSD value in the KDE plots in **Figure S.2.d** for both glycosylated and non-glycosylated systems, precludes defining any clear role of the *N*-glycans in affecting the rigidity of the V-set domain relative to the adjacent C2-domain.

As an important note, the binding to the GM1 is not affected by the *N*-glycosylation neither directly, nor indirectly. This is further supported by the MD simulations of the fully glycosylated Siglec-6/GM1 complex, which consistently show stable anchoring on the membrane through the interactions with Trp127 and Lys126.

Membrane-insertion analyses of Trp127 and Lys126 (**Figure S.2.f.**) show similar depth and persistence in glycosylated and non-glycosylated complexes, confirming that ligand engagement is unchanged; the main glycan effect is to decrease d1–d2 mobility, yielding more ordered, less variable, behavior in the RMSD/KDE results. In replica MD1, Trp127 briefly disengaged from the bilayer after about 1 μs, coinciding with the loss of the Arg–Neu5Ac contact and destabilization of the complex.

In replica MD2, the Arg–Neu5Ac interaction was lost after 700 ns, as in the non-glycosylated run, but the complex remained stable thanks to persistent Trp127/Lys126 anchoring and a recurring contact between the terminal Gal of GM1 and Asp70 in the C–C′ loop, as shown in **Figure S.2.e-f**. Distance traces (**Figure S.2.e.**) and insertion profiles

(**Figure S.2.f.**) indicate that V-set domain glycans do not affect GM1 binding, but they reduce inter-domain flexibility and enhance membrane interaction stability.

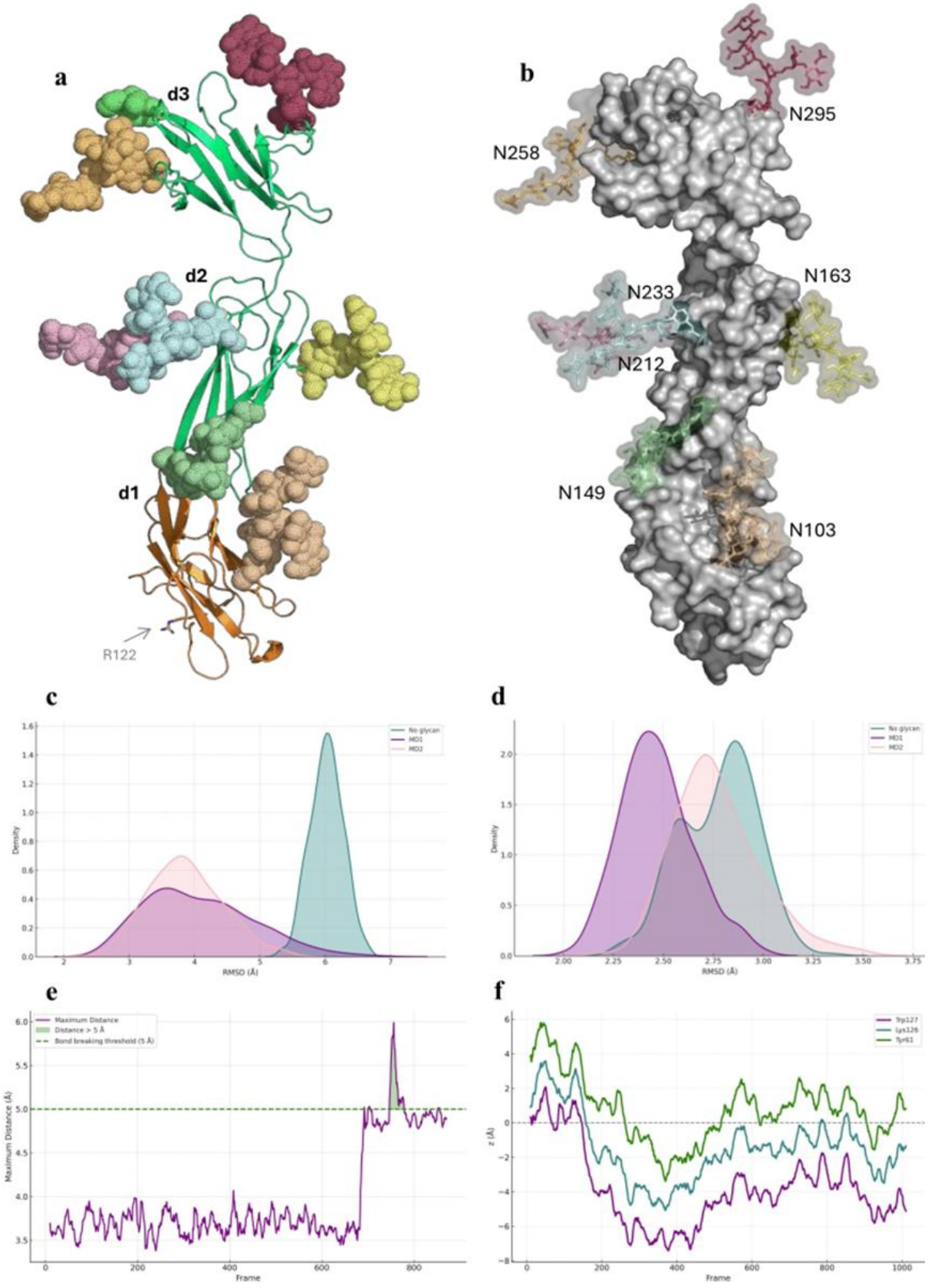


Figure S.2. a. 3D structure of the Siglec-6 extracellular region from AF model AF-O43699-F1, shown as cartoons. The terminal V-set binding domain (d1) is shown in orange and the C2-set domains (d2 and d3) are shown in green. *N*-linked glycans are rendered with atoms as vdW spheres and color-coded individually. The seven *N*-glycosylation sequons were assessed as occupied with the GlcNAc Scanning tool in GlycoShape ReGlyco[5]. **b.** Representative snapshot of the Siglec-6 from the MD simulation with the N-glycosylation sites labelled. **c.** Kernel density estimates (KDE) distributions of the backbone RMSD values along the trajectory, corresponding to the Siglec-6 calculated for d1 and d2. The results corresponding to the non-glycosylated Siglec-6 are shown in green and for the fully glycosylated Siglec-6 in purple and pink for MD1 and MD2, respectively). **d.** Kernel density estimates (KDE) distributions of the backbone RMSD values calculated along the MD trajectories for the backbone of the Siglec-6 extracellular region, i.e. d1 to d3. **e.** Time evolution along the MD trajectory of the distance (Å) between the Arg122 and the Neu5Ac carboxylate group. Data points correspond to the largest distance value calculated between four pairs of atoms R122-NH1(2) and O11(12)-Neu5Ac. A distance of 5 Å was chosen as a threshold for the formation of the salt bridge. **f.** Membrane-insertion depth (z, Å) of Trp127, Lys126 relative to the DSPC phosphorus plane (z = 0, dashed). Negative values indicate insertion into the bilayer. Molecular rendering with *pymol* (www.pymol.org) and KDE plots with *seaborn* and *Matplotlib* (https://seaborn.pydata.org/).

Assessing the potential involvement of non-canonical Arg residues in the V-set domain of Siglec-6 in the recognition and binding of GM1os. Previous work by our collaborators[1] clearly showed that Siglec-6 binds gangliosides with a mechanism independent of the canonical Arg, namely Arg122. Due to the presence of multiple Arg and Lys residues in the V-set domain of Siglec-6, it is reasonable to hypothesize that one of those may have replaced or could complement the function of the conserved Arg, as our collaborators recently demonstrated in the case of Siglec-10[2]. To address this matter we ran a set of three uncorrelated molecular dynamics (MD) simulations of the non-glycosylated Siglec-6 (see results in the main text) in an equilibrated water box, in the presence of seven GM1os molecules. The starting structure in all replicas was built with one of the GM1os bound to the canonical Arg122 to occupy the canonical site, while another was positioned in proximity to Arg92, , see **Figure S.1,** which has been previously highlighted as a potential (non-canonical) binding Arg[1]. We focused our attention to the Arg in the V-set domain, but monitored contacts with all Arg on the Siglec-6 accessible surface, as these could be involved in both *cis,* and *trans* binding events. The results show that none of the GM1os, except the one bonded to the canonical Arg122, engage in stable interactions, see **Figure S.1**. The Arg92 is engaged in a stable salt bridge with Asp115, which is never disrupted during any of the three independent MD simulations. As a caveat, it is important to underline that additive force fields, such

as the one we used to run the MD simulations in this work, are notoriously biased towards enhancing electrostatic contacts[3] and thus tend to overestimate the stability of already strong hydrogen bonding interactions, such as salt bridges[4]. In this specific case an analysis of the Siglec-6 3D structure indicates that such salt bridge interaction may be essential to the correct folding of the V-set domain, and mutation of Arg92 (or of Asp115) would compromise the folding, which explains the low expression levels of the R92A mutant reported in earlier work[1].

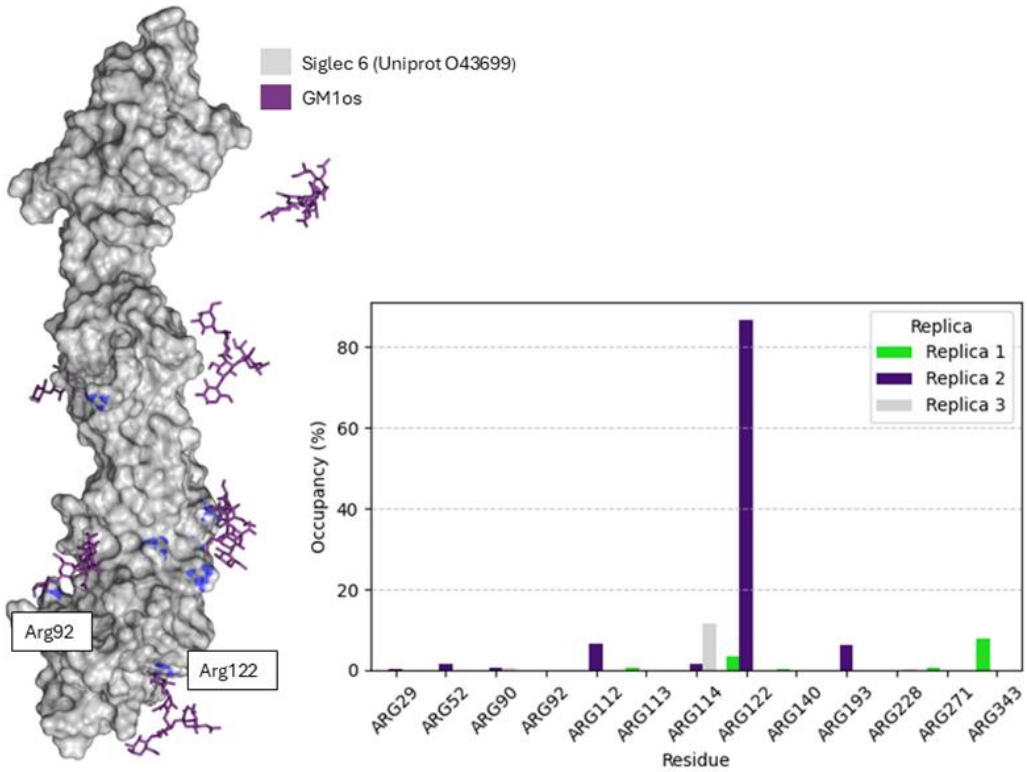


Figure S.1. Left: 3D surface representation of Siglec-6 (AF-O43699, gray) with seven GM1os molecules (purple sticks), including one placed at the canonical sialic acid binding site (Arg122) and others distributed in possible non canonical regions of the protein. **Right:** Bar plot showing the occupancy of GM1os interactions with selected ARG residues across three independent MD replicas. Occupancy is assessed in terms of distance between guanidinium C (CZ) of Arg and the carboxyl carbon (C1) of the Neu5Ac with a threshold of 5 Å for binding. Molecular rendering with VMD (https://www.ks.uiuc.edu/Research/vmd/) and bar plot with MS Excel.

Experimental Methods

Nanodisc preparation

Nanodiscs (NDs), consisting of 10% GM1 and DMPC, were prepared using the protocol described by Sligar and coworkers [12]. Briefly, the lipids were diluted in methanol at the desired molar ratios, dried under gentle vacuum to form a lipid film and then resuspended in a buffer (pH 7.4) containing 20 mM TrisHCl, 0.5 mM EDTA, 100 mM NaCl and 25 mM sodium cholate (Sigma-Aldrich Canada, Oakville, Canada). The membrane scaffold protein MSP1E1 was added to the mixture at a MSP1E1:(GM1+DMPC) molar ratio of 1:100. The ND self-assembly process was initiated by adding pre-washed biobeads (Bio-Rad, Mississauga, Canada) and the mixture was incubating at room temperature overnight on an orbital shaker. After incubation, the supernatant was recovered and the NDs purified using a Superdex 200 10/300 size-exclusion column (GE-Healthcare Life Sciences, Piscataway, NJ) equilibrated with 200 mM ammonium acetate (pH 7.4). Finally, the ND fraction was collected, concentrated and dialyzed into 200 mM ammonium acetate (pH 7.4) using an Amicon microconcentrator (EMD Millipore, Billerica, MA) with a 30 kDa MW cut-off. All ND stock solutions were stored at –80 °C prior to analysis. Each ND sample consists of two copies of MSP1E1 and ~200 lipids. Therefore, ND concentration was estimated based on the UV absorption of MSP1E1 at 280 nm with the extinction coefficient of $\varepsilon_{280nm,MSP1E1} = 32{,}430\ cm^{-1} \cdot M^{-1}$, and assuming [ND] = 1/2×[MSP1E1]. Each ND approximately contained 20 GM1 molecules and 180 DMPC molecules.

Concentration independent native mass spectrometry with catch-and-release

The Concentration Independent native mass spectrometry (COIN-nMS) assay, performed with Catch-and-Release (CaR), was implemented in negative ion mode using a Q-Exactive Ultra-High Mass Range (UHMR) Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a modified nanoflow electrospray ionization (nanoESI) source, as described elsewhere[13].

Catch-and-release native mass spectrometry with ion mobility separation

CaR-nMS, performed with Ion Mobility Separation (IMS), was implemented in negative ion mode using a Waters Synapt G2S quadrupole-ion mobility separation-time of flight (Q-IMS-TOF) mass spectrometer (Waters, Manchester, UK) equipped with a NanoLockSpray ion source. A nanoESI voltage of –0.8 kV, a source temperature of 80 °C and a cone voltage of 70 V were used. Argon was used in the Trap and Transfer ion guides at pressures of 2.77×10^{-2} and 2.84×10^{-2} mbar, respectively, with the Trap and Transfer voltages of 5 and 2 V, respectively. For IMS, a wave height of 40 V and a wave velocity of 650 m s^{-1} were applied along with a helium and nitrogen (IMS gas) gas flow

of 120 and 90 mL min^{-1}, respectively. Collision-induced dissociation (CID) by isolating ions with m/z 7,000±100 in the quadrupole, subjecting them to IMS, and then collisional activation using a 100 V collision energy in the Transfer region [14]. All data was processed using MassLynx software (v4.1) in combination with DriftScope v2.5.

Mutagenesis. Mutagenesis was achieved by three consecutive polymerase chain reactions using gene-overlap extension mutagenesis. The first reaction used a forward primer at the start of the gene containing a 5' *NheI* site and reverse primer that was centered on the mutation site and the complete gene of interest at the template. The second reaction used a forward prime that was centered on the mutation site and a reverse primer at the end of the gene that featured and 3' *AgeI* site and used the gene of interest as the template. The third reaction used the products of the first two PCR reactions as a template and the primers at the start and end of the gene. Following size validation using agarose gel electrophoresis, the final PCR products were digested with *NheI* and *AgeI* and then ligated into pCDNA5. Following ligation, the ligation product was then transformed into *E. coli* DH5α. Colonies were then grown in LB overnight at 37 °C and then miniprepped. The minipreps were then validated by restriction digest and Sanger sequencing.

Table S.1: Primers used in this work.

Primer Name	**Sequence**
Siglec-6 Fwd	AGC AGC GCT AGC ATG CAG GGA GCC CAG GAA GCC
Siglec-6 Rvs	AGC AGC ACC GGT TCA CTT GTG TAT CTT GAT TTC
Sig-6 K124A Fwd	CGG TTG AAG TCC GCA TGG ATG AAA TAC
Sig-6 K124A Rvs	GTA TTT CAT CCA TGC GGA CTT CAA CCG
Sig-6 Rvs 3D	AGC AGC ACC GGT CCT GCC TTC TGG TTT CCA ATG
Sig-6 K126A Fwd	CGG TTG AAG TCC GCG TGG ATG AAA TAC GG

Sig-6 K126A Rvs	CCG TAT TTC ATC CAC GCG GAC TTC AAC CG
Sig-6 W127A Fwd	GTT GAA GTC CAA AGC CAT GAA ATA CGG TTA TAC
Sig-6 W127A Rvs	GTA TAA CCG TAT TTC ATG GCT TTG GAC TTC AAC
Sig-6 K129A Fwd	GTC CAA ATG GAT GGC ATA CGG TTA TAC
Sig-6 K129A Rvs	GTA TAA CCG TAT GCC ATC CAT TTG GAC

Cell culture. Flp-In Chinese hamster ovary (CHO) cells were cultured as previously described[1], but in short, CHO cells were cultured in DMEM/F12 Media (Gibco) supplemented with 5% (V/V) fetal bovine serum (Gibco), penicillin (100 U/mL), and streptomycin (100 μg/mL). Cells were grown at 37 °C and 5% CO_2 in tissue culture dishes.

Stable transfection of Flp-In CHO Cells. Flp-In Chinese[1].

Liposomes preparation. Liposomes were prepared by hydrating lipid thin consisting of DSPC (61.5 mol%), cholesterol (38 mol%), and PEG_{45}-DSPE (0.4 mol) and AF647-PEG_{45}-DSPE films with PBS pH 7.4. Glycolipids were added at the expense of DSPC. The lipid solutions were extruded using an Avanti-mini extruder using 800 nm and 100 nm filters respectively. Liposomes were then validated with dynamic light scattering. More details regarding liposome production can be found elsewhere[1],[15].

Siglec-Fc expression and purification. The procedure for Siglec-Fc expression and purification can be found elsewhere[16])

Cell assay. 150,000 cells were added to a 96 well U-bottom microplate. Liposome solutions were prepared at 50 μM in 1% (g:mL) BSA PBS pH 7.4. 50 μL of liposome solution was added to the cells and the cells were incubated with the liposomes for 30 min at 37 °C. The cells were then washed with PBS twice. Following the washes, the cells were then resuspended in anti-Siglec antibody solution (1/250 V:V antibody in 500 μM EDTA, 1% BSA (g:mL)). The cells were then incubated at 4 °C for 30 min. The cells

were again washed and then analyzed by flow cytometry. Antibody information can be found in Table S.2, and general gating strategy can be found in Figure S.3.

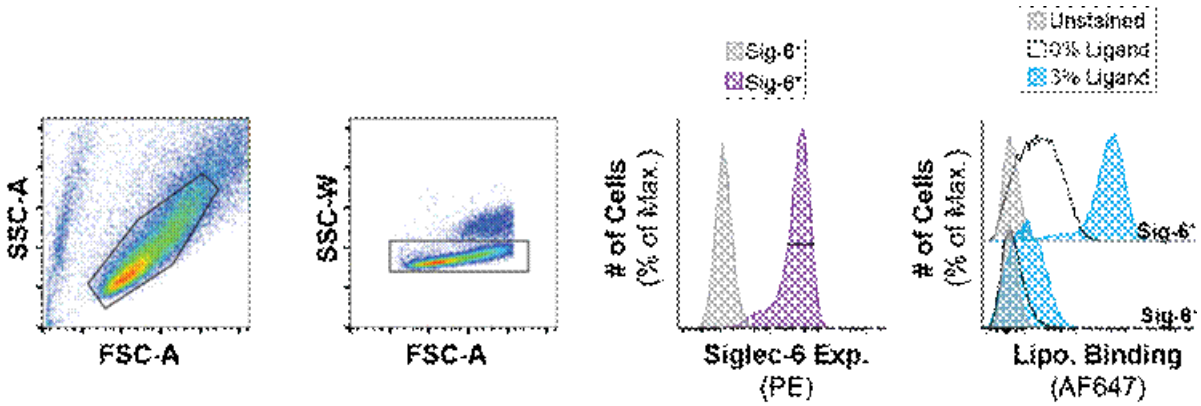


Figure S.3. Gating strategy used to measure glycolipid liposome binding to Chinese hamster ovary cells expressing Siglec-6 mutants. Liposome binding was measured using 3 mol% compound NGL1 [1].

Chemical structure of neoglycolipid ligand (NGL1) used in this experiment. Synthesized as described previously [1].

Table S.2: Antibodies used in this work

Antibody	**Supplier**	**Cat. No.**	**Label**	**Clone**	**Isotype**	**Dilution**
anti-Siglec-6	R&D Systems	FAB2859P	PE	767329	Mouse IgG2A	1:250 (V:V)

Flow cytometry. Flow cytometry measurements were collected on a 5-laser Fortessa X-20 (BD Bioscience). All the resulting data were analyzed using FlowJo (10.5.3) software (BD Biosciences)

Siglec-Fc ELISA. A detailed description of the approach can be found elsewhere [1,15]. In short, 50 µL of 50 µM glycolpid solution was transferred to a 96 well ELISA microplate. The plate was left at 37 °C for 2 to remove the ethanol. The wells were blocked with 5% (g:mL) bovin serum albumin (BSA) for 1 h at room temperature. During the blocking

step, the Siglec-Fc-Strep-Tactin solution (Siglec-Fc solution: 2 µg/mL, Strep-Tactin-HRP: 0.13 µg/mL-IBA) was prepared and complexed at room temperature for at least 30 min at room temperature. The blocking buffer was then removed, and the Siglec-Fc solution was added for 2 h at room temperature. Unbound Siglec-Fc solution was then removed, and the amount of binding was measured using TMB substrate. The reaction was then quenched using 1 M H_3PO_4 and the Abs_{450} was measured using a Molecular Devices SpectraMAX iD5.

References

(1) Schmidt, E. N.; Lamprinaki, D.; McCord, K. A.; Joe, M.; Sojitra, M.; Waldow, A.; Nguyen, J.; Monyror, J.; Kitova, E. N.; Mozaneh, F.; Guo, X. Y.; Jung, J.; Enterina, J. R.; Daskhan, G. C.; Han, L.; Krysler, A. R.; Cromwell, C. R.; Hubbard, B. P.; West, L. J.; Kulka, M.; Sipione, S.; Klassen, J. S.; Derda, R.; Lowary, T. L.; Mahal, L. K.; Riddell, M. R.; Macauley, M. S. Siglec-6 Mediates the Uptake of Extracellular Vesicles through a Noncanonical Glycolipid Binding Pocket. *Nat. Commun.* **2023**, *14* (1), 2327.

(2) Sobczak, K.; Antoñana-Vildosola, A.; Valverde, P.; Travecedo, M. A.; Jame-Chernaboo, Z.; Schmidt, E. N.; D'Andrea, S.; Valdaliso-Díez, E.; Oyenarte, I.; Laugieri, M. E.; Joe, M.; Mozaneh, F.; Lin, S.-Y.; Bosch, A.; Moure, M. J.; Franconetti, A.; Lee, S. Y.; de Durana, J. E.-D.; Pérez-Gutierrez, L.; Palazón, A.; Marcelo, F.; Fadda, E.; Corzana, F.; Gimeno, A.; Macauley, M. S.; Jiménez-Barbero, J.; Ereño-Orbea, J. The Unique Molecular Recognition Features of Siglec-10: Structural Insights into Sialoglycan and Antibody Interactions. *bioRxiv*, 2025. https://doi.org/10.1101/2025.06.10.658867.

(3) Molecular Simulations of Complex Carbohydrates and Glycoconjugates. *Current Opinion in Chemical Biology* **2022**, *69*, 102175.

(4) Ahmed, M. C.; Papaleo, E.; Lindorff-Larsen, K. How Well Do Force Fields Capture the Strength of Salt Bridges in Proteins? *PeerJ* **2018**, *6*, e4967.

(5) Ives, C. M.; Singh, O.; D'Andrea, S.; Fogarty, C. A.; Harbison, A. M.; Satheesan, A.; Tropea, B.; Fadda, E. Restoring Protein Glycosylation with GlycoShape. *Nat Methods* **2024**, *21* (11), 2117–2127.

(6) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D.

Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

(7) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50* (D1), D439–D444.

(8) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859–1865.

(9) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., 3rd; MacKerell, A. D., Jr; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **2016**, *12* (1), 405–413.

(10) Case, D. A.; Metin Aktulga, H.; Belfon, K.; Ben-Shalom, I.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Andrés Cisneros, G.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2022*; University of California, San Francisco, 2022.

(11) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38, 27–28.

(12) Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G. Chapter 11 - Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol* **2009**, *464*, 211–231.

(13) Bui, D. T.; Favell, J.; Kitova, E. N.; Li, Z.; McCord, K. A.; Schmidt, E. N.; Mozaneh, F.; Elaish, M.; El-Hawiet, A.; St-Pierre, Y.; Hobman, T. C.; Macauley, M. S.; Mahal, L. K.; Flynn, M. R.; Klassen, J. S. Absolute Affinities

from Quantitative Shotgun Glycomics Using Concentration-Independent (COIN) Native Mass Spectrometry. *ACS Cent. Sci.* **2023**, *9* (7), 1374–1387.

(14) Li, J.; Fan, X.; Kitova, E. N.; Zou, C.; Cairo, C. W.; Eugenio, L.; Ng, K. K. S.; Xiong, Z. J.; Privé, G. G.; Klassen, J. S. Screening Glycolipids Against Proteins in Vitro Using Picodiscs and Catch-and-Release Electrospray Ionization-Mass Spectrometry. *Anal Chem* **2016**, *88* (9), 4742–4750.

(15) Schmidt, E. N.; Guo, X. Y.; Bui, D. T.; Jung, J.; Klassen, J. S.; Macauley, M. S. Dissecting the Abilities of Murine Siglecs to Interact with Gangliosides. *J. Biol. Chem.* **2024**, *300* (7), 107482.

(16) Schmidt, E. N.; Jung, J.; Macauley, M. S. Flow Cytometry-Based Detection of Siglec Ligands. *Methods Mol. Biol.* **2023**, *2657*, 181–193.

Appendix II: **S. D'Andrea***, A Satheesan*, E. Patterson, CA. Fogarty, A. Maggioni, M. von Itzstein, D. Kolarich and E.Fadda. *Effect of bisection on the 3D structure and recognition of N-glycans during maturation: Does MGAT3 introduce a stop codon?* (in preparation)

Appendix II

The following is adapted from Dr. A. Satheesan's thesis and included here for completeness.

Assessing the suitability of A2B and A3 as substrates for β4GalT1

β4-galactosyltransferase 1 (β4GalT1, EC 2.4.1.38) is a transmembrane Golgi protein that transfers galactose from a UDP-α-D-galactose donor to a terminal N-acetyl-β-D-glucosamine.

My colleague, Akash Satheseen, followed a protocol similar to the FUT8 study to assess the suitability of A2B and A3 as substrates for β4GalT1.

We used the X-ray structure PDB 4EE4 solved at 1.95 Å and co-crystallized with a tetrasaccharide from lacto-N-neohexose [GlcNAcβ1–3(GlcNAcβ1–6)Galβ1–4Glc]. Recognition complexes were built by aligning the terminal β2-GlcNAc of the most populated A2B and A3 MD conformers onto the crystallographic acceptor. Structural superpositions show that, in the presence of a bisecting GlcNAc, steric clashes arise between the bisected core and residues lining the β4GalT1 acceptor-binding pocket, preventing productive accommodation and thus disfavoring A2B as a substrate (see **Figure 6.7**). Notably, when the alignment is performed on the α3 arm, the N-glycan core flips toward the protein, rendering the substrate inaccessible not only because of the collision with Y291 but also due to the incorrect orientation of the core relative to the catalytic geometry, see **Figure 6.7**.

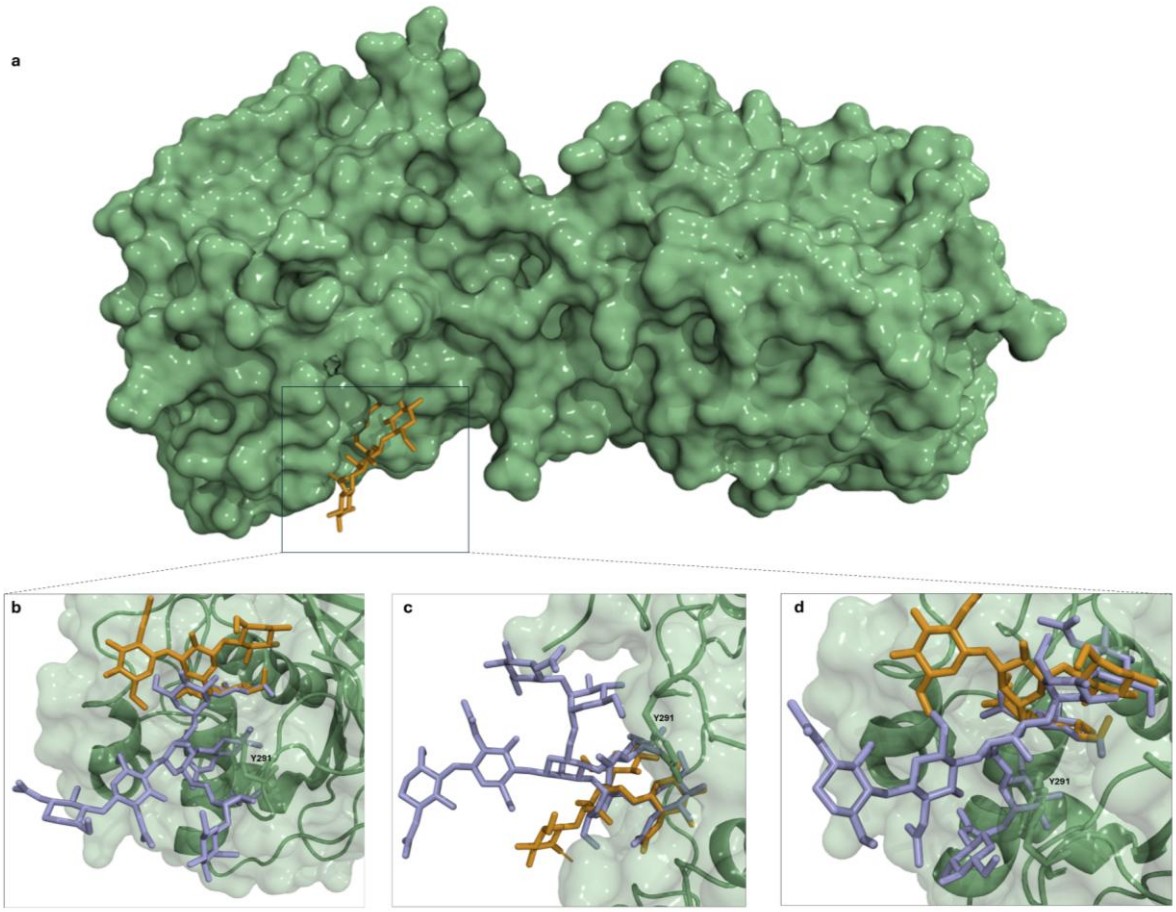


Figure 6.7. a. Overall view of β4GalT1 (PDB 4EE4) shown as surface (green). The structure was solved at 1.95 Å and crystallized with a tetrasaccharide from lacto-N-neohexose, GlcNAcβ1–3(GlcNAcβ1–6)Galβ1–4Glc shown as sticks in yellow. **b-d.** Close-ups of the β4GalT1 acceptor site: the bisected N-glycan conformations for all three arms (from MD) are shown as light-blue sticks and overlaid on the crystallographic acceptor (orange), highlighting steric clashes with pocket residues (e.g., Y291, labelled). Renderings generated with PyMOL (www.pymol.org).

Regarding the structural alignment of the terminal β2-GlcNAc on the third antenna the 'open' A3 conformer with the co-crystallised ligand produced a good fit where in the arms were able to exist in complex with the β4GalT1 structure without any steric clashes. Results from the 1μs of MD simulation are shown in **Figure 6.8.** The A3 structure remained bound for the duration of the simulation with an average RMSD of 4.56 Å and a standard deviation of 0.48 Å. The A3 structure was able to interact with residues lining the binding site, namely Q173, D203, D204 and R244, see **Figure 6.8.**

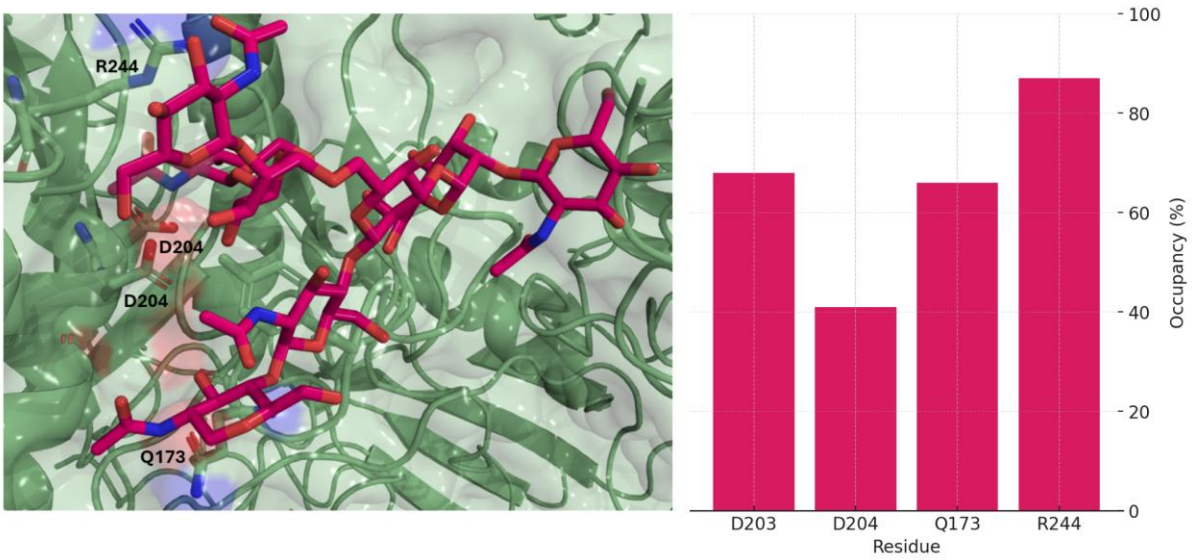


Figure6.8. Left. Representative snapshot from the MD trajectory of the recognition complex between A3 and β4GalT1. β4GalT1 rendered as surface (green) and A3 rendered as sticks in dark green. **Right.** Barplot indicating the persistence of the hydrogen bonding interactions between the A3 structure and specific binding site residues calculated during the MD trajectory.