

Emerging out of the “blur”: exploration, evaluation  
and significance of 3D N-glycans’ structure through  
molecular dynamic studies



**Maynooth  
University**

National University  
of Ireland Maynooth

**Aoife Harbison** B.Sc. (Hons)

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

**Doctor of Philosophy**

By

Aoife Harbison B. Sc.

Department of Chemistry

Maynooth University

**February 2021**

**Research supervisor:** Dr. Elisa Fadda

**Head of Department:** Dr. Denise Rooney

# Table of Contents

Acknowledgements .....	i
Declaration.....	ii
Abbreviations .....	iii
Abstract.....	1
List of Publications .....	1
Chapter 1: Introduction .....	4
References.....	12
Chapter 2: Computational Method .....	17
References.....	26
Chapter 3: Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study .....	29
3.1 Introduction .....	29
3.2 Computational Methods .....	31
3.3 Results.....	32
3.4 Discussion .....	43
3.5 Conclusions .....	46
References.....	47
Chapter 4: How and why plants and human N-glycans are different: Insight from molecular dynamics into the “glycoblocks” architecture .....	51
4.1 Introduction .....	51
4.2 Computational Methods .....	54
4.3 Results.....	55
4.4 Discussion .....	62
4.5 Conclusions .....	65
References.....	66
Chapter 5: An Atomistic Perspective on Antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 N-glycans from enhanced sampling molecular dynamics .....	69
5.1 Introduction .....	69
5.2 Computational Methods .....	71
5.3 Results.....	72
5.4 Discussion .....	79
5.5 Conclusions .....	80
References.....	81

Chapter 6: All-Atom Molecular Dynamics Simulations of SARS-CoV-2 Spike Protein with glycan variation.....	84
6.1 Introduction .....	85
6.2 Computational Method .....	89
6.3 Results.....	94
6.4 Discussion .....	106
6.5 Conclusions .....	108
References.....	109
Chapter 7: Conclusions .....	113
References.....	115
Appendices	

## Acknowledgements

First, I would like to thank my supervisor Dr. Elisa Fadda for the immeasurable support you've given to me over the last few years through this PhD journey. I really appreciate the faith you had in me to succeed, even when I found some obstacles near impossible to overcome. I've truly enjoyed being your PhD student, and getting to know you as a person.

I would also like to thank the other members of our research group, past and present. I would particularly like to thank Matthew and Carl, for indulging me in all the endless chatting, group Halloween costumes and tea breaks. It was great to be part of the team, and learn from each other.

I would like to thank Maynooth University Graduate Studies office for their support of my research with the John Hume Scholarship, and the HEA and D/FHERIS for additional funding to finish my research.

Many thanks to the members of staff from the Chemistry Department and the Hamilton Institute, who have helped me, gave me advice and had good discussions with me in the last four years.

To all the postgrads of the Chemistry Department and Hamilton Institute, thanks for bringing fun and comradery to this experience. I'll always remember the parties, board game nights, collective junk food binges, lunch time chats and general letting off of steam that made the harder parts of being a PhD student more bearable.

I would like to thank my parents, Kevina and Tom, and my brother Eoin, for all the support they've provided for me. Thanks for standing by me through all my highs and lows, and for your unconditional love, I'm forever grateful. Thanks to all my friends who have rooted me on and kept me going and somewhat sane. Many thanks to Pauline for being a great listener and providing me with perspective.

To Joe, thank you for not only putting up with the stress of your own PhD, but also putting up with the stress of me doing one concurrently. You've been the calm in the storm when I've needed it, a stalwart support, and a source of laughter and reason in equal measure. We've journeyed through this 3<sup>rd</sup> level experience together and I hope we can keep supporting each other and learning more (preferably in a less intense setting).

Finally, I would like to thank Library Cat, for just being a cat. Maynooth University wouldn't be the same without you.



## Declaration

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

Aoife Harbison

*Aoife Harbison*

---

## Abbreviations

ACE2	Angiotensin-Converting Enzyme 2
ADCC	Antibody Dependant Cellular Cytotoxicity
AMBER	Assisted Model Building with Energy Refinement
Asn	Asparagine
CH Domain	Chain (Heavy) Domain
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CoV	Coronavirus
Cryo-EM	Cryogenic Electron Microscopy
DFT	Density Functional Theory
Fc region	Fragment Crystallizable Region
fs	femtosecond
Fuc	Fucose
GlcNAc	N-acetylglucosamine
Gal	Galactose
GUI	Graphic User Interface
HPC	High Performance Computing
ICHEC	Irish Centre for High End Computing
IgG	Immunoglobulin G
KDE	Kernel Density Estimation
kJ/mol	kiloJoules per mole
Le <sup>x</sup>	Lewis X
sLe <sup>x</sup>	Sialyl-Lewis X
Le <sup>A</sup>	Lewis A
sLe <sup>A</sup>	Sialyl-Lewis A

Man	Mannose
MD	Molecular Dynamics
MERS	Middle Eastern Respiratory Syndrome
μs	microseconds
MS	Mass Spectrometry
NAMD	Nanoscale Molecular Dynamics
NeuAc	N-Acetylneuraminic acid (also known as Sialic Acid)
NMR	Nuclear Magnetic Resonance
ns	nanoseconds
NTD	N-Terminus Domain
PCA	Principal Component Analysis
PDB	Protein Data Bank
PES	Potential Energy Surface
ps	picosecond
QM	Quantum Mechanics
RBM	Receptor Binding Motif
RBD	Receptor Binding Domain
REMD	Replica Exchange Molecular Dynamics
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Square Deviation
SARS	Severe Acute Respiratory Syndrome
Sia	Sialic Acid (see Neu5Ac)
Ser	Serine
Thr	Threonine
Xyl	Xylose

## Abstract

Glycosylation is the most abundant and diverse post-translational modification of proteins, contributing to protein folding, trafficking, structural stability and dynamics, and function. Complex N-glycans are a class of glycans found in eukaryotes, sharing a common pentasaccharide core structure. The functionalization of the arms and the branching patterns are specific to different species, while the complexity of the cellular biosynthetic pathways contribute to the broad variety and to the heterogeneity of N-glycan sequences. By understanding at an atomistic level of detail the structural implications of glycan sequence, we can relate the glycan sequence to its function in a given glycoprotein environment.

With this ultimate goal in mind I conducted, through conventional and enhanced molecular dynamics (MD) methods, a series of systematic studies of mammalian, plant and invertebrate glycosylation patterns, in order to characterize the intrinsic 3D architecture of different sets of commonly found and synthetic (non-natural) glycan structures. From these results, we were able to disentangle the complexity of N-glycans structure and dynamics through a new 3D representation, which describes N-glycans not only in terms of the monosaccharides sequence, but that also includes anomeric configurations and linkage specificity. Within this framework, we defined N-glycans as structured by specific groups of monosaccharide units, named “glycoblocks”. This formulation incorporates 3D structural information and uniquely dictates the overall conformational landscape of any given N-glycan.

With this expanded viewpoint of sequence-to-structure dependencies in complex N-glycans, we applied this knowledge to glycoproteins, where variation of glycan composition affects its functional capabilities. In the two cases presented in this thesis, we determined how changes in the sequence of the N-glycans in the Fc region of IgG1 antibodies affect its effector function, and discovered for the first time a unique functional role of the glycan shield in the SARS-CoV-2 spike protein. In both cases, we observed that the conformational equilibria of complex N-glycans change to promote conformers that can accommodate interactions with the glycoprotein environment, but this adaption does not interfere with the intrinsic 3D glycan architecture, shifting a paradigm commonly assumed in structural biology, where the protein dictates the glycan conformation by actively morphing it.

The work presented in this thesis shows an alternative atomistic perspective of N-glycans structure and dynamics, where glycans play a starring role rather than a cameo as a simple protein “decoration”, while the knowledge and insight gained could inform the *ad-hoc* design and modulation of sequence-to-structure-to-function relationships of complex N-glycans, with applications in glycoengineering and therapeutic and diagnostic strategies.

## List of Publications

- i. E. C. Thomson, L. E. Rosen, J. G. Shepherd, R. Spreafico, A. da Silva Filipe, J. A. Wojcechowskyj, C. Davis, L. Piccoli, D. J. Pascall, J. Dillen, S. Lytras, N. Czudnochowski, R. Shah, M. Meury, N. Jesudason, A. De Marco, K. Li, J. Bassi, A. O'Toole, D. Pinto, R. M. Colquhoun, K. Culap, B. Jackson, F. Zatta, A. Rambaut, S. Jaconi, V. B. Sreenu, J. Nix, I. Zhang, R. F. Jarrett, W. G. Glass, M. Beltramello, K. Nomikou, M. Pizzuto, L. Tong, E. Cameroni, T. I. Croll, N. Johnson, J. Di Iulio, A. Wickenhagen, A. Ceschi, **A. M. Harbison**, D. Mair, P. Ferrari, K. Smollett, F. Sallusto, S. Carmichael, C. Garzoni, J. Nichols, M. Galli, J. Hughes, A. Riva, A. Ho, M. Schiuma, M. G. Semple, P. J. M. Openshaw, E. Fadda, J. K. Baillie, J. D. Chodera, S. J. Rihn, S. J. Lycett, H. W. Virgin, A. Telenti, D. Corti, D. L. Robertson, G. Snell, Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* (2021), doi:10.1016/j.cell.2021.01.037
- ii. L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, **A. M. Harbison**, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, R. E. Amaro, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).
- iii. C. A. Fogarty\*, **A. M. Harbison\***, A. R. Dugdale, E. Fadda, How and why plants and human N-glycans are different: Insight from molecular dynamics into the 'glycoblocks' architecture of complex carbohydrates. *Beilstein J. Org. Chem.* **16**, 2046–2056 (2020)  
[\*Shared first authorship]
- iv. **A.M. Harbison**, E. Fadda, An atomistic perspective on antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics. *Glycobiology.* **30**, 407–414 (2020)
- v. **A.M. Harbison**, L. P. Brosnan, K. Fenlon, E. Fadda, Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: A molecular dynamics study. *Glycobiology.* **29**, 94–103 (2019).

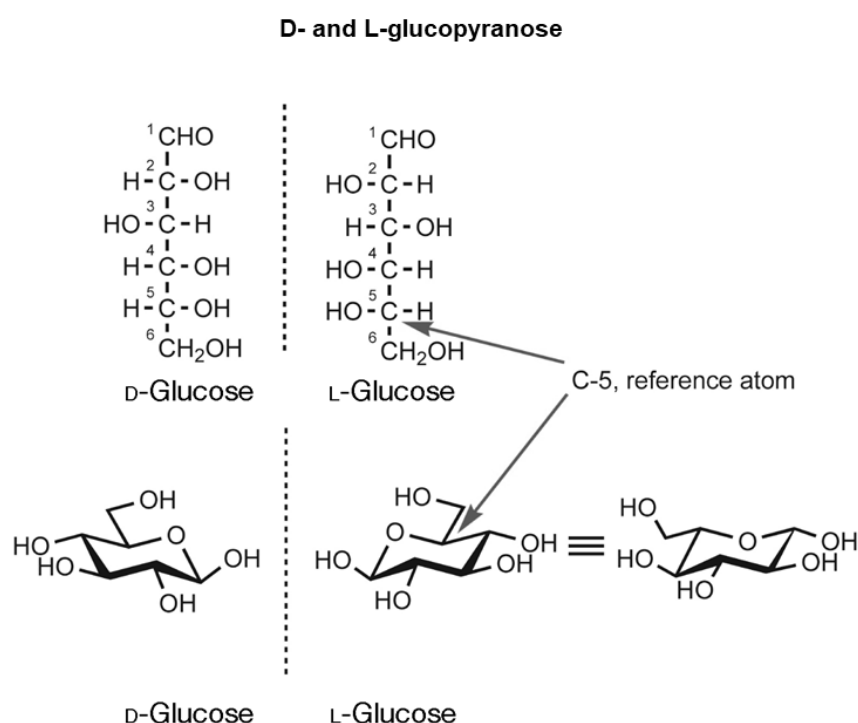
- vi. R. O’Flaherty, **A. M. Harbison**, P. J. Hanley, C. H. Taron, E. Fadda, P. M. Rudd, Aminoquinoline Fluorescent Labels Obstruct Efficient Removal of N-Glycan Core  $\alpha(1-6)$  Fucose by Bovine Kidney  $\alpha$ -1-Fucosidase (BKF). *J. Proteome Res.* **16**, 4237–4243 (2017).

**Under peer-review:**

- i. M. I. Zimmerman, J. R. Porter, M. D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C. E. Kuhn, J. H. Borowsky, R. P. Wiewiora, M. F. D. Hurley, **A. M. Harbison**, C. A. Fogarty, J. E. Coffland, E. Fadda, V. A. Voelz, J. D. Chodera, G. R. Bowman, *bioRxiv*, in press, doi:10.1101/2020.06.27.175430.

## Chapter 1: Introduction

Complex carbohydrates, also known as glycans, are the most abundant biomolecules in nature, and highly diverse, given to the variety of possible sequences, branching patterns and 3D structures. This high degree of polymorphism allows glycans to facilitate and regulate a wide range of biological functions in the myriad of cellular pathways they are involved in. Monosaccharides are the basic units of a glycan structure. These are defined by the general formula,  $(\text{CH}_2\text{O})_n$  and can be classified as polyhydroxyl aldehydes or ketones, see **Figure 1.1**. A monosaccharide's configuration is defined by the orientation of the furthest carbon centre from the carbonyl C1, also known as the anomeric carbon, and is denoted as D- or L-, see **Figure 1.1**.

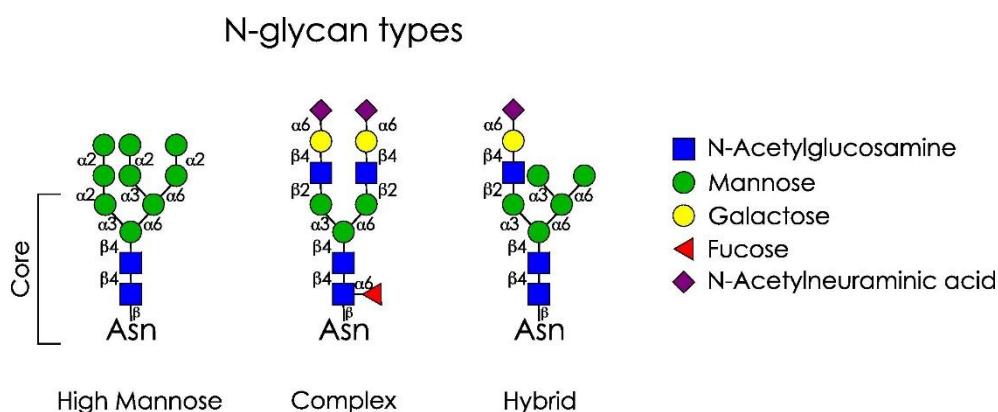


**Figure 1.1.** D- and L- configurations of glucose in both, its acyclic form, shown as a Fischer projection above, and in its cyclic form, in the chair conformation, Image adapted from Bertozzi *et al.*<sup>1</sup>.

In solution, monosaccharides are found in their more stable cyclic forms counting five or six membered rings, see **Figure 1.1**. The most stable ring conformation is known as “chair”, which can usually interconvert between two accessible forms, namely  ${}^4\text{C}_1$  to  ${}^1\text{C}_4$ , through a mechanism known as ring puckering. Puckering determines the change of the orientation of the hydroxy groups in the plane of the ring, i.e. from equatorial to axial, or vice versa. Cyclization produces an additional asymmetric stereocentre at the C1 anomeric carbon, which can be ( $\alpha$ -) or ( $\beta$ -) with reference to the orientation of the hydroxyl group on the stereocentre that determines the sugar's absolute configuration. More specifically, if in a Fisher projection the two hydroxy groups are on the same side, then the configuration

is ( $\alpha$ -), if not it is ( $\beta$ -). Monosaccharides can differ based on the stereochemistry at any one of its chiral carbon atoms in the chain, with structures differing by one chiral centre referred to as epimers, or by the functionalization of one or more hydroxy groups<sup>2</sup>.

Glycosidic bonds, which link monosaccharides together, are formed by a condensation reaction involving the hydroxy group at the C1 of one monosaccharide and any one of the hydroxy groups of another monosaccharide. Contrary to proteins or nucleic acids, where the combination of two amino acids or of two nucleotides has only two possible outcomes, taking into account the two anomeric configurations, the combination of two different hexoses, e.g. glucose and galactose, determines a theoretical 16 possible disaccharides. The progressive increase of the number of monosaccharides in an oligo- or polysaccharide, determines a combinatorial explosion of possible regio- and stereoisomers, to reach an enormous number of distinct glycans<sup>3</sup>. However, fortunately, glycan sequences are not found to be random in biological systems, suggesting that sequence and branching play a specific role in glycans' functions in biology. This thesis summarizes my four year-long research effort in studying the molecular basis for sequence-to-structure-to-function relationships in complex carbohydrates, more specifically in N-glycans, by means of high-performance computing (HPC)-based molecular simulations.



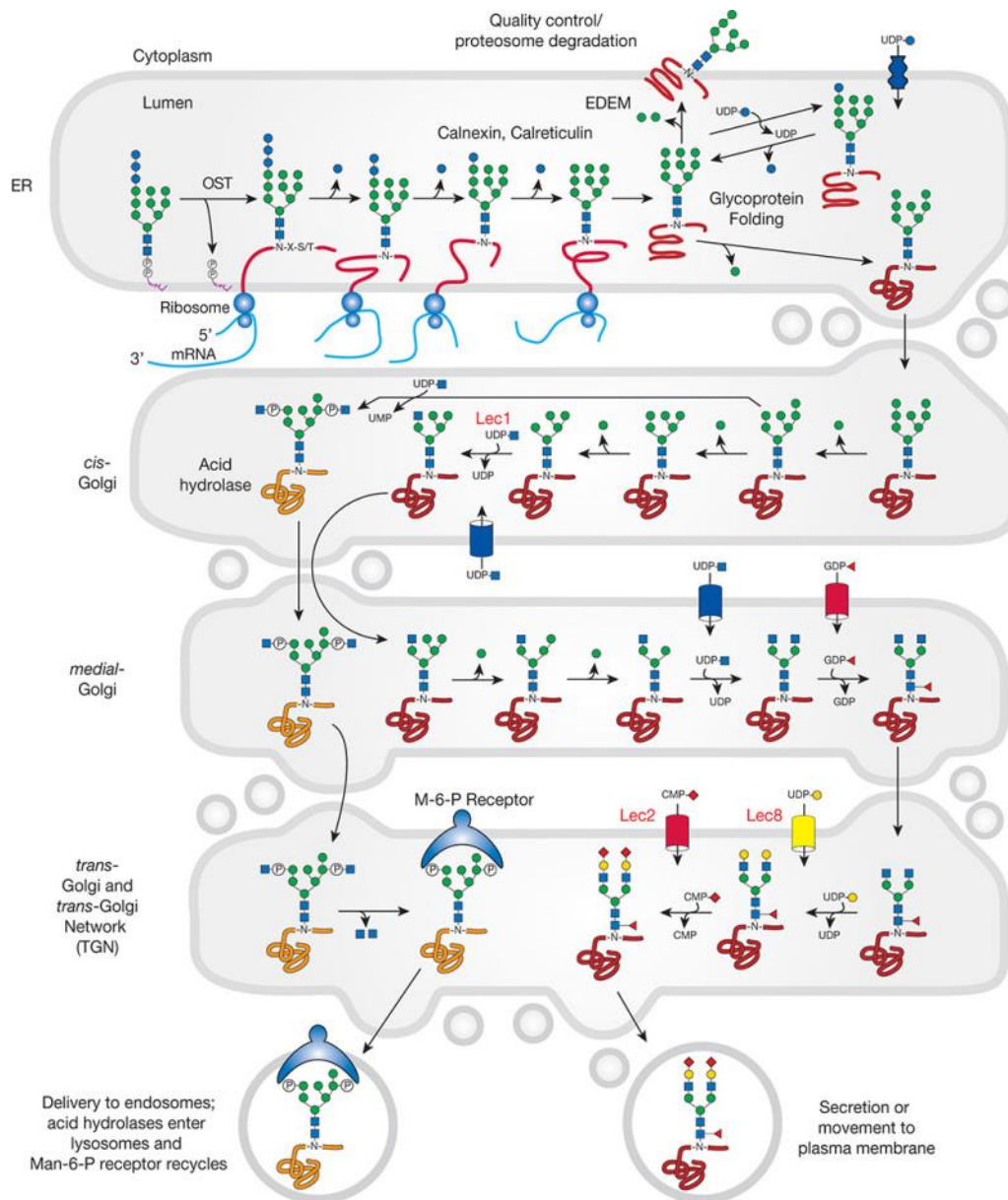
**Figure 1.2.** Three types of N-glycans, with the common core of  $\text{Man}\alpha 1-3(\text{Man}\alpha 1-6)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1$  linked to Asn from the sequon NX(S/T). The three sub-categories are named based on the glycosylation sequence of their additional antennae. Adapted from Higel *et al.*<sup>4</sup>

The types of glycosylation found in biological systems vastly differ between eukaryotes and bacteria<sup>5</sup>, with eukaryotic glycans classified based on their core glycan sequence and on the amino acid in a glycoprotein that they are covalently linked to. In an N-glycan, the carbohydrate part (or glycone) is covalently linked through a glycosidic bond to the amide nitrogen of an Asn side chain (aglycone), typically within the sequon N-X-(S/T), where X is any amino acid but Pro. The “common core” sequence of an N-glycan consists of the pentasaccharide  $\text{Man}\alpha 1-3(\text{Man}\alpha 1-6)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}\beta 1-\text{Asn}$ , shown in **Figure 1.2**. N-glycans can be further diversified in function of their



terminal, branching residues, and classified into three types, namely oligomannose (or high mannose), complex and hybrid, based on the residue types added to the core and to the “arms”, see **Figure 1.2**.

Unlike proteins, glycans' synthesis is not template driven but it depends on the nature, existence and expression levels of over 600 enzymes, between transferases and hydrolases, in mammalian cells<sup>6</sup>. This feature confers glycans the ability to retain a remarkable evolutionary plasticity, that allows them to modulate glycoproteins' activity, reflecting the physiological state of the cell<sup>7</sup>. Glycosylation of proteins is a post-translational modification, involving sequential concerted steps, shown in **Figure 1.3**, in the endoplasmic reticulum (ER) and the Golgi apparatus of the cell, resulting in the glycosylation of over 85% of secreted proteins<sup>8,9</sup>.



**Figure 1.3. Biosynthetic pathways of N-glycans in the cell.** Addition of glycan to the protein glycosylation site, and following processing and maturation as it is transported through the ER and Golgi apparatus. Adapted from Essentials of Glycobiology<sup>10</sup>.

When a nascent protein is translocated into the ER, a 14 residues glycan ( $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ ) assembled on the lipid carrier Dol-P (Dolichol phosphate) is transferred by the oligosaccharyltransferase (OST) to Asn residues that are part of N-X-S/T sequons<sup>11</sup>. The glycan is sequentially trimmed by ER  $\alpha$ -glucosidases and  $\alpha$ -mannosidases, and specific features of the trimmed glycan are recognised by ER chaperones to regulate protein folding<sup>12</sup>. Further trimming is performed in the *cis* Golgi, see **Figure 1.3**, where high mannose glycans, namely Man9 to Man7 can be trimmed down to a Man5 ( $\text{Man}_5\text{GlcNAc}_2$ ). Some oligomannose N-glycans avoid further modifications, while others undergo maturation in the *medial* and *trans* Golgi<sup>13</sup>. This step is initiated by the transfer of N-acetylglucosylamine (GlcNAc) to

both antennae by the GlcNAc-T1 enzyme, with consequent removal of terminal mannose residues from the Man5. Notably, hybrid-type N-glycans, see **Figure 1.3**, are exposed to the activity of GlcNAc-T1, but only on one arm, so the other arm retains its mannose residues<sup>14</sup>. Levels of core-fucosylation, further branching, galactosylation and terminal sialylation are dependent on the activity of the relevant transferases, and it is also dependent by the physical accessibility of the N-glycosylation site<sup>15</sup>. The latter is uniquely determined by the surrounding protein's landscape, which has reached folding maturation in the later Golgi<sup>16</sup>. Notably, the size of the glycoprotein and the location of N-glycan sites near  $\beta$ -turns is major factor in determining site-specific accessibility<sup>15</sup>. The products of this complex biosynthetic process is heterogeneous on a macro and micro scale, as glycosylation in the same site can vastly differ<sup>17,18</sup>. In order to understand the functional role of site-specific N-glycans in glycoproteins, one must first accurately identify the highest populated N-glycan sequences and the degree of heterogeneity at a given N-glycosylation site.

Despite the great advances in experimental techniques to determine glycans' sequences in biological systems<sup>19</sup>, there are still serious technological limitations hindering the rapid advancement of glycomics. The most commonly used technique for glycan characterisation is mass spectrometry (MS) with fluorescent labelling, usually preceded by separation chromatography<sup>20</sup>. Glycans must first be selectively cleaved from the protein surface with endoglycosidases<sup>21-23</sup>, producing intact N-glycans, which are further degraded by exoglycosidase digestion, taking advantage of enzymatic selectivity to identify specific anomericity, linkage type, and monosaccharide type in sequence<sup>24</sup>. It is important to note that enzymatic digestion is susceptible to changes in the 3D structure of the glycans, which can arise from modification of the structure upon addition of synthetic labels necessary for analysis<sup>25</sup>. Additionally, these methods can indicate the type of N-glycan, but are unable to provide any information on its 3D structure and cannot distinguish between positional isomers.

Similarly, because of their high flexibility and dynamic nature, the structural characterization of glycans is rarely straightforward if at all possible. Indeed, the rotation about the glycosidic torsion angle, namely phi ( $\phi$ ), psi ( $\psi$ ), and in 1-6 linkages, omega ( $\omega$ ), can generate a large set of distinct 3D conformations that are relatively free to interconvert at room temperature, giving rise to complex conformational ensembles. Thus, from a structural biology point of view, glycans appear to be "unstructured" or "intrinsically disordered". Because of all these difficulties inherent with glycoanalytics and with glycans' structure characterization, the field of glycomics has and is lagging considerably behind the fields of proteomics and genomics<sup>26</sup>. The two major tools for structural characterization of complex carbohydrates at the atomistic level of detail are NMR spectroscopy<sup>27,28</sup> and X-ray crystallography<sup>29</sup>. NMR takes advantage of naturally abundant isotopes (<sup>1</sup>H and <sup>13</sup>C), and signals are usually enhanced by isotopic enrichment<sup>30</sup>. However, this enrichment technique is laborious and not atom-specific, making it difficult to identify local structural features of the populated conformers<sup>26</sup>. Crystallographic studies can provide a great deal of structural information on glycans, such as bond lengths, bond angles and the

3D occupancy of the glycan in space. Yet only small and relatively rigid carbohydrate species can be crystallized with ions in solution<sup>31</sup>. For the systems we are interested in, namely highly glycosylated glycoproteins, crystallization is often not feasible as larger glycans retain a high degree of flexibility even in cryogenic environments relative to the folded protein they are linked to. If any electron density corresponding to the glycan is obtained, and the glycan sequence at that site is known, a glycan structure can be reconstructed by fitting to the electron density map. Yet, because of the broad features of these maps and the number of accessible degrees of freedom, it is a common occurrence to find in the PDB incorrect glycan structures due to overfitting of interpretive errors<sup>32</sup>.

Thanks to the rapid development of high performance computing (HPC) in the last two decades, large scale computational studies of biomolecular systems have become more feasible and accessible to a wider cohort of researchers<sup>33,34</sup>, supporting experimental analysis and as an independent method of investigation in its own right. Molecular simulations allow us to observe the dynamics and interactions of a system at the appropriate timescales and at an atomistic level of detail. Furthermore, in a virtual setting, we can overcome some of the limitations of experimental synthesis or design system modifications experimentally difficult to achieve, with little difference in computational cost. In the context of glycoscience, molecular simulations allow us to study the conformational propensity of specific glycan sequences and understand the intrinsic dynamics and architecture that dictates their overall structure and subsequent function when applied to larger biomolecules. However, much like the wider field of glycomics, progress in computational glycoscience has been slow to develop, with the majority of glycan-specific force fields created after 1990, and have been improved upon in the last twenty years<sup>35</sup>. The development of force fields to represent glycan dynamics has the added difficulty of capturing the flexibility of these biomolecules, and the branched nature of the carbohydrate structures, relative to proteins and other biomolecules. In the beginning, computational studies were limited to monosaccharide and disaccharide simulations at very low timescales<sup>36,37</sup>, but this has expanded with the advancement and widened accessibility to HPC, allowing for microseconds of large glycoprotein simulation. An in-depth discussion of the theory and details on the set-up of molecular simulation by molecular dynamics is presented in **Chapter 2**.

Previous computational research into isolated glycan dynamics and conformational analysis has been conducted on a wide range of small disaccharides and trisaccharides<sup>38-44</sup>, and some select N-glycan sequences<sup>45-48</sup>. While these studies have investigated the conformational propensity of the simulated glycans, none have systematically broken down how the presence of each monosaccharide in the sequence affects the overall conformational propensity, with Nishima *et al*<sup>46</sup> to be the only other group commenting on the change in populated conformers with monosaccharide alteration, using MD simulation. My work has further explored this concept, with exhaustive sampling of isolated N-glycans with a systematic approach of adding onto the glycan sequence, in order to ascertain the definitive effect of each alteration in monosaccharide addition and branching. Throughout my PhD I have studied the

sequence-to-structure relationships in complex N-glycans, an indissoluble connection widely accepted for proteins, but not for other biomolecules and in particular not for carbohydrates. Starting with isolated N-glycans in solution, we were able to identify the intrinsic interactions between monosaccharide residues, as well as the significant conformational changes dependent on specific monosaccharide additions, detailed in **Chapter 3** for mammalian N-glycans and **Chapter 4** for plant and for (some) invertebrates N-glycans. By expanding the concept of “sequence” from single monosaccharides to specific groups of monosaccharides, we are able to describe the 3D structural ensemble and preferred conformations of distinct motifs. This alternative description of glycans’ architecture with grouped glycan residues, or “glycoblocks”, can explain more clearly the glycotopes’ exposure and/or presentation, their activity within a complex glycoprotein and their recognition. This concept highlights the influence that the local spatial environment has on the conformational propensity of these highly flexible biomolecules.

The new perspective gained from our investigation into isolated N-glycan dynamics and its sequence-to-structure relationship was applied to glycoproteins of interest. The functional role of N-glycans within a glycoprotein environment has been studied more widely with simulation<sup>49–54</sup>, as the presence of N-glycans and their effect on protein interaction and structure is undeniable. My interests into glycoprotein simulation centred on how the glycan structures adapt and functionally complement glycoproteins environments, which I explored in **Chapters 5** and **6**. More specifically, we explored how the glycans in glycoproteins are fundamental, from the initial glycosylation in the ER that guides protein folding<sup>12,16</sup>, to their active roles in recognition and in the mediation of intrinsic and extrinsic biomolecular interactions<sup>55</sup>, and even to camouflage viral epitopes from the host’s immune system<sup>56,57</sup>. We also found evidence that the interaction with the protein and binding events can shift the glycans’ intrinsic conformational equilibria observed in isolated/unlinked glycans, yet not actively shape glycans into unobserved conformations. More specifically, in our studies the proteins are never found to alter the inherent conformation of the linked or bound glycans, which depends exclusively on their sequence and branching, but only shift their inherent conformational equilibrium<sup>58</sup>.

In **Chapter 5**, we examined the role that different glycan sequences and structures play in the dynamics of the Fc region of IgG1 antibodies. Our results provided a greater perspective of the role that core-fucosylation and sialylation (in the context of core-fucosylation) play in reducing antibody-dependent cell cytotoxicity (ADCC), by affecting Fc dynamics and carbohydrate interactions within the Fc region, as well as impeding access of the Fc  $\gamma$  receptors, potentially affecting the binding free energy between the receptors and the antibodies. Our understanding of the role of glycans in glycoproteins was further expanded when our group participated in the global effort to discover the mechanical function(s) of the SARS-CoV-2 spike protein, a large (180 kDa) type I fusion glycoprotein. As we discuss in detail in **Chapter 6**, through multi-microsecond molecular dynamics simulations, we were able to identify key glycosylation sites that play a functional role in the infection mechanism, unique to SARS-CoV,

mediating and supporting the opening mechanism of the spike's receptor-binding domain (RBD). By variation of the type of glycosylation and the extent of processing at these specific sites, we see a modulation of the structural support that the different glycoforms can provide, potentially weakening or strengthening the binding affinity of the spike's RBD to the ACE2 primary host cell receptor.

Based on the results and discussion presented from Chapters 3 to 6, I will conclude with some perspective for future work.

## References

1. Bertozzi, C. . & Rabuka, D. Structural Basis of Glycan Diversity. in *Essentials of Glycobiology, 3rd Edition* (eds. Varki, A., Cummings, R. D. & Esko, J. D.) (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press 2015-2017, 2009).
2. Witczak, Z. J. Monosaccharides: Occurrence, Significance, and Properties BT - Glycoscience: Chemistry and Chemical Biology. in (eds. Fraser-Reid, B. O., Tatsuta, K. & Thiem, J.) 815–840 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-30429-6\_18.
3. Laine, R. A. Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields  $1.05 \times 10^{12}$  structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759–767 (1994).
4. Higel, F., Seidl, A., Sörgel, F. & Friess, W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **100**, 94–100 (2016).
5. Dell, A., Galadari, A., Sastre, F. & Hitchen, P. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *Int. J. Microbiol.* **2010**, 148178 (2010).
6. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).
7. Rudd, P. M. & Dwek, R. A. Glycosylation: Heterogeneity and the 3D Structure of Proteins. *Crit. Rev. Biochem. Mol. Biol.* **32**, 1–100 (1997).
8. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
9. Zielinska, D. F., Gnad, F., Wiśniewski, J. R. & Mann, M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **141**, 897–907 (2010).
10. Varki, A. *et al.* Essentials of Glycobiology. in (2015). doi:NBK453020.
11. Silberstein, S. & Gilmore, R. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **10**, 849–858 (1996).

12. Caramelo, J. J. & Parodi, A. J. A sweet code for glycoprotein folding. *FEBS Lett.* **589**, 3379–3387 (2015).
13. Helenius, A. & Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
14. Kornfeld, R. & Kornfeld, S. ASSEMBLY OF ASPARAGINE-LINKED OLIGOSACCHARIDES. *Annu. Rev. Biochem.* **54**, 631–664 (1985).
15. Thaysen-Andersen, M. & Packer, N. H. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* **22**, 1440–1452 (2012).
16. Parodi, A. J. Protein Glucosylation and Its Role in Protein Folding. *Annu. Rev. Biochem.* **69**, 69–93 (2000).
17. Mathew, C. *et al.* Glycan-Protein Interactions Determine Kinetics of <math>\text{N}</math>-Glycan Remodeling. *bioRxiv* 2020.12.01.406371 (2020) doi:10.1101/2020.12.01.406371.
18. Schachter, H. Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochem. Cell Biol.* **64**, 163–181 (1986).
19. Mariño, K., Bones, J., Kattla, J. J. & Rudd, P. M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **6**, 713–723 (2010).
20. Stöckmann, H., O’Flaherty, R., Adamczyk, B., Saldova, R. & Rudd, P. M. Automated, high-throughput serum glycoprofiling platform. *Integr. Biol.* **7**, 1026–1032 (2015).
21. MIZUOCHI, T., AMANO, J. & KOBATA, A. New Evidence of the Substrate Specificity of Endo- $\beta$ -N-Acetylglucosaminidase D1. *J. Biochem.* **95**, 1209–1213 (1984).
22. Maley, F., Trimble, R. B., Tarentino, A. L. & Plummer, T. H. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem.* **180**, 195–204 (1989).
23. Collin, M. & Olsén, A. EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. *EMBO J.* **20**, 3046–3055 (2001).
24. O’Flaherty, R., Trbojević-Akmačić, I., Greville, G., Rudd, P. M. & Lauc, G. The sweet spot for biologics: recent advances in characterization of biotherapeutic glycoproteins. *Expert Rev. Proteomics* **15**, 13–29 (2018).
25. O’Flaherty, R. *et al.* Aminoquinoline Fluorescent Labels Obstruct Efficient Removal of N-



- Glycan Core  $\alpha(1-6)$  Fucose by Bovine Kidney  $\alpha$ -L-Fucosidase (BKF). *J. Proteome Res.* **16**, 4237–4243 (2017).
26. Council, N. R. *Transforming Glycoscience: A Roadmap for the Future*. (The National Academies Press, 2012). doi:10.17226/13446.
  27. Battistel, M. D., Azurmendi, H. F., Yu, B. & Freedberg, D. I. NMR of glycans: Shedding new light on old problems. *Prog. Nucl. Magn. Reson. Spectrosc.* **79**, 48–68 (2014).
  28. Gimeno, A., Valverde, P., Ardá, A. & Jiménez-Barbero, J. Glycan structures and their interactions with proteins. A NMR view. *Curr. Opin. Struct. Biol.* **62**, 22–30 (2020).
  29. Nagae, M. & Yamaguchi, Y. Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* **13**, 8398–8429 (2012).
  30. Yu, F. & Prestegard, J. H. Structural Monitoring of Oligosaccharides through  $^{13}\text{C}$  Enrichment and NMR Observation of Acetyl Groups. *Biophys. J.* **91**, 1952–1959 (2006).
  31. Jeffrey, G. A. & Sundaralingam, M. Bibliography of Crystal Structures of Carbohydrates, Nucleosides, and Nucleotides\*\*Work supported by NIH Grants GM-17378 and GM-24526 and the College of Agricultural and Life Sciences, University of Wisconsin, Madison. The authors express their gratitude . in (eds. Tipson, R. S. & Horton, D. B. T. -A. in C. C. and B.) vol. 43 203–421 (Academic Press, 1985).
  32. Agirre, J. Strategies for carbohydrate model building, refinement and validation. *Acta Crystallogr. Sect. D, Struct. Biol.* **73**, 171–186 (2017).
  33. Bolnykh, V., Rothlisberger, U. & Carloni, P. Biomolecular Simulation: A Perspective from High Performance Computing. *Isr. J. Chem.* **60**, 694–704 (2020).
  34. Löffler, H. & Winn. Large biomolecular simulation on HPC Platforms I. Experiences with AMBER, Gromacs and NAMD. in (2009).
  35. Fadda, E. & Woods, R. J. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today* **15**, 596–609 (2010).
  36. Brady, J. W. Molecular dynamics simulations of carbohydrate molecules. *Adv. Biophys. Chem.* **1**, 155–202 (1990).
  37. Brady, J. W. Molecular dynamics simulations of  $\alpha$ -D-glucose. *J. Am. Chem. Soc.* **108**, 8153–8160 (1986).
  38. Sayers, E. W. & Prestegard, J. H. Solution Conformations of a Trimannoside from Nuclear Magnetic Resonance and Molecular Dynamics Simulations. *Biophys. J.* **79**, 3313–3329 (2000).

39. Almond, A. Towards understanding the interaction between oligosaccharides and water molecules. *Carbohydr. Res.* **340**, 907–920 (2005).
40. Landström, J. & Widmalm, G. Glycan flexibility: insights into nanosecond dynamics from a microsecond molecular dynamics simulation explaining an unusual nuclear Overhauser effect. *Carbohydr. Res.* **345**, 330–333 (2010).
41. Perić-Hassler, L., Hansen, H. S., Baron, R. & Hünenberger, P. H. Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling. *Carbohydr. Res.* **345**, 1781–1801 (2010).
42. Kim, H.-M., 최영진, 이종현 TA - Lee, J.-H., 정갑주 TA - Jeong, K. & 정선호. Conformational Analysis of Trimannoside and Bisected Trimannoside Using Aqueous Molecular Dynamics Simulations. *Bull. Korean Chem. Soc.* **30**, 2723–2728 (2009).
43. Kirschner, K. N. & Woods, R. J. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci.* **98**, 10541 LP – 10545 (2001).
44. Fernandes, C. L., Sachett, L. G., Pol-Fachin, L. & Verli, H. GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins. *Carbohydr. Res.* **345**, 663–671 (2010).
45. Yang, M., Huang, J. & MacKerell, A. D. Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *J. Chem. Theory Comput.* **11**, 2855–2867 (2015).
46. Nishima, W., Miyashita, N., Yamaguchi, Y., Sugita, Y. & Re, S. Effect of Bisecting GlcNAc and Core Fucosylation on Conformational Properties of Biantennary Complex-Type N-Glycans in Solution. *J. Phys. Chem. B* **116**, 8504–8512 (2012).
47. Re, S., Nishima, W., Miyashita, N. & Sugita, Y. Conformational flexibility of N-glycans in solution studied by REMD simulations. *Biophys. Rev.* **4**, 179–187 (2012).
48. André, S., Kožár, T., Kojima, S., Unverzagt, C. & Gabius, H.-J. From structural to functional glycomics: core substitutions as molecular switches for shape and lectin affinity of N-glycans. **390**, 557–565 (2009).
49. Woods, R. J. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem. Rev.* **118**, 8005–8024 (2018).
50. Gavrillov, Y., Shental-Bechor, D., Greenblatt, H. M. & Levy, Y. Glycosylation May Reduce Protein Thermodynamic Stability by Inducing a Conformational Distortion. *J. Phys. Chem.*

- Lett.* **6**, 3572–3577 (2015).
51. 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).
  52. Lu, D., Yang, C. & Liu, Z. How Hydrophobicity and the Glycosylation Site of Glycans Affect Protein Folding and Stability: A Molecular Dynamics Simulation. *J. Phys. Chem. B* **116**, 390–400 (2012).
  53. Jo, S., Lee, H. S., Skolnick, J. & Im, W. Restricted N-glycan Conformational Space in the PDB and Its Implication in Glycan Structure Modeling. *PLOS Comput. Biol.* **9**, e1002946 (2013).
  54. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* **5**, 8926 (2015).
  55. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
  56. Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta - Gen. Subj.* **1863**, 1480–1497 (2019).
  57. Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).
  58. Fogarty, C. A. & Fadda, E. The oligomannose N-glycans 3D architecture and its response to the FcγRIIIa structural landscape. *bioRxiv* 2021.01.11.426234 (2021)  
doi:10.1101/2021.01.11.426234.

## Chapter 2: Computational Method

This chapter details the different molecular simulation techniques I used in my work, starting by the description of the fundamental concepts and principles underlying classical mechanics-based models and molecular dynamic (MD) simulations. Following from this, I will present the set-up of an MD simulation and explain some of the factors that can affect it and the parameters involved. I will also discuss the cases in which enhanced sampling methods within a MD framework are useful, focusing on replica exchange molecular dynamics (REMD), an approach I used to study the structure and dynamics of glycosylated IgG1s, see **Chapter 5**.

### **Molecular Mechanics and Empirical Force Fields**

According to the fundamental principles of quantum mechanics (QM), all molecular properties of a system can be uniquely defined by the electronic wave function of the system, or by its electron density in density functional theory (DFT)<sup>1,2</sup>. The exact information needed from these calculations cannot be obtained exactly because of the mathematical formulation of these theories, whereby approximations are needed to find solutions. Different levels of theory can be applied to the molecular system in order to increase the accuracy. However, this comes at the cost of computational time and resources, as most of the *ab-initio* QM or DFT-based approaches do not scale linearly with the number of atoms (N) in the system. The most efficient scaling achieved with DFT schemes is  $N^3$ , which allow us to handle routinely systems up to a few hundred atoms<sup>2</sup>. Nevertheless, when working with biomolecular systems, such as glycoproteins, this approach becomes unfeasible. Additionally, most QM and DFT calculations are run in the gas phase (*in vacuo*) or with implicit solvent, which does not accurately represent the properties of molecules at physiological temperatures or in the presence of heterogeneous solvents. By implementing the Born-Oppenheimer (BO) approximation, we can consider the nuclear motions of a molecular system separately from the electronic motions, due to the difference in mass between these subatomic particles. With this approximation, we can use molecular mechanics to represent our system with Newtonian laws of motion, which scale linearly with the number of atoms, thus modelling the properties of the full system in a reasonable amount of time.

Within a classical molecular mechanics framework, atoms are assumed to be a hard, impenetrable spheres, with the bonds between atoms represented as springs. This allows the model to incorporate both the stability of a chemical structure, while allowing for flexibility of the covalent bonds in their vibrational degrees of freedom. In order for these molecular models to be realistic, their structural and interaction properties are represented by parameters fitted to experimental data and QM data as part of what we call an empirical forcefield. Empirical force fields define molecular systems by internal and external potential energy (PE) terms that are used to calculate the specific potential energy of the system and the atomic motions, in function of the position of its atoms. The internal (bonding/covalent) PE terms represent the bond length and angles that are approximated by Hooke's Law, while torsions that are treated as a sinusoidal function. External (non-bonding/non-covalent) terms are applied to atoms in a molecule more than four atoms apart, and include the Van der Waals attraction and repulsion interactions between atoms, usually modelled with a Lennard-Jones potential, and the electrostatic interactions represented by a Coulomb potential, see **Equation 2.1**.

$$\begin{aligned}
 V(r^N) = & \sum_{bonds} \frac{K_i}{2} (l_i - l_{i,o})^2 + \sum_{angles} \frac{K_i}{2} (\theta_i - \theta_{i,o})^2 \\
 & + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \\
 & + \sum_i^n \sum_{j=i+1}^n \left\{ 4\varepsilon_{i,j} \left( \left[ \frac{\sigma_{i,j}}{r_{i,j}} \right]^{12} - \left[ \frac{\sigma_{i,j}}{r_{i,j}} \right]^6 \right) + \frac{q_i q_j}{4\pi\varepsilon_o r_{i,j}} \right\}
 \end{aligned} \tag{2.1}$$

Force fields can be separated into additive and polarizable force fields, where the partial charges ( $q$ ) of an atom is either centred on the representative atom, or the partial charge is variable in order to account for the 3D occupancy of the electron density surrounding an atom, respectively.

Parameter sets are used in empirical force fields to minimise the number of unique parameters, making the calculation of these PE terms more efficient and manageable. This is achieved by assigning an "atom type" to atoms in similar chemical environments, which have a set of bonded and non-bonded parameters associated to them. This approach affects accuracy and determines the specificity of the force field to different biomolecules, i.e. carbohydrates, proteins, nucleic acids.

For proteins, there are many commonly used force fields developed within ‘families’, such as AMBER and CHARMM. Specific carbohydrate force fields have been developed so they can be used in conjunction with protein force fields in simulations, in order to give an accurate representation of the behaviour of glycoproteins. The most commonly used carbohydrate force fields are GLYCAM06<sup>3</sup> and CHARMM36<sup>4</sup>. GLYCAM06 is the first complete and stand-alone all-atom force field for carbohydrates and glycoproteins<sup>5</sup>. The GLYCAM06 parameter set was independently developed but in a way that was consistent with AMBER parameters, allowing it to be used in combination with other protein force fields. AMBER does however include specialised atom types for the Asn and Thr/Ser residues linked to glycans in glycoprotein structures, thus making GLYCAM06 and AMBER protein force fields suitable to be paired for the simulation of glycoproteins. Another feature of GLYCAM06 is the absence of different atom types for  $\alpha$  and  $\beta$  anomers, allowing the ring to “pucker”, i.e. to move through the different ring conformations accessible to different monosaccharides<sup>6</sup>. Note, puckering is an important feature to retain correctly as it is implicated in many oligosaccharide conformational changes<sup>7-10</sup> and in enzymatic reactions<sup>11</sup>. The covalent PE terms for GLYCAM06 were derived from gas-phase QM calculations on appropriate molecular fragments, with torsional energy barriers fitted to minimize the error throughout the rotational PE curve, ensuring the force field could reproduce the flexibility of oligo- and polysaccharides<sup>3</sup>. Electrostatic point charges were obtained from QM restrained electrostatic potential (RESP) calculations<sup>5</sup>. While 1-4 electrostatic interactions are scaled in the AMBER family of protein force fields, in GLYCAM06 they are not, due to the unique structural flexibility of the 1-6 linkage. Lennard-Jones potentials from the AMBER PARM94 were directly transferred to GLYCAM06 without modification.

The CHARMM36 additive parameter set for carbohydrates and glycoconjugates<sup>4,12-14</sup> was developed to be compatible with the CHARMM all-atom biomolecular force fields. Similar to GLYCAM06, parameters were derived for carbohydrate fragments and then applied to monosaccharides, with missing parameters developed and set adjusted to gas phase QM and condensed phase experimental data. These parameters were further refined with comparison to MD simulation of infinite crystals and of differently diluted solutions, aimed at reproducing crystal geometries of monosaccharides<sup>12,13</sup> and disaccharides<sup>4</sup>. Notably, CHARMM has the first polarizable carbohydrate force field based on the Drude oscillator model<sup>15</sup>, which adds a charge polarization effect to the existing atom point charges in the CHARMM empirical force field.

In addition to parameter sets that describe protein and carbohydrate atoms, the quality of a simulated system greatly depends on the choice of water force field, if the system is solvated. Ideally, the chosen water model should match the water model used in the parameterization and validation of the protein force field itself. TIP3P<sup>16</sup>, the three-point water model, was used in development of AMBER and GLYCAM, making it a clear candidate for simulation with AMBER and GLYCAM force fields.

## Molecular Dynamics

Molecular Dynamics (MD) is based on a classical mechanics approach to describe the distribution of atoms in space and how their position, through Newtonian motion, evolves over time. The work presented in this thesis was completed by using both conventional MD and an enhanced sampling technique, namely temperature replica exchange molecular dynamics (REMD)<sup>17</sup>.

Before starting a simulation, an initial starting conformation has to be selected, ideally from experimental sources, such as NMR, x-ray crystallography or cryogenic electron microscopy (cryo-EM). In terms of glycans, this can be an issue, as experimental techniques do not always resolve atoms of high flexibility and dynamics, or as in x-ray crystallography the atoms coordinates are obtained through fitting to an electron density map, sometimes without a proper consideration of the correct glycan structure. To supply for the absence of glycan structural information there are glycan building tools available, developed by GLYCAM and CHARMM teams, such as glycam-web (<https://dev.glycam.org/>) and CHARMM-GUI<sup>18</sup> (<http://www.charmm-gui.org/>). For the simulations of biomolecules, the biomolecular system needs to be solvated by water molecules, filling a simulation box within which counterions can be added to the desired concentration. Following this, energy minimisation and equilibration steps need to be run to bring the system to a thermodynamic equilibrated state where thermodynamics and structural properties can be sampled through the integration of the equation of motion as described below.

Empirical force fields provide the PE terms of the system as a function of the position of all its atoms, denoted by  $V(r)$ . The forces acting on the atoms are produced as described by **Equation 2.2**, which gives the gradient of the PES.

$$F = - \frac{\delta V(r)}{\delta r} \quad 2.2$$

Accelerations are calculated using Newton's second law of motion, shown in **Equation 2.3**,

$$F = ma \quad 2.3$$

where  $m$  is the atom mass. Newton's equation of motion are integrated numerically in order to determine the time-evolution (or trajectory) of positions and velocities of the atoms. A classic integrator, like the Verlet algorithm shown in **Equations 2.4 to 2.6**, takes the positions and accelerations at time  $t$ , with the previous positions (at time  $t - \delta t$ ) and calculates the new positions at time  $t + \delta t$ .

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad 2.4$$

$$r(t - \delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad 2.5$$

$$\therefore r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \quad 2.6$$

Another option is the leap frog integrator, that uses velocities calculated at  $t + 1/2\delta t$  to then calculate the accelerations at time  $t$ , while the positions are calculated at  $r(t + \delta t)$  using  $r(t)$  and  $v(t + 1/2\delta t)$  shown in **Equations 2.7 to 2.8**.

$$r(t + \delta t) = r(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t \quad 2.7$$

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t \quad 2.8$$

This overlapping "leap frog" generation of velocities and positions gives this integrator method an advantage over the Verlet integrator, as the velocities are explicitly calculated. Other integrators are available like the velocity Verlet method<sup>19</sup> or the Runge-Kutta method<sup>20</sup>. In my work I used a stochastic leap frog integrator when working with AMBER (**Chapters 3, 4 and 6**), and a symplectic Verlet integrator with NAMD (**Chapter 5**). Temperature and pressure are regulated with a thermostat and a barostat, respectively. Langevin dynamics with a collision

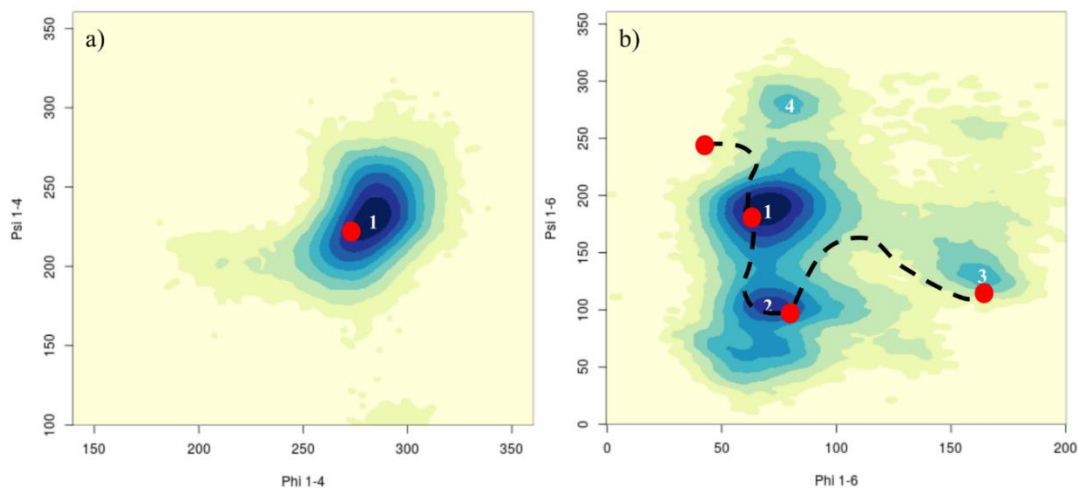


frequency of  $2.0 \text{ ps}^{-1}$  was used in all of my simulations for temperature control. Pressure was regulated with the Berendsen barostat<sup>21</sup> in AMBER, and with a modified Nosé-Hoover method<sup>22,23</sup> selected in NAMD.

A very important aspect of the set-up of MD simulations is the time-step,  $\delta t$ , as it determines how long the simulation will take to run depending on the computational resources available. The size of the time-step is defined by an order of magnitude shorter than the shortest vibration of the system; in most systems, this is usually the vibration of covalent hydrogen-carbon bonds, approximating to 10 fs, leading to a time step of 1 fs. However, most simulation software, there is an option to constrain the vibration of covalent hydrogen bonds allowing for a longer time step of 2 fs. Both NAMD and AMBER software packages implement the SHAKE algorithm<sup>24</sup> to constrain these bonds.

### **Conformational sampling**

Once the molecular simulation is set-up to run, the commonly asked question (or one that we should ask) is “how long of a trajectory is long enough?”. Indeed, the simulation length determines the reproducibility, thus the validity of the simulations results. Although the latter hinges on the choice of force field and methodology implemented, exhaustive (or sufficient) conformational sampling is key, though knowing when convergence will be reached is not intuitive, and can be unique for each chosen starting structure. In the case of complex carbohydrates, due to their high degree of flexibility, especially when unbound and isolated, complete sampling requires considerable computational resources for relatively small systems. Indeed, looking at the free energy associated to different conformations of two common linkages in complex carbohydrates, see **Figure 2.1**, we can illustrate the problem of insufficient sampling.



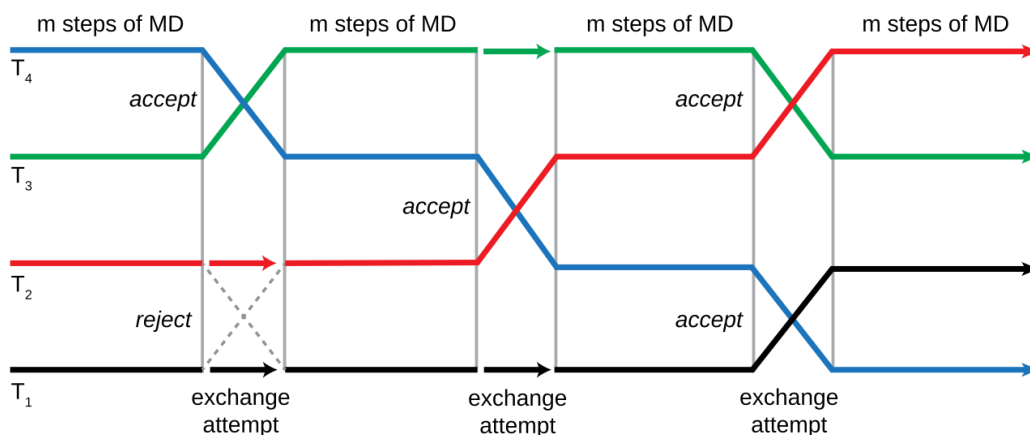
**Figure 1.1.** Conformational heat maps corresponding to **a)**  $\alpha(1-4)$  linkage and **b)**  $\alpha(1-6)$  linkage obtained for a N-glycan through complete conformational mapping performed with a series of conventional MD simulations run in parallel. The red dots and black dashed lines represent a hypothetical walk in time space obtained by MD simulation. The numbers correspond to the minima identified during complete sampling<sup>8</sup>, with the darkest shade of blue corresponding to the most stable structure. Maps rendered with RStudio (<https://rstudio.com/index2/>).

MD simulations with a conventional implementation explore the force-field defined PES through a continuous walk in time. For the  $\beta(1-4)$  linkage, simulation from one starting structure is adequate in estimating the flexibility and stability of the structure (from population frequency and standard deviation measurements). However, the  $\alpha(1-6)$  linkage has an extra torsion angle, providing the linkage with extra flexibility and therefore a vaster and more complicated conformational landscape. Starting from one structure (red point), the MD trajectory may identify some minima of the PES, but not all, with the inability to cross high energy barriers at 300 K and thus a longer simulation is needed for greater exploration. An alternative approach is to set up a sufficient number of uncorrelated MD runs from different starting structures, with shorter MD sampling time for each structure, but an estimation of convergence based on structural interconversion between separate trajectories. We implemented this approach in the work discussed in **Chapters 3** and **4**, successfully characterizing the conformational propensity and dynamics of free carbohydrates in solvated systems.

As an alternative that does not always require a previous knowledge of the conformational space, there are conformational sampling techniques available, referred to as “enhanced

sampling” approaches, that overcome the shortcomings of conventional deterministic sampling. Enhanced sampling schemes are especially useful in cases where conformational degrees of freedom are restrained and PES barriers may be too high to cross at room temperature. Indeed, in complex systems, such as glycosylated biomolecules, glycans can be restricted by interactions with residues on the protein’s surface and/or can be part of complex networks of carbohydrate-carbohydrate interactions that also confine their dynamics to a smaller subset of possible conformational occupancies<sup>25</sup>. The scenario is made even more complex when starting structures from experimental sources have glycans in low populated, unnaturally distorted or even in incorrect conformations<sup>26</sup>. Enhanced sampling methods in this case may provide a strategy to resolve these issues by providing the particular system with enough energy to escape particularly tricky energy wells and/or barriers. Stochastic methods such as metadynamics<sup>27</sup> (MTD) and umbrella sampling<sup>28</sup> apply biases to the PES, to lower the energy barriers, and thus allowing the system to traverse the PES more freely. Deterministic methods like temperature REMD<sup>29</sup> perform the simulation partially at higher temperatures within a defined temperature range, allowing easier traversal of the PES. Our specific implementation of temperature REMD is outlined in **Chapter 4** and the basic principles of the method are discussed below.

The REMD algorithm developed by Sugita and Okamoto<sup>17</sup> combines the parallel tempering method used in Monte Carlo simulations with an MD approach, and is widely used as an enhanced sampling technique. The process of exchange between replicas is represented in **Figure 2.2**.



**Figure 2.2.** Overview of temperature REMD, showing exchanges being attempted after every  $m$  simulation steps. Exchanges can only happen between adjacent temperatures<sup>30</sup>.

Using temperature as an exchange variable, multiple isothermal (conventional) MD simulations run in parallel, with a set number of replicas ( $n$ ) generated in given temperature range, and each assigned a simulation temperature in ascending order ( $T_0, T_1, \dots, T_{n-1}$ ). After  $m$  number of steps, the system will try to swap adjacent replicas  $i$  and  $j$ , with an acceptance ratio dependant on the energy between replicas shown in **Equation 2.9**.

$$\min \left\{ 1, e^{(E_i - E_j) \left( \frac{1}{kT_i} - \frac{1}{kT_j} \right)} \right\}$$

2.9

If the distribution of the differential temperatures is set, in the selected temperature range, this ensures a better exchange probability, as the acceptance ratio is the same between all pairs of replicas<sup>31,32</sup>. This can be achieved easily with tools developed for temperature REMD<sup>33</sup>, which gives a good indication of the viability of a simulation for temperature REMD based on the predicted exchange ratio. Hamiltonian REMD<sup>34</sup> can also be performed, where arbitrary perturbations to the Hamiltonian of the system is used rather than temperature as an exchange variable, and allows for smaller number of replicas with larger differences between individual replica Hamiltonians. The easy set up and implementation of REMD comparative to other enhanced sampling techniques is an advantage, but REMD is computationally very costly, given the large number of replicas needed, which is dependent on the size of the system, including explicit solvent<sup>35</sup>.

## References

1. van Mourik, T., Bühl, M. & Gaigeot, M.-P. Density functional theory across chemistry, physics and biology. *Philos. Trans. A. Math. Phys. Eng. Sci.* **372**, 20120488 (2014).
2. Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **136**, 150901 (2012).
3. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622–655 (2008).
4. Guvench, O. *et al.* CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput.* **7**, 3162–3180 (2011).
5. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622–655 (2008).
6. Mayes, H. B., Broadbelt, L. J. & Beckham, G. T. 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* **136**, 1008–1022 (2014).
7. Harbison, A. M., Brosnan, L. P., Fenlon, K. & Fadda, E. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94–103 (2019).
8. Fogarty, C. A., Harbison, A. M., Dugdale, A. R. & Fadda, E. How and why plants and human N-glycans are different: Insight from molecular dynamics into the ‘glycoblocks’ architecture of complex carbohydrates. *Beilstein J. Org. Chem.* **16**, 2046–2056 (2020).
9. Alibay, I., Burusco, K. K., Bruce, N. J. & Bryce, R. A. Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *J Phys Chem B* **122**, 2462–2474 (2018).
10. Topin, J. *et al.* The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *ACS Chem Biol* **11**, 2011–2020 (2016).
11. Ardèvol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **137**, 7528–7547 (2015).
12. Guvench, O. *et al.* Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564 (2008).

13. Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W. & Mackerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353–2370 (2009).
14. Mallajosyula, S. S., Jo, S., Im, W. & MacKerell Jr, A. D. Molecular dynamics simulations of glycoproteins using CHARMM. *Methods Mol. Biol.* **1273**, 407–429 (2015).
15. Lemkul, J. A., Huang, J., Roux, B. & MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **116**, 4983–5013 (2016).
16. Jorgensen, W. L. & Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **19**, 1179–1186 (1998).
17. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
18. 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).
19. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).
20. Janezic, D. & Orel, B. Implicit Runge-Kutta method for molecular dynamics integration. *J. Chem. Inf. Comput. Sci.* **33**, 252–257 (1993).
21. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
22. Posch, H. A., Hoover, W. G. & Vesely, F. J. Canonical dynamics of the Nosé oscillator: Stability, order, and chaos. *Phys. Rev. A* **33**, 4253–4265 (1986).
23. Hoover, W. G. & Holian, B. L. Kinetic moments method for the canonical ensemble distribution. *Phys. Lett. A* **211**, 253–257 (1996).
24. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
25. Fogarty, C. A. & Fadda, E. The oligomannose N-glycans 3D architecture and its response to the FcγRIIIa structural landscape. *bioRxiv* 2021.01.11.426234 (2021)  
doi:10.1101/2021.01.11.426234.

26. Ferrara, C. *et al.* Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fcγ<sub>3</sub> and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669–12674 (2011).
27. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).
28. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
29. Okamoto, Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graph. Model.* **22**, 425–439 (2004).
30. Rowley, C. Schematic of a replica exchange molecular dynamics simulation.
31. Ballard, A. J. & Jarzynski, C. Replica exchange with nonequilibrium switches. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12224–12229 (2009).
32. Rosta, E. & Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* **131**, 165102 (2009).
33. Patriksson, A. & van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **10**, 2073–2077 (2008).
34. Yang, M., Huang, J. & MacKerell, A. D. Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *J. Chem. Theory Comput.* **11**, 2855–2867 (2015).
35. Meli, M. & Colombo, G. A Hamiltonian replica exchange molecular dynamics (MD) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *Int. J. Mol. Sci.* **14**, 12157–12169 (2013).

## Chapter 3: Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study

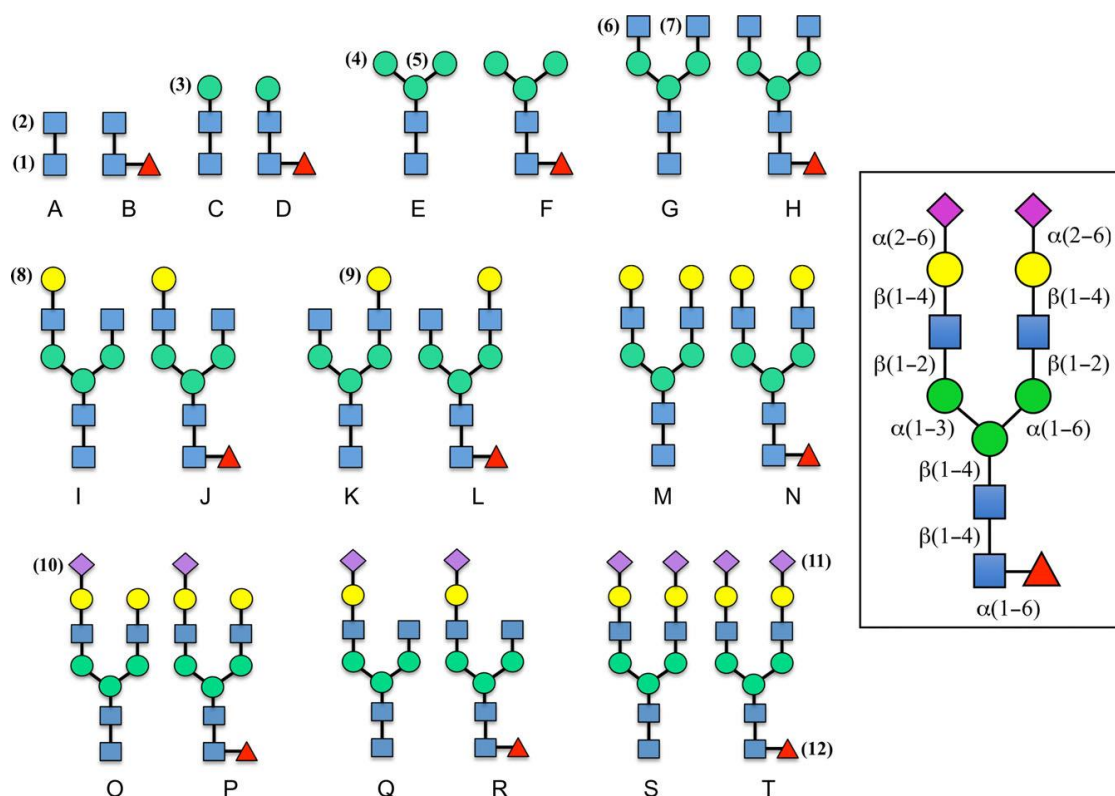
**Paper citation:** Harbison, A. M., Brosnan, L. P., Fenlon, K. & Fadda, E. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94–103 (2019).

### 3.1 Introduction

N-glycosylation of the immunoglobulin G (IgG) fragment crystallizable (Fc) region is essential for its structural stability and function<sup>1-4</sup>. The sequence and branching of the Fc N-glycoforms, bound at the highly conserved Asn 297 in both CH<sub>2</sub> domains of the Fc region, strongly affect the antibody-mediated effector function<sup>5-7</sup> by modulating the binding to the immune cells' Fc receptors, thus the antibody-mediated immune response<sup>8, 9</sup>. In this context the effects of core-fucosylation, sialylation and of galactosylation are particularly interesting. Between 81% and 98.7% of the Fc N-glycans in human IgGs are core-fucosylated<sup>10</sup>. Even though it may appear like a very subtle change to the glycan sequence, especially relative to the size of the whole IgG, core fucosylation, where fucose is  $\alpha(1-6)$  linked to the chitobiose core of the complex N-glycan, greatly affects the IgGs antibody-dependent cell-mediated cytotoxicity (ADCC) function. More specifically, a strongly enhanced ADCC corresponds to non-fucosylated Fc N-glycan species<sup>6, 11-17</sup>. This information has found wide interest and applications in cancer immunotherapy, especially in regards to engineering non-fucosylated antibodies with higher efficacy<sup>7, 13, 17-20</sup>. The molecular basis for this phenotype is not entirely clear. It has been linked to the stronger binding between IgG and the Fc  $\gamma$  receptor IIIa (Fc $\gamma$ RIIIa)<sup>7, 21, 22</sup>, determined by a more effective contact between the IgG and Fc $\gamma$ RIIIa glycans in the absence of core fucose<sup>21</sup>. No significant structural changes in the IgG structure have been detected in function of the presence or absence of fucose<sup>11</sup>. Sialylation of the Fc glycans is known to reverse the antibody inflammatory effect from pro to anti<sup>23</sup>. Only about 20% of core-fucosylated biantennary Fc N-glycans are sialylated<sup>10</sup>, meanwhile the majority of Fc N-glycans in human IgGs are galactosylated, with neutral glycans without galactose slightly below 40%, neutral glycans with one terminal galactose slightly above 40%, and neutral glycans with two terminal galactoses contributing to 20% of the neutral IgG glycome<sup>10</sup>. The abundance of galactosylated glycoforms has been directly linked to aging, with decreasing levels correlated with aging, and to immune activation<sup>10, 24-26</sup>. Notably, the risk of developing rheumatoid arthritis is correlated with low levels of galactosylation<sup>27-29</sup>.



The phenotypes linked to different Fc N-glycan sequences are likely to be determined by the modulation of the interaction of the IgG with cell surface receptors, which is a difficult topic to address as a whole due to the complexity of the systems involved. NMR spin relaxation data provide evidence that despite the close contact with the protein, both arms of the N-glycans at the IgG Fc remain flexible and accessible<sup>30</sup>, suggesting that the intrinsic conformational propensity of the Fc N-glycan in function of its sequence may play a role in regulating the molecular interaction with the cell surface receptors. Therefore, to understand the implications on the N-glycans structural and dynamics of their size and sequence, we conducted a complete conformational study by extensive molecular dynamics (MD) simulations of progressively long complex biantennary Fc N-glycans most commonly expressed in human IgGs<sup>10,31</sup>. All the glycoforms we have analysed in this work are shown in **Figure 3.1**.



**Figure 3.1.** Schematic representations of all N-glycans, fucosylated and non-fucosylated in pairs, analysed in this study. Letters are used as shorthand notation for the identification of each sugar. Numbers are used to identify residues. The graphical representation follows the guidelines indicated in (Varki, A., Cummings, R.D., et al. 2015).

Because of the complexity of the N-glycans dynamics and the high flexibility of  $\alpha(1/2-6)$  linkages, we chose a sampling method based on single (conventional) MD trajectories, ran in parallel, all started from different combinations of  $\alpha(1/2-6)$  conformers, namely 3 trajectories for one  $\alpha(1-6)$  linkage, 9 for two  $\alpha(1-6)$  linkages, 12 for one  $\alpha(1-6)$  and one  $\alpha(2-6)$  linkages, 24 for two  $\alpha(1-6)$  and one  $\alpha(2-6)$

linkages, and finally 72 for two  $\alpha(1-6)$  and two  $\alpha(2-6)$  linkages, for a total cumulative sampling time in excess of 62  $\mu\text{s}$ . This approach allowed us to systematically and directly sample by construction regions of the potential energy surface corresponding to rare conformations of the different  $\alpha(1/2-6)$  linkages and to assess their relative stability. Our results show that a) the highest conformational flexibility concerns primarily the  $\alpha(1-6)$  arm, while the  $\alpha(1-3)$  arm remains mostly extended, b) core fucosylation and sialylation do not affect the conformational equilibrium of the  $\alpha(1-6)$  arm in the isolated glycan in solution, meanwhile c) galactosylation of the  $\alpha(1-6)$  arm alone greatly shifts the conformational propensity of the arm from outstretched, to folded over the chitobiose core. These findings provide important insight into the differences in the molecular recognition of biantennary complex N-glycans by enzymes and lectins in function of their sequence. Implications in the molecular recognition of the different glycoforms studied here when isolated (unbound), Fc-linked, or on glycan arrays are discussed in the sections below.

### 3.2 Computational Methods

All glycans were built with the carbohydrate builder tool on GLYCAM-WEB (<http://www.glycam.org>). All combinations of rotamers for the  $\alpha(1/2-6)$  linkages have been considered as starting structures for the MD simulations, namely 3 trajectories for one  $\alpha(1-6)$  linkage, 9 for two  $\alpha(1-6)$  linkages, 12 for one  $\alpha(1-6)$  and one  $\alpha(2-6)$  linkages, 24 for two  $\alpha(1-6)$  and one  $\alpha(2-6)$  linkages, and finally 72 for two  $\alpha(1-6)$  and two  $\alpha(2-6)$  linkages. The GLYCAM-06h-12SB version of the GLYCAM06 force field<sup>32</sup> was used to represent the carbohydrate atoms, TIP3P parameters<sup>33</sup> were used to represent water, while amber99SB parameters<sup>34</sup> were used for the counterions, added in a number sufficient to neutralize the charge in the case of sialylated species. All simulations were run with versions 12 and 16 of the AMBER molecular simulation package<sup>35</sup>. Dispersion interactions were cutoff at a distance of 13 Å. Electrostatics interactions were treated with Particle Mesh Ewald (PME). A constant pressure of 1 atm was maintained by isotropic position scaling with a relaxation time of 2 ps, while a constant temperature of 300 K was regulated by Langevin dynamics using a collision frequency of 1.0  $\text{ps}^{-1}$ . The SHAKE algorithm was used to restrain bonds with hydrogen atoms and an integration time step of 2 fs was used throughout. For each sugar, each starting structure was analysed for at least 250 ns, with exceptions of sugar D, which was analysed for 3  $\mu\text{s}$ , because of the complexity of its dynamics, which will be discussed in the sections below, and sugar E for 500 ns on single trajectories. Further details on the minimization, equilibration and production phases are included in **Appendix I**. High performance computing (HPC) resources were provided by the Irish Centre for High-End Computing (ICHEC).

As an interesting note, because of the better scaling on our machines of v. 4.6.3 and 5.0.x of GROMACS (GMX) for the calculations on these relatively small systems, we ran some tests on the medium-sized sugar H, see **Figure 3.1**, and compared the results with the GLYCAM/AMBER set-up. The starting structure and parameter files obtained from the carbohydrate builder on GLYCAM-WEB were converted to GMX format with the AnteChamber PYthon Parser interface (*acpype.py*) tool<sup>36</sup>. It is important to note that in all GMX simulations the 1-4 scaling was re-set to “1” as required by the GLYCAM force field<sup>32, 37</sup>. Equal amount of sampling was done with both GMX and AMBER, preceded by a very similar set-up and equilibration protocols. Details of the GMX protocol are provided in **Appendix I**. The results indicate large differences in  $\alpha(1-6)$  torsions populations between GMX and GLYCAM/AMBER, shown in **Tables S.1** and **S.2** in **Appendix I**. The reason for this may be problems in the transfer of torsional parameters from a GLYCAM/AMBER format to the GMX format. More specifically we found that simulations of sugar H with GMX do not reproduce the correct conformer populations, or give energetically disfavoured conformers, such as the *tg* in the core fucose  $\alpha(1-6)$  linkage as the highest populated for sugar H2, see **Table S.2** in **Appendix I**.

### 3.3 Results

The results below are presented function of the different N-glycan linkages for clarity. Notable effects on the conformational propensity of different linkages determined by the N-glycan size and sequence are indicated within. All torsion angles discussed throughout correspond to the following nomenclature,  $\phi$  ( $O_5C_1O_xC_x$ ),  $\psi$  ( $C_1O_xC_xC_{x+1}$ ), and  $\omega$  ( $O_6C_6C_5C_4$ )<sup>38</sup>. The method we have chosen to number the monosaccharides and to name the different N-glycans is indicated in **Figure 3.1**. A summary of the results obtained as averages over all the simulations for all fucosylated and non-fucosylated species is shown in **as 3.1** and **3.2**, respectively. Results obtained for each system studied are shown in **Tables S.3-23** in **Appendix I**.

**Table 3.1** Conformational propensities of different linkages calculated for all core-fucosylated species shown in **Figure 3.1**. The torsion angle values are shown in degrees and calculated as averages over all N-glycans. Data were collected and analysed at 100 ps intervals. Errors are shown in parenthesis and are averages of standard deviations calculated for each N-glycan. Relative populations (%) are indicated in red. All torsion angles discussed throughout correspond to the following nomenclature,  $\phi$  ( $O_5-C_1-O_x-C_x$ ),  $\psi$  ( $C_1-O_x-C_x-C_{x+1}$ ), and  $\omega$  ( $O_6-C_6-C_5-C_4$ ).

Linkage	$\phi$	$\psi$	$\omega$
GlcNAc(2)- $\beta$ (1-4)-GlcNAc(1)	-77.9 (11.3) <b>100</b>	-127.5 (15.1) <b>100</b>	-
Fuc(12)- $\alpha$ (1-6)-GlcNAc(1)	-77.3 (15.9) <b>100</b>	-185.2 (19.5) <b>78/ 94.5 (19.4)</b> <b>19/ -97.3 (16.5) 2</b>	47.1 (12.3) <b>92/ -168.5 (16.5)</b> <b>7/ -54.8 (30.5) 1</b>
Man(3)- $\beta$ (1-4)-GlcNAc(2)	-78.1 (18.2) <b>97/ -</b> 159.3 (18.5) <b>3</b>	-125.5 (15.3) <b>96/ 72.1 (12.9) 4</b>	-
(1-3) branch: Man(4)- $\alpha$ (1-3)- Man(3)	72.9 (12.2) <b>99/</b> 123.7 (22.0) <b>1</b>	147.5 (15.2) <b>60/ 102.2 (14.5)</b> <b>39</b>	-
(1-3) branch: GlcNAc(6)- $\beta$ (1- 2)-Man(4)	-80.9 (16.5) <b>100</b>	161.9 (15.6) <b>84/ 105.4 (12.4)</b> <b>16</b>	-
(1-3) branch: Gal(8)- $\beta$ (1-4)- GlcNAc(6)	-75.1 (15.9) <b>100</b>	-122.3 (16.5) <b>95/ 78.7 (21.2)</b> <b>5</b>	-
(1-3) branch: Sia(10)- $\alpha$ (2-6)- Gal(8)	65.1 (11.0) <b>90/ -50.4</b> (15.2) <b>10</b>	-182.6 (24.5) <b>85/ -99.3 (16.7)</b> <b>11/ 104.7 (17.9) 3</b>	-63.7 (15.6) <b>64/ -165.5 (15.0)</b> <b>31/ 58.2 (14.3) 5</b>
(1-6) branch: Man(5)- $\alpha$ (1-6)- Man(3) (F,H,J,R)	73.7 (15.8) <b>100</b>	85.3 (17.2) <b>49/ -178.5 (18.7)</b> <b>50/ -93.0 (17.4) 2</b>	51.9 (10.7) <b>81/ -173.6 (16.5)</b> <b>15/ -77.5 (19.6) 4</b>
(1-6) branch: Man(5)- $\alpha$ (1-6)- Man(3) (L,N,P,T)	76.3 (15.0) <b>100</b>	85.3 (15.8) <b>76/ -185.1 (22.1)</b> <b>23/ -99.3 (13.6) 0</b>	50.5 (9.8) <b>88/ -172.8 (14.9)</b> <b>11/ -80.2 (18.1) 2</b>
(1-6) branch: GlcNAc(7)- $\beta$ (1- 2)-Man(5)	-81.4 (14.9) <b>100</b>	162.0 (13.1) <b>92/ 107.8 (11.8)</b> <b>9</b>	-
(1-6) branch: Gal(9)- $\beta$ (1-4)- GlcNAc(7)	-74.8 (13.7) <b>100</b>	-122.6 (15.7) <b>99/ 153.3 (9.1) 1</b>	-
(1-6) branch: Sia(11)- $\alpha$ (2-6)- Gal(9)	64.9 (11.8) <b>88/ -52.3</b> (19.1) <b>13</b>	-182.0 (25.2) <b>88/ -101.3</b> (16.6) <b>9/ 90.3 (16.3) 4</b>	-64.9 (16.2) <b>56/ -163.5 (15.2)</b> <b>34/ 60.3 (14.4) 8</b>

**Table 3.2** Conformational propensities of different linkages calculated for all non-fucosylated species shown in **Figure 3.1**. The torsion angle values are shown in degrees and calculated as averages over all N-glycans. Data were collected and analysed at 100 ps intervals. Errors are shown in parenthesis and are averages of standard deviations calculated for each N-glycan. Relative populations (%) are indicated in red. All torsion angles discussed throughout correspond to the following nomenclature,  $\phi$  ( $O_5-C_1-O_x-C_x$ ),  $\psi$  ( $C_1-O_x-C_x-C_{x+1}$ ), and  $\omega$  ( $O_6-C_6-C_5-C_4$ ).

Linkage	$\phi$	$\psi$	$\omega$
GlcNAc(2)- $\beta$ (1-4)- GlcNAc(1)	-78.7 (11.1) <b>100</b>	-130.8 (15.7) <b>99</b> / 69.0 (12.4) <b>3</b>	-
Man(3)- $\beta$ (1-4)-GlcNAc(2)	-79.6 (20.6) <b>97</b> / -157.9 (18.2) <b>1</b> / 56.7 (9.1) <b>1</b>	-125.4 (15.4) <b>96</b> / 66.8 (11.8) <b>3</b>	-
(1-3) branch: Man(4)- $\alpha$ (1-3)- Man(3)	72.8 (12.1) <b>100</b>	146.5 (13.9) <b>61</b> / 102.3 (14.2) <b>39</b>	-
(1-3) branch: GlcNAc(6)- $\beta$ (1- 2)-Man(4)	-81.0 (16.3) <b>100</b>	161.8 (15.9) <b>84</b> / 105.4 (12.3) <b>16</b>	-
(1-3) branch: Gal(8)- $\beta$ (1-4)- GlcNAc(6)	-75.7 (17.3) <b>100</b>	-122.1 (16.4) <b>95</b> / 74.5 (17.6) <b>5</b>	-
(1-3) branch: Sia(10)- $\alpha$ (2-6)- Gal(8)	65.1 (11.2) <b>89</b> / -50.0 (14.7) <b>11</b>	-177.2 (22.9) <b>80</b> / -100.3 (15.9) <b>18</b> / 105.9 (15.8) <b>3</b>	-63.2 (15.5) <b>66</b> / -164.3 (14.8) <b>31</b> / 60.3 (13.1) <b>4</b>
(1-6) branch: Man(5)- $\alpha$ (1-6)- Man(3) (E,G,I,Q)	73.2 (16.5) <b>100</b>	82.1 (18.4) <b>45</b> / -177.7 (18.6) <b>54</b> / -95.6 (17.7) <b>2</b>	51.8 (11.0) <b>80</b> / -172.3 (17.0) <b>15</b> / -75.7 (17.9) <b>5</b>
(1-6) branch: Man(5)- $\alpha$ (1-6)- Man(3) (K,M,O,S)	75.1 (14.9) <b>100</b>	82.9 (17.3) <b>74</b> / -182.2 (21.9) <b>26</b> / -101.8 (13.4) <b>1</b>	58.4 (10.3) <b>99</b> / -173.5 (18.5) <b>1</b>
(1-6) branch: GlcNAc(7)- $\beta$ (1- 2)-Man(5)	-79.8 (19.6) <b>100</b>	162.1 (13.6) <b>90</b> / 106.5 (12.4) <b>10</b>	-
(1-6) branch: Gal(9)- $\beta$ (1-4)- GlcNAc(7)	-76.2 (15.1) <b>100</b>	-125.7 (15.4) <b>97</b> / 120.3 (7.2) <b>3</b>	-
(1-6) branch: Sia(11)- $\alpha$ (2-6)- Gal(9)	64.5 (11.2) <b>90</b> / -50.7 (13.7) <b>10</b>	-181.0 (25.2) <b>88</b> / -98.6 (17.9) <b>9</b> / 106.5 (16.0) <b>3</b>	-62.6 (15.6) <b>60</b> / -166.6 (14.3) <b>33</b> / 58.8 (13.6) <b>7</b>

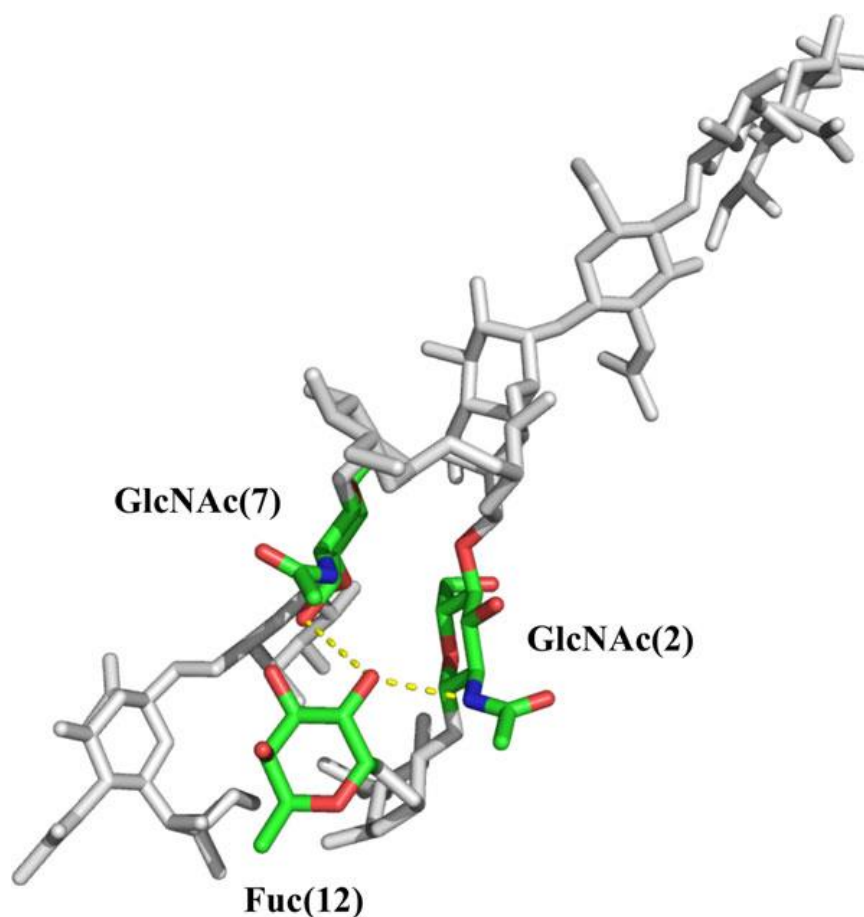
### GlcNAc(2)- $\beta$ (1-4)-GlcNAc(1) linkage

As shown in see **Tables 3.1** and **3.2**, this linkage is conformationally rigid with only one rotamer populated in all the N-glycans analysed. The average  $\phi$  angle values are  $-78.7^\circ$  and  $-77.9^\circ$ , while the  $\psi$  angle values are  $-130.8^\circ$  and  $-127.5^\circ$  for non-fucosylated and for fucosylated species, respectively. In the core-fucosylated species this conformation of the chitobiose favours the formation of a hydrogen bond between the O2 of the fucose and the NH of GlcNAc(2). No significant deviations from these torsion angle values have been observed, except in the case of the core-fucosylated tetrasaccharide sugar D, shown in **Figure 3.1**, where the GlcNAc(1) ring pucker has been observed to transition from the

more stable  ${}^4C_1$  to a  ${}^1C_4$  chair, causing a change to a  $\psi$  angle value of  $-74.0^\circ$  for 13% of the time over a 3  $\mu$ s trajectory. Because this transition has been seen exclusively for sugar D and it is caused by the reorientation of the GlcNAc(1) C5-C6 bond following the  ${}^4C_1$  to  ${}^1C_4$  re-puckering, from an equatorial to an axial position, sugar D has been excluded from the averaging. The conformational propensity of sugar D is discussed in detail in a dedicated subsection below.

### **Fuc(12)- $\alpha$ (1-6)-GlcNAc(1) linkage**

The flexibility of the  $\alpha$ (1-6) linkage between the fucose and the GlcNAc(1) is unsurprisingly higher relative to the other core monosaccharides. Notably, the conformational space sampled is independent of the length or the size of the N-glycan. The highest populated  $\alpha$ (1-6) conformer, i.e.  $\phi -77.3^\circ$  (100%),  $\psi -185.2^\circ$  (78%), and  $\omega 47.1^\circ$  (89%), with relative average populations indicated in brackets, corresponds to a structure where the fucose O2 forms a hydrogen bond with the NH of GlcNAc(2). An example of this conformation is shown in the case of the dodecasaccharide sugar T in **Figure 3.2**. As the size of the glycan increases the fucose is also able to extend the hydrogen bond network by forming interactions with the GlcNAc(7) O3 when the  $\alpha$ (1-6) arm is folded over the chitobiose core, as shown in **Figure 3.2**. A deviation from this conformation with 19% of the total population, consistently for all N-glycans, has the same  $\phi$  and  $\omega$  torsions, but a  $\psi$  value of  $94.5^\circ$ . In this conformer the fucose is not hydrogen bonded. A unique case where the  $\omega$  torsion deviates from its *gg* stable configuration, i.e.  $\omega$  of  $47.1^\circ$ , is found only in the case of the tetrasaccharide sugar D, where the reducing GlcNAc(1)  ${}^1C_4$  chair conformation has the C5 in an axial position and the  $\omega$  torsion to adopt a  $-64.5^\circ$  value. This  $\omega$  value corresponds to an energetically prohibited *tg* conformer when the GlcNAc(1) is in the standard  ${}^4C_1$  chair. The *gt*  $\omega$  conformation, corresponding to an average value of  $-168.7^\circ$ , has relative populations ranging between 3% (sugar T) and 15% (sugar F) in a pattern that doesn't seem to be dependent on the branching, the length, nor on the sequence.



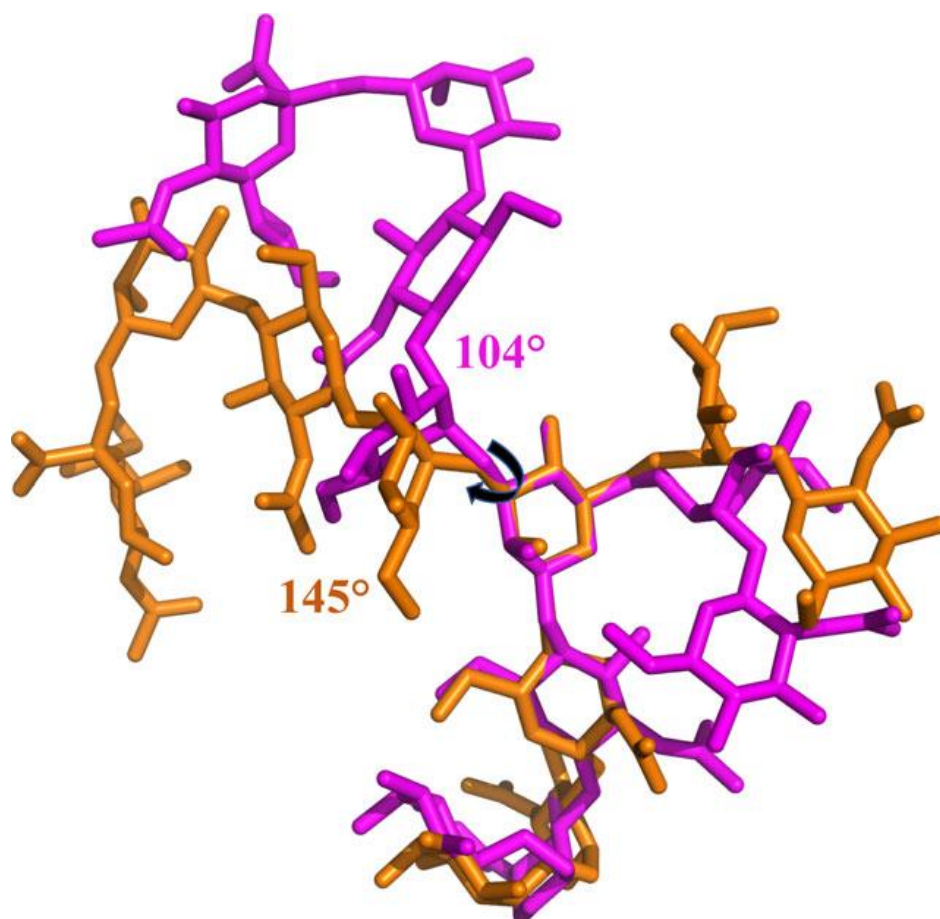
**Figure 3.2.** Hydrogen bond between the Fuc(12) O2 and the amide N of GlcNAc(2) and GlcNAc(7) in the dodecasaccharide sugar T is highlighted with yellow dots. Aside from Fuc(12) and GlcNAc(2) all other monosaccharides are shown in light grey. Note: the conformation shown, the highest populated in sugar T, has the  $\alpha(1-6)$  arm folded over the chitobiose core. The image was generated with pymol.

### **Man(3)- $\beta(1-4)$ -GlcNAc(2) linkage**

This linkage has one largely preferred conformation for both, fucosylated and non-fucosylated species, irrespectively of the length and sequence of the N-glycan. For non-fucosylated species  $\phi$  and  $\psi$  have average values of  $-79.6^\circ$  (97%) and  $-125.4^\circ$  (96%), respectively. The relative populations calculated over the entire simulation time are indicated in parentheses. For core-fucosylated species  $\phi$  and  $\psi$  have average values of  $-78.1^\circ$  (97%) and  $-125.5^\circ$  (96%), respectively. Conformational changes from this prevalent configuration amount to 3% to 4% of the whole cumulative simulation time and represent changes of  $\phi$  to values around  $-160^\circ$  and of  $\psi$  to values around  $60^\circ$  for both, non- and core-fucosylated species, see **Tables 3.1** and **3.2**. These  $\phi$  and  $\psi$  excursions are not correlated to each other and confer a slight flexibility in terms of rotations of the plane containing the  $\alpha(1-3)$  and  $\alpha(1-6)$  branches around the chitobiose core. In the most stable conformation the planes containing the chitobiose and the trimannose core are parallel to each other.

### $\alpha(1-3)$ branch: Man(4)- $\alpha(1-3)$ -Man(3) linkage

The conformational dynamics of the  $\alpha(1-3)$  branch at the trimannose level is not particularly complex and it is not affected by core fucosylation. The  $\phi$  torsion is prevalently found in a single conformation,  $72.8^\circ$  in 100% of all non-fucosylated species analysed and  $72.9^\circ$  in 99% of all core-fucosylated species analysed. The slight difference is due to the more complex dynamics of sugar P and sugar T, which both have sialylated  $\alpha(1-3)$  arms, and for which in 5% and 1% of the sampling time the  $\phi$  torsion value is an average of  $123.7^\circ$ , see **Table S.9** in **Appendix I**. The  $\psi$  torsion defines two distinct basins, identical in case of both fucosylated and non-fucosylated species. In case of non-fucosylated N-glycans the highest populated  $\psi$  value is  $146.5^\circ$  (61%) and the lowest is  $102.3^\circ$  (39%). In the case of fucosylated species the  $\psi$  values are,  $147.5^\circ$  with a relative population of 60% and  $102.2^\circ$  with a population of 39%. The remaining 1% is contributed by the more complex dynamics of sugar P, where the third  $\psi$  value corresponds to  $182.0^\circ$ . The two distinct conformations of this linkage confer a degree of flexibility to the  $\alpha(1-3)$  branch with a range of movement shown in **Figure 3.3**.



**Figure 3.3.** The conformational dynamics of the  $\alpha(1-3)$  arm, shown here for sugar R as an example, is restricted to two basins corresponding to  $\psi$  values of  $145^\circ$  (orange) and  $104^\circ$  (purple). The image was generated with pymol.



### **$\alpha(1-3)$ branch: GlcNAc(6)- $\beta(1-2)$ -Man(4) linkage**

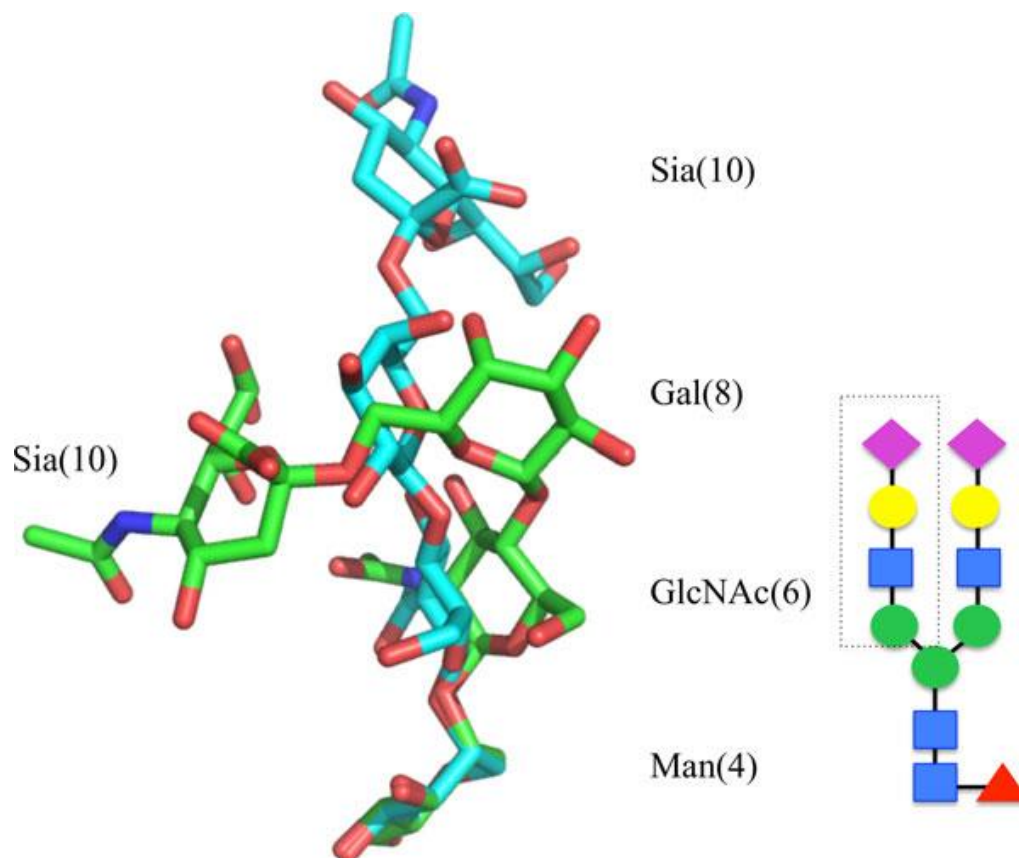
The conformation of this linkage is restricted to two conformations, identical for both fucosylated and non-fucosylated species. In case of the non-fucosylated glycans the average  $\phi$  value is  $-81.0^\circ$ , while the  $\psi$  values are  $161.8^\circ$  (84%) and  $105.4^\circ$  (16%). In case of the fucosylated glycans the average  $\phi$  value is  $-80.9^\circ$ , while the  $\psi$  values are  $161.9^\circ$  (84%) and  $105.4^\circ$  (16%).

### **$\alpha(1-3)$ branch: Gal(8)- $\beta(1-4)$ -GlcNAc(6) linkage**

This linkage is the least flexible in the branch with one conformation accounting for over 95% of the population of both fucosylated and non-fucosylated N-glycans. In case of the non-fucosylated glycans the average  $\phi$  value is  $-75.7^\circ$ , while the  $\psi$  value is  $-122.1^\circ$  (95%). In case of the fucosylated glycans the average  $\phi$  value is  $-75.1^\circ$ , while the  $\psi$  value is  $-122.3^\circ$  (95%). The 5% difference corresponds to excursions to  $\psi$  values around  $75^\circ$  for both fucosylated and non-fucosylated species, see **Tables 3.1** and **3.2**.

### **$\alpha(1-3)$ branch: Sia(10)- $\alpha(2-6)$ -Gal(8) linkage**

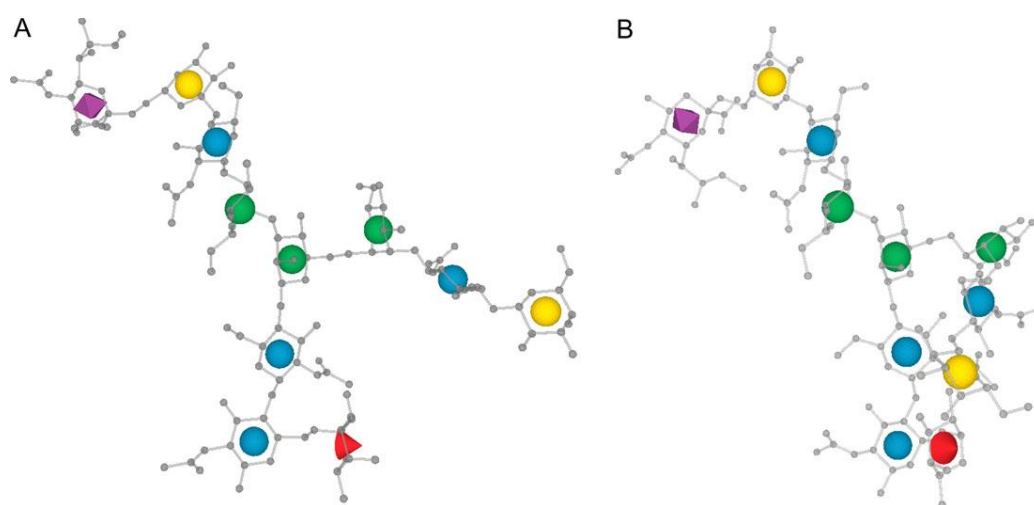
The linkage to the terminal sialic acid is the most flexible of the  $\alpha(1-3)$  arm and all the conformations visited are independent of core-fucosylation, both in terms of torsion angle values and populations, see **Tables 3.1** and **3.2**. The most flexible torsion is the  $\omega$  angle with values around  $-60^\circ$  (65%) and  $-165^\circ$  (30%). A small contribution of around 5% is given by a value of  $60^\circ$ . These two  $\omega$  values define the conformations shown in **Figure 3.4**, with the highest populated corresponding to the completely out-stretched  $\alpha(1-3)$  arm.



**Figure 3.4.** The two highest populated conformations accessible to the terminal sialic acid here shown for sugar T on the  $\alpha(1-3)$  arm. Note, only the section of the  $\alpha(1-3)$  arm from Man(4) is shown for clarity. The conformations corresponding to  $\omega$  torsion values of  $-60^\circ$  (65%) and  $-165^\circ$  (30%) are shown with C atoms in cyan and green, respectively. The image was generated with pymol.

### **$\alpha(1-6)$ branch: Man(5)- $\alpha(1-6)$ -Man(3) linkage**

The conformational dynamics of the Man(5)- $\alpha(1-6)$ -Man(3) linkage is the most interesting and the only one that has a clear dependence on sequence among the IgG Fc N-glycans analysed here. Unlike the  $\alpha(1-3)$  arm that has a relatively restricted dynamics, the  $\alpha(1-6)$  arm can adopt two very distinct conformations, one extended, or ‘outstretched’, corresponding to  $\phi$  values around  $-180^\circ$ , and one where the arm is folded over the chitobiose core, which we’ll refer to as ‘folded-over’, corresponding to  $\phi$  values around  $80^\circ$ , see **Tables 3.1, 3.2** and **Figure 3.5**. In sugar E (F), with an  $\alpha(1-6)$  arm terminating with Man, over 80% of the conformer have the arm in an outstretched conformation, see **Tables S.16** and **S.17** in **Appendix I**. The addition of the GlcNAc in  $\beta(1-2)$  shifts this equilibrium, so for sugars G (H), I (J), and Q (R), where the fucosylated species are indicated in parenthesis, the outstretched and the folded-over conformations are equally populated, namely 45% (49%) for the folded-over conformation, and 55% (51%) for the outstretched conformation. Interestingly, galactosylation of the  $\alpha(1-6)$  arm shifts this equilibrium further with a high majority of the galactosylated sugars, namely K (L), M (N), O (P) and S (T), in a folded-over conformation. The relative average populations are 74% and 76% in a folded-over conformation for non-fucosylated and core-fucosylated sugars, respectively, while 26% and 23% in an outstretched conformation for non-fucosylated and core-fucosylated sugars, respectively. This equilibrium is not affected either by the type of glycosylation in the  $\alpha(1-3)$  arm, nor by the sialylation of the  $\alpha(1-6)$  arm, and it does not depend on core-fucosylation either, despite the fucose in its most stable conformation interacts effectively through hydrogen bonding with the  $\alpha(1-6)$  arm when folded-over, as shown in **Figure 3.2**.



**Figure 3.5.** Two main conformations accessible to the  $\alpha(1-6)$  arm represented for sugar P as an example. The ‘outstretched’ conformation is shown on panel a) and the ‘folded-over’ on panel b). The image was generated with LiteMol with the 3D-SNFG rendering<sup>39</sup>.

#### **$\alpha(1-6)$ branch: GlcNAc(7)- $\beta(1-2)$ -Man(5) linkage**

As seen in the case of the  $\alpha(1-3)$  arm, the conformational dynamics of this linkage is largely restricted to  $\phi$  and  $\psi$  values of  $-80^\circ$  and  $162^\circ$ , respectively, both in case of fucosylated and non-fucosylated species, see **Tables 3.1** and **3.2**. The only significant alternative conformer accounting for 10% total contribution is characterized by  $\psi$  values of  $106^\circ$  for non-fucosylated species and  $108^\circ$  for fucosylated species. The same linkage on the  $\alpha(1-3)$  arm has an analogous conformational equilibrium of the  $\psi$  torsion, with a 85:15 ratio.

#### **$\alpha(1-6)$ branch: Gal(9)- $\beta(1-4)$ -GlcNAc(7) linkage**

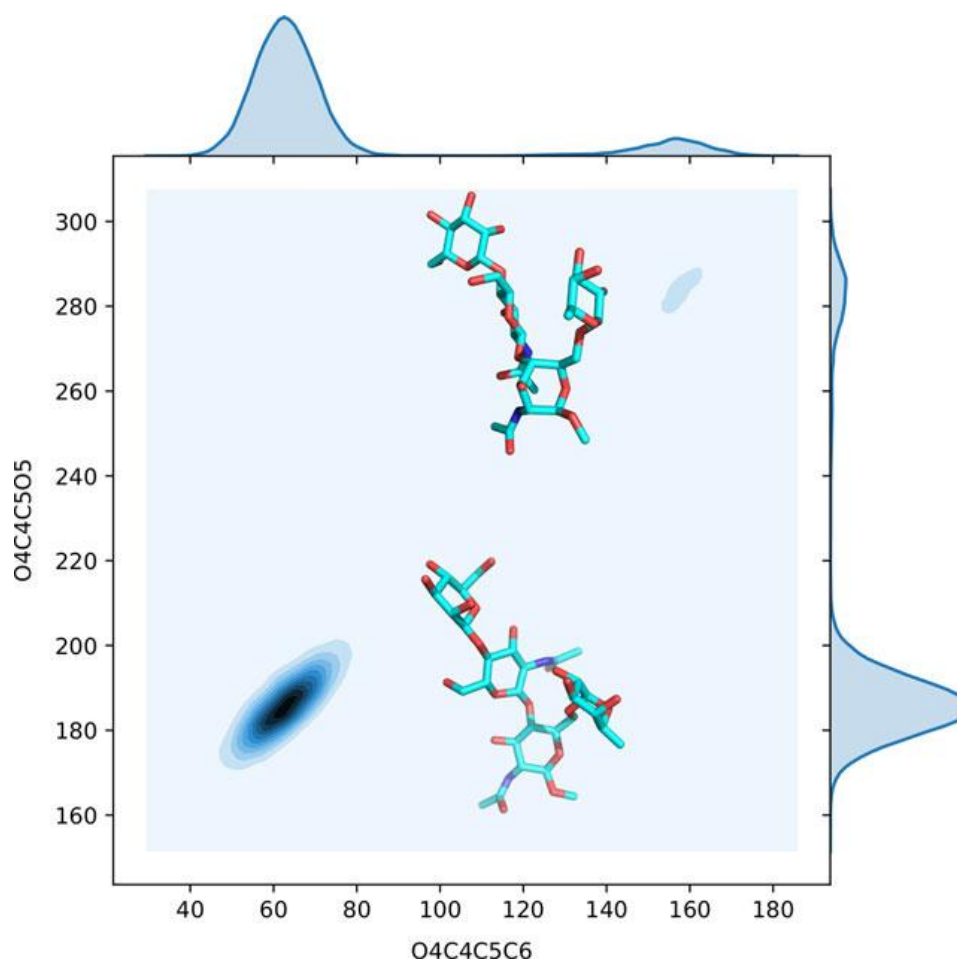
Similarly to the same linkage on the  $\alpha(1-3)$  arm, the Gal(9)- $\beta(1-4)$ -GlcNAc(7) torsional space is highly restricted to a single conformation with  $\phi$  and  $\psi$  values around  $-75^\circ$  and  $-125^\circ$ , respectively, both in the cases of fucosylated and non-fucosylated species.

### **$\alpha(1-6)$ branch: Sia(11)- $\alpha(2-6)$ -Gal(9) linkage**

The conformational dynamics of the terminal sialic acid on the  $\alpha(1-6)$  arm is virtually identical to the one observed for the terminal sialic acid on the  $\alpha(1-3)$  arm. The linkage is highly flexible with, the  $\phi$  angle predominantly around  $65^\circ$  (90%), with a  $\sim 10\%$  contribution of  $-52^\circ$ . The  $\psi$  angle has also a preferred conformation around  $-180^\circ$  (88%). Small contributions of  $\psi$  values around  $-100^\circ$  (9%), and  $+100^\circ$  (3-4%) have been also detected, both in case of fucosylated and of non-fucosylated species. The  $\omega$  angle is quite flexible, with values of  $-65^\circ$  as the preferred conformation (60%) and of  $-162^\circ$  (33%). A small contribution of 7% corresponds to an  $\omega$  value of  $60^\circ$ , both in case of fucosylated and of non-fucosylated species.

### **Sugar D: Man- $\beta(1-4)$ -GlcNAc- $\beta(1-4)$ -[ $\alpha(1-6)$ -Fuc]-GlcNAc**

The conformational dynamics of the tetrasaccharide Man- $\beta(1-4)$ -GlcNAc- $\beta(1-4)$ -[ $\alpha(1-6)$ -Fuc]-GlcNAc, named here sugar D for short, see **Figure 3.1**, is quite complex showing occasional transitions of the reducing GlcNAc ring from the stable  ${}^4C_1$  chair conformation to a less stable  ${}^1C_4$  chair. Similar conformational changes have been reported recently for Lewis X ( $Le^x$ ) and sialyl  $Le^x$  ( $sLe^x$ )<sup>40,41</sup>, where the linkage to the fucose is a less flexible  $\beta(1-3)$ . Although we observed the first  ${}^4C_1$  to  ${}^1C_4$  chair transition within the first 100 ns of the simulation, we extended the trajectory to 3  $\mu$ s to evaluate more accurately the relative populations of the two different conformers. During the 3  $\mu$ s trajectory the  ${}^1C_4$  chair conformation is visited 5 times and remains stable during intervals ranging from 30 ns up to 200 ns. Based on these data, as shown in **Figure 3.6**, sugar D is found in the stable  ${}^4C_1$  chair for over 87% of the time. In terms of relative populations, the corresponding free energy for the  ${}^4C_1$  to  ${}^1C_4$  chair transition corresponds to 4.7 kJ/mol at 300 K.



**Figure 3.6.** Contour plot showing the conformational propensity of the reducing GlcNAc(1) of sugar D in function of two torsion angles O4C4C5C6 and O4C4C5O5 on the x and y axis, respectively. The structures corresponding to the highest populated  ${}^4C_1$  chair conformation (87%) and the lowest populated  ${}^1C_4$  chair conformation (13%) over 3  $\mu$ s (30,000 data points) are also shown in line with the corresponding histograms on the y axis. Torsion angles are in degrees. 2D contour plots are done with seaborn ([seaborn.pydata.org](https://seaborn.pydata.org)) and the structures are rendered with pymol.

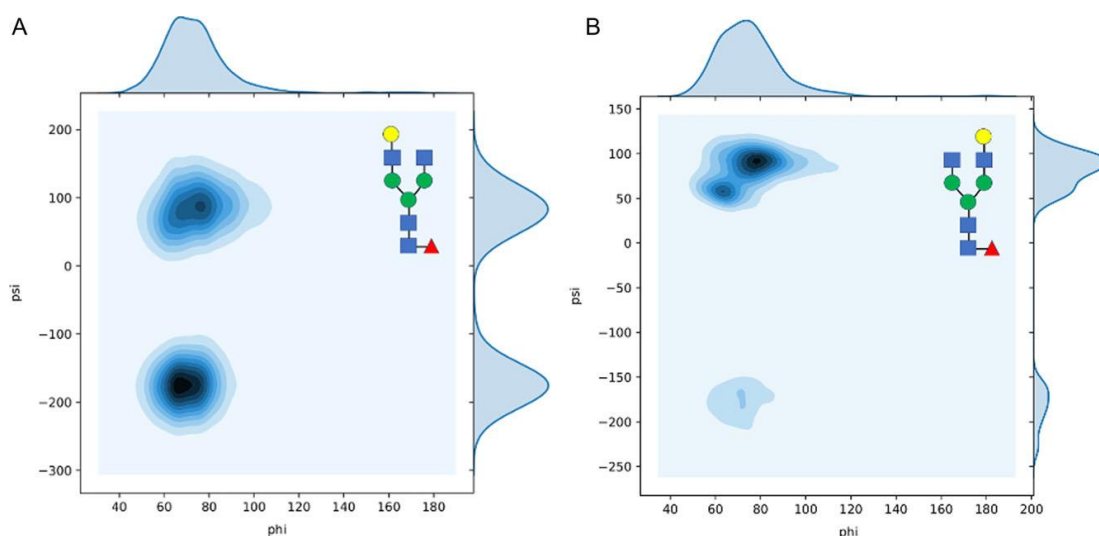
### 3.4 Discussion

The sequence of the N-glycans expressed at Asn 297 has been shown to have a significant effect in modulating the IgG effector function. In particular core-fucosylation, and galactosylation and sialylation of both  $\alpha(1-6)$  and  $\alpha(1-3)$  arms in Fc biantennary complex N-glycans have been identified as key players, with different roles in regulating the ADCC function<sup>5, 13, 16</sup> and the onset of inflammation<sup>24, 26</sup>. In particular levels of galactosylation have been linked to aging<sup>31</sup> and to the risk of developing conditions such as rheumatoid arthritis<sup>25, 27-29</sup>. The molecular basis underlying the link between the glycan sequence and these different phenotypes is very complex to understand, as it most likely depends on the dynamics and energetics of the interaction between the glycosylated antibody and

the receptors on the immune cells<sup>21</sup>. As a first step we wanted to determine if and how the monosaccharide sequence affects the N-glycan dynamics in the absence of the protein, so we analysed the intrinsic conformational propensity of the biantennary complex N-glycans most commonly expressed at Asn 297 on the IgG Fc<sup>5,10</sup>. Although we do expect that contacts with residues on the IgG Fc surface are likely to affect the conformational equilibrium relative to the isolated glycan, it has been shown that both  $\alpha(1-6)$  and  $\alpha(1-3)$  arms retain high degrees of flexibility in Fc-linked glycans<sup>30</sup>, therefore their intrinsic conformational propensity may play a role in their contact with the receptors. The results of our study based on extensive conformational sampling in excess of 62  $\mu$ s by MD simulations, show that while core-fucosylation and sialylation do not affect the conformational dynamics of the N-glycan as a whole, the effect of galactosylation of the  $\alpha(1-6)$  arm is remarkable, as it clearly regulates its conformation. Indeed, we have found that regardless of the sequence of the  $\alpha(1-3)$  arm, of core-fucosylation, or of sialylation of the  $\alpha(1-6)$  arm, the presence of galactose  $\beta(1-4)$  linked to GlcNAc shifts the conformational propensity of the whole  $\alpha(1-6)$  arm from an ‘outstretched’ conformation to a ‘folded-over’ conformation where the  $\alpha(1-6)$  arm is stacked against the chitobiose core, see **Figure 3.5**. The shift in the conformational equilibrium we observe for galactosylated  $\alpha(1-6)$  arms is likely due to a more effective interaction in terms of both hydrogen bonding and of stacking with the chitobiose core, regardless of the presence of the core-fucose. This is surprising as fucose does contribute to the hydrogen bonding network by bridging the GlcNAc(2) of the chitobiose to the GlcNAc(7) of the folded-over  $\alpha(1-6)$  arm, as shown in **Figure 3.2**. Conversely, the dynamics of the  $\alpha(1-3)$  arm is more restricted due to the intrinsic rigidity of the Man- $\alpha(1-3)$ -Man relative to the Man- $\alpha(1-6)$ -Man. Nevertheless, it undergoes excursions between the two close rotamers, as shown in **Figure 3.3**. The conformational dynamics of both arms is locally enhanced by the sialic acid due to the intrinsic flexibility of the  $\alpha(2-6)$  linkage.

Although the existence of folded-over (and variation there-of) and of outstretched conformations of the  $\alpha(1-6)$  arm have been identified and discussed previously<sup>42, 43</sup>, our systematic study highlights the unique dependence of this conformational propensity on the galactosylation of the  $\alpha(1-6)$  arm. This behaviour can explain the recent evidence of differential recognition of positional isomers in glycan array screening<sup>44</sup>. As an example, the positional isomers sugar J and L have a galactose on either the  $\alpha(1-3)$  or the  $\alpha(1-6)$  arm, respectively, which confers a different conformational propensity of the  $\alpha(1-6)$  arm. Indeed, in sugar J, shown in **Figure 3.7.A**, the folded-over and the outstretched conformations are equally populated during the MD simulations, i.e. 49% and 50%, respectively (note: the 1% corresponds to a  $\psi$  value of  $-99.3^\circ$ ), while in sugar L, shown in **Figure 3.7.B**, the folded-over  $\alpha(1-6)$  arm is the dominant conformation with a relative population of 81% over the simulation time. As a further implication of this behaviour, a different structural propensity shifted towards a folded-over  $\alpha(1-$

6) arm and an outstretched  $\alpha(1-3)$  arm is in agreement with the higher affinity of the sialyltransferase for the  $\alpha(1-3)$  galactose, more accessible in isolated glycans<sup>30</sup>. Interestingly, this selectivity is only moderately changed when the N-glycans are linked to the IgG Fc<sup>30</sup>.



**Figure 3.7.** Contour plot showing the different conformational propensities of the  $\alpha(1-6)$  arm in sugar J (panel a) and L (panel b) in function of galactosylation. The ‘folded over’ conformation corresponds to a  $\psi$  value around  $80^\circ$ , while the outstretched conformation corresponds to a  $\psi$  value around  $-180^\circ$ . Torsion angles are in degrees. 2D contour plots are done with seaborn ([seaborn.pydata.org](http://seaborn.pydata.org)) with 2,500 data points.

The systematic conformational analysis of progressively longer fucosylated and non-fucosylated N-glycans has allowed us to highlight an additional and unique complexity in the dynamics of the tetrasaccharide sugar D, namely  $\text{Man-}\beta(1-4)\text{-GlcNAc-}\beta(1-4)\text{-}[\alpha(1-6)\text{-Fuc}]\text{-GlcNAc}$ . During a single (unbiased) MD trajectory extended to  $3 \mu\text{s}$ , we have observed the reducing GlcNAc ring transitioning from the most stable  ${}^4\text{C}_1$  chair to a  ${}^1\text{C}_4$  chair a total of 5 times, where the  ${}^1\text{C}_4$  chair remains stable for intervals between 30 to 200 ns. This relative ease of interconversion, corresponding to a free energy of 4.7 kJ/mol, is somewhat in contrast with the results obtained for  $\text{Le}^{\text{x(a)}}$  and  $\text{sLe}^{\text{x(a)}}$  where opening events are much more rare and much more difficult to sample through conventional MD simulations<sup>40, 41</sup>. The greater accessibility of the open state in sugar D is most likely related to the higher flexibility of the  $\alpha(1-6)$  linkage to the fucose, relative to the  $\alpha(1-3)$  and  $\alpha(1-4)$  linkages in  $(\text{s})\text{Le}^{\text{x}}$  and  $(\text{s})\text{Le}^{\text{a}}$ , respectively. Indeed, in each of the 5 events sampled in this study and in agreement with previous work<sup>40</sup>, the  ${}^4\text{C}_1$  to  ${}^1\text{C}_4$  transition involves high energy intermediate steps, where the reducing GlcNAc ring adopts boat and skewed-boat conformations, while the  $\beta(1-4)\text{-GlcNAc}$  and the  $\alpha(1-6)\text{-Fuc}$  change from an equatorial to an axial configuration. It is reasonable that the energy cost involved in this latter equatorial-to-axial conversion is less for the more flexible sugar D, relative to the more rigid  $(\text{s})\text{Le}^{\text{x}}$  and



(s)Le<sup>a</sup>. The conformational propensity of the  $\alpha(1-6)$  linkage through the 3  $\mu$ s trajectory is shown in **Figure S.1 in Appendix I**.

### 3.5 Conclusions

Here we have shown the results of extensive sampling obtained through unbiased MD simulations of the conformational space accessible to increasingly larger biantennary complex N-glycans, commonly expressed at the Asn 297 in the IgG Fc region. Our data indicate that while core-fucosylation and sialylation do not affect the overall conformation of the isolated N-glycan as a whole, but contribute to its local dynamics, galactosylation of the  $\alpha(1-6)$  arm shifts its conformational equilibrium towards a structure where the arm is ‘folded over’ the chitobiose core. This effect is determined by more effective hydrogen bonding and stacking interactions between the chitobiose and the ‘longer’ galactosylated  $\alpha(1-6)$  arm and it is independent of the sequence of the  $\alpha(1-3)$  arm, of core-fucosylation, and of sialylation of the  $\alpha(1-6)$  arm. These results can explain the differential recognition of positional isomers<sup>44</sup> and with the preference of sialyltransferases for the  $\alpha(1-3)$  arm in both, isolated and Fc-linked N-glycans<sup>30</sup>. Currently we are in the process of determining how the dynamics and the conformational equilibria we have discussed here for the isolated biantennary N-glycans are affected by the presence of the IgG Fc protein surface. As a note, a relatively long simulation of the tetrasaccharide Man- $\beta(1-4)$ -GlcNAc- $\beta(1-4)$ -[ $\alpha(1-6)$ -Fuc]-GlcNAc, named here sugar D, has highlighted an equilibrium between a high populated ‘closed’ form of the sugar, where GlcNAc(1) is in the stable <sup>4</sup>C<sub>1</sub> chair conformation, and an ‘open’ form, where GlcNAc(1) is in the <sup>1</sup>C<sub>4</sub> chair conformation. We find that the accessibility to the open state is significantly higher for sugar D, relative to (s)Le<sup>x(a)</sup>, as reported in recent work<sup>40, 41</sup>. This is most likely due to the higher flexibility of the  $\alpha(1-6)$ -linked fucose relative to the  $\alpha(1-3/4)$ .

## References

1. S. Krapp, Y. Mimura, R. Jefferis, R. Huber, P. Sondermann, Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J Mol Biol* **325**, 979-989 (2003).
2. J. Fang, J. Richardson, Z. M. Du, Z. Q. Zhang, Effect of Fc-Glycan Structure on the Conformational Stability of IgG Revealed by Hydrogen/Deuterium Exchange and Limited Proteolysis. *Biochemistry-Us* **55**, 860-868 (2016).
3. J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd, R. A. Dwek, The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50 (2007).
4. A. Kobata, The N-linked sugar chains of human immunoglobulin G: their unique pattern, and their functional roles. *Biochim Biophys Acta* **1780**, 472-478 (2008).
5. T. S. Raju, Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Curr Opin Immunol* **20**, 471-478 (2008).
6. G. P. Subedi, A. W. Barb, The immunoglobulin G1 N-glycan composition affects binding to each low affinity Fc gamma receptor. *MAbs* **8**, 1512-1524 (2016).
7. J. M. Hayes *et al.*, Identification of Fc gamma receptor glycoforms that produce differential binding kinetics for rituximab. *Mol Cell Proteomics*, (2017).
8. R. L. Shields *et al.*, High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. *J Biol Chem* **276**, 6591-6604 (2001).
9. M. H. Tao, S. L. Morrison, Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region. *J Immunol* **143**, 2595-2601 (1989).
10. M. Pucic *et al.*, High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* **10**, M111 010090 (2011).
11. S. Matsumiya *et al.*, Structural comparison of fucosylated and nonfucosylated Fc fragments of human immunoglobulin G1. *J Mol Biol* **368**, 767-779 (2007).
12. M. Ratner, Genentech's glyco-engineered antibody to succeed Rituxan. *Nat Biotechnol* **32**, 6-7 (2014).
13. M. Satoh, S. Iida, K. Shitara, Non-fucosylated therapeutic antibodies as next-generation therapeutic antibodies. *Expert Opin Biol Th* **6**, 1161-1173 (2006).

14. Y. Kanda *et al.*, Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: the high-mannose, hybrid, and complex types. *Glycobiology* **17**, 104-118 (2007).
15. C. L. Chen *et al.*, Crystal Structure of a Homogeneous IgG-Fc Glycoform with the N-Glycan Designed to Maximize the Antibody Dependent Cellular Cytotoxicity. *Acs Chem Biol* **12**, 1335-1345 (2017).
16. T. Li *et al.*, Modulating IgG effector function by Fc glycan engineering. *Proc Natl Acad Sci U S A* **114**, 3485-3490 (2017).
17. W. R. Strohl, Optimization of Fc-mediated effector functions of monoclonal antibodies. *Curr Opin Biotechnol* **20**, 685-691 (2009).
18. S. Iida *et al.*, Two mechanisms of the enhanced antibody-dependent cellular cytotoxicity (ADCC) efficacy of non-fucosylated therapeutic antibodies in human blood. *Bmc Cancer* **9**, (2009).
19. S. Iida *et al.*, Nonfucosylated therapeutic IgG1 antibody can evade the inhibitory effect of serum immunoglobulin g on antibody-dependent cellular cytotoxicity through its high binding to Fc gamma RIIIa. *Clin Cancer Res* **12**, 2879-2887 (2006).
20. E. Oflazoglu, L. P. Audoly, Evolution of anti-CD20 monoclonal antibody therapeutics in oncology. *MAbs* **2**, 14-19 (2010).
21. C. Ferrara *et al.*, Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc gamma RIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669-12674 (2011).
22. A. Okazaki *et al.*, Fucose depletion from human IgG1 oligosaccharide enhances binding enthalpy and association rate between IgG1 and Fc gamma RIIIa. *J Mol Biol* **336**, 1239-1249 (2004).
23. R. M. Anthony *et al.*, Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG fc. *Science* **320**, 373-376 (2008).
24. S. E. de Jong *et al.*, IgG1 Fc N-glycan galactosylation as a biomarker for immune activation. *Sci Rep-Uk* **6**, (2016).
25. R. Parekh *et al.*, A comparative analysis of disease-associated changes in the galactosylation of serum IgG. *J Autoimmun* **2**, 101-114 (1989).
26. C. M. Karsten *et al.*, Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of Fc gamma RIIB and dectin-1. *Nat Med* **18**, 1401-1406 (2012).
27. I. Gudelj *et al.*, Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10years of follow-up. *Biochim Biophys Acta* **1864**, 2034-2039 (2018).
28. M. Pasek *et al.*, Galactosylation of IgG from rheumatoid arthritis (RA) patients - changes during therapy. *Glycoconjugate J* **23**, 463-471 (2006).

29. R. B. Parekh *et al.*, Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452-457 (1985).
30. A. W. Barb, J. H. Prestegard, NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic. *Nat Chem Biol* **7**, 147-153 (2011).
31. T. S. Raju, J. B. Briggs, S. M. Borge, A. J. S. Jones, Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* **10**, 477-486 (2000).
32. K. N. Kirschner *et al.*, GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622-655 (2008).
33. W. L. Jorgensen, C. Jenson, Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry* **19**, 1179-1186 (1998).
34. V. Hornak *et al.*, Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712-725 (2006).
35. R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198-210 (2013).
36. A. W. Sousa da Silva, W. F. Vranken, ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes* **5**, 367 (2012).
37. E. Fadda, R. J. Woods, Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today* **15**, 596-609 (2010).
38. M. R. Wormald *et al.*, Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem Rev* **102**, 371-386 (2002).
39. D. Sehnal *et al.*, LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat Methods* **14**, 1121-1122 (2017).
40. I. Alibay, K. K. Burusco, N. J. Bruce, R. A. Bryce, Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *J Phys Chem B* **122**, 2462-2474 (2018).
41. J. Topin *et al.*, The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *Acs Chem Biol* **11**, 2011-2020 (2016).
42. A. Canales *et al.*, Breaking pseudo-symmetry in multiantennary complex N-glycans using lanthanide-binding tags and NMR pseudo-contact shifts. *Angew Chem Int Ed Engl* **52**, 13789-13793 (2013).
43. R. Galvelis, S. Re, Y. Sugita, Enhanced Conformational Sampling of N-Glycans in Solution with Replica State Exchange Metadynamics. *J Chem Theory Comput* **13**, 1934-1942 (2017).

44. B. Echeverria *et al.*, Chemoenzymatic Synthesis of N-glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-type Lectin Receptors. *Acs Chem Biol*, (2018).

## Chapter 4: How and why plants and human N-glycans are different: Insight from molecular dynamics into the “glycoblocks” architecture

**Paper citation:** Fogarty, C. A.\*, Harbison, A. M.\*, Dugdale, A. R. & Fadda, E. How and why plants and human N-glycans are different: Insight from molecular dynamics into the ‘glycoblocks’ architecture of complex carbohydrates. *Beilstein J. Org. Chem.* **16**, 2046–2056 (2020).

\*Shared first authorship

### 4.1 Introduction

Complex carbohydrates (or glycans) are an essential class of biomolecules, directly implicated in the cell’s interactions with its environment, facilitating communication and infection<sup>1, 2</sup>. These processes are often initiated by molecular recognition involving carbohydrate-binding proteins (lectins) or by glycan-glycan interactions<sup>1, 3-5</sup>, all events that hinge on specific structural and dynamic features of the glycans. This makes the 3D complementarity of the glycans architecture key towards the success of these processes and an essential piece of information for us to have in order to understand glycan recognition. Because of their chemical nature, glycans are intrinsically flexible and highly dynamic at room temperature, thus their characterization through experimental structural biology methods is hardly straightforward even in cryogenic environments<sup>6</sup>. As an additive layer of difficulty, glycosylation is only indirectly dependent on the genome, which often results in a micro- (or macro-) heterogeneity of glycan sequences at specific sites<sup>7</sup>. These complexities are very difficult to resolve, requiring high levels of expertise and multi-layered orthogonal approaches<sup>7-10</sup>. Within this framework, the contribution of glycoinformatics tools and databases represents an essential resource to advance glycomics<sup>11-15</sup>, while molecular simulations fit in very well as complementary and orthogonal techniques to support and advance structural glycobiology research. Indeed, current high performance computing (HPC) technology allows us to study realistic model systems<sup>16,17</sup> and to reach experimental timescales<sup>18</sup>, so that computing can now contribute as one of the leading research methods in structural glycobiology.

One of the most interesting and remarkably challenging areas in glycoscience research that HPC simulations can address is the study of the links between glycans sequence and 3D structure. This direct relationship is a well-recognized and broadly accepted concept in proteins’ structural biology, according

to which the amino acid sequence dictates the functional 3D fold and its stability.\* However, the same notion is not generally invoked when discussing other biopolymers or complex carbohydrates. In the specific case of glycans, the structural complexity, in terms of the diversity of monosaccharides, the linkages' stereochemistry and the branched scaffolds, makes the already difficult case even more intricate. Nevertheless, the fact that glycoforms follow recurrent sequence patterns, clearly suggests that the glycans 3D structure is also non-random and very likely sequence-determined. We use computer modelling to gain insight into these relationships and to define a framework to understand how subtle modifications to the glycans sequence can alter their 3D structure and conformational dynamics, ultimately regulating recognition<sup>19</sup>. In this work we use molecular dynamics (MD) simulations to analyse the effects of the inclusion of motifs typically found in plants and invertebrates N-glycans which are immunogenic in mammals<sup>20-23</sup>. More specifically, we investigate how core  $\alpha(1-3)$ -linked fucose (Fuc) and  $\beta(1-2)$ -linked xylose (Xyl) affect the structure and dynamics of plants N-glycoforms<sup>23</sup> and of hybrid constructs with mammalian N-glycoforms<sup>24</sup>.

At first glance plants protein N-glycosylation<sup>23</sup> is quite similar to the one of higher species<sup>25</sup>, carrying the distinctive trimannose core (Man3), which can be further functionalised with  $\beta(1-2)$  linked GlcNAc residues on the arms. As a trademark feature, shown in **Figure 4.1**, plants N-glycans can also have a  $\beta(1-2)$ -Xyl linked to the central mannose and core  $\alpha(1-3)$ -Fuc, instead of the  $\alpha(1-6)$ -Fuc commonly found in mammalian complex N-glycans. Additionally, the arms can be further functionalised with terminal galactose (Gal) in  $\beta(1-3)$  instead of  $\beta(1-4)$ <sup>23</sup>, commonly found in vertebrates, which forces the addition of fucose in the  $\alpha(1-4)$  position of the GlcNAc and results in the occurrence of Lewis A (LeA) instead of Lewis X (LeX) terminal motifs on the arms<sup>23,26</sup>. In a previous study, we characterized through extensive sampling the structure and dynamics of complex biantennary N-glycans commonly found in the human IgGs Fc region<sup>24</sup>. The results of this study indicated a clear sequence-to-structure relationships, especially in the context of the dynamics of the (1-6) arm. More specifically, we found that the outstretched (open) conformation of the (1-6) arm gets progressively less populated as the functionalization of the arm grows, i.e. from 85% in Man3, to 52% in (F)A2, (F)A2[3]G1, and (F)A2[3]G1S1, where the (F) indicates the presence or absence of  $\alpha(1-6)$  core fucosylation, to 24% in all structures with (1-6) arm terminating with Gal- $\beta(1-4)$ -GlcNAc or Sia- $\alpha(2-6)$ -Gal- $\beta(1-4)$ -GlcNAc, irrespective of the functionalization of the (1-3) arm<sup>24</sup>. As a practical implication of these results, positional isomers, such as (F)A2[3]G1 and (F)A2[6]G1, have different conformational propensities, the latter with a much lower population of outstretched (1-6) arm and therefore quite different 3D average structures, which ultimately explains their differential recognition in glycan arrays<sup>27</sup>.

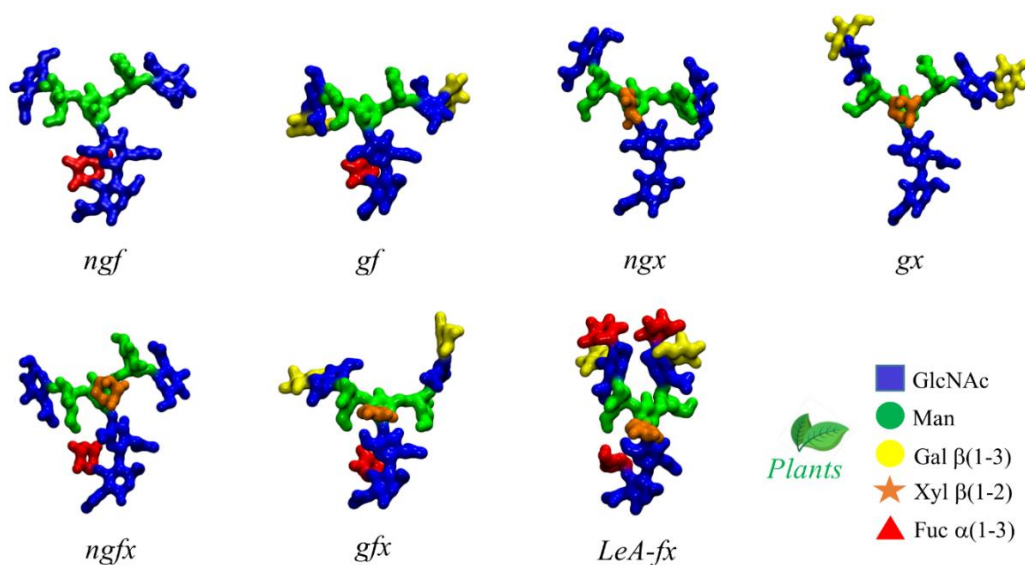
---

\* First shown in the preliminary findings of Christian Anfinsen (Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* (80-. ). **181**, 223 LP – 230 (1973).)

Additionally, the different conformation of the arms explains the known difficulties in sialylating the (1-6) arm by ST6-Gal1, relatively to the (1-3) arm<sup>28</sup>. Also, the different 3D conformational propensity of the arms in function of sequence can have important implications in terms of the N-glycans biosynthesis and biodegradation<sup>29</sup>. As an additional interesting point, we found that the folding of the (1-6) arm over the chitobiose region is completely independent of core  $\alpha(1-6)$  fucosylation<sup>24</sup>, with the result that core-fucosylated and non-core fucosylated N-glycans with the same sequence in the (1-6) arm correspond to the same structural ensemble.

In this work we discuss how core  $\alpha(1-3)$ -Fuc and  $\beta(1-2)$ -Xyl regulate the conformational propensity of the (1-6) arm to push a predominantly outstretched (open) conformation when the arms are functionalized with terminal  $\beta(1-3)$ -Gal. Within this framework, we explored the possibility of integrating these motifs in the context of mammalian sequences as an exploratory strategy towards the design of N-glycans with the desired 3D structure. For simplicity in the presentation and discussion of the results, we refer to N-glycans as either “plant” or “hybrid” separately. Nevertheless, it is important to underline that some of these motifs, such as  $\beta(1-2)$  xylosylation and difucosylated core are also found in invertebrate N-glycosylation<sup>30</sup>. Finally, we discuss these findings within a framework where the different N-glycoforms can be represented as a combination of spatial self-contained units, named “glycoblocks”, rather than in terms of monosaccharides and linkages. We find that this approach helps our understanding of N-glycans architecture in terms of equilibrium structures and relative populations and also of how specific modifications affect molecular recognition.





**Figure 4.1.** Representative structures of the plant N-glycans studied in this work with corresponding nomenclature. The letters f, x, and g indicate the presence of Fuc, Xyl and  $\beta(1-3)$  Gal, respectively, and ng the absence of  $\beta(1-3)$  Gal. LeA stands for Lewis A antigen. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The plants N-glycan characteristic linkages are indicated in the legend. Rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>).

## 4.2 Computational Methods

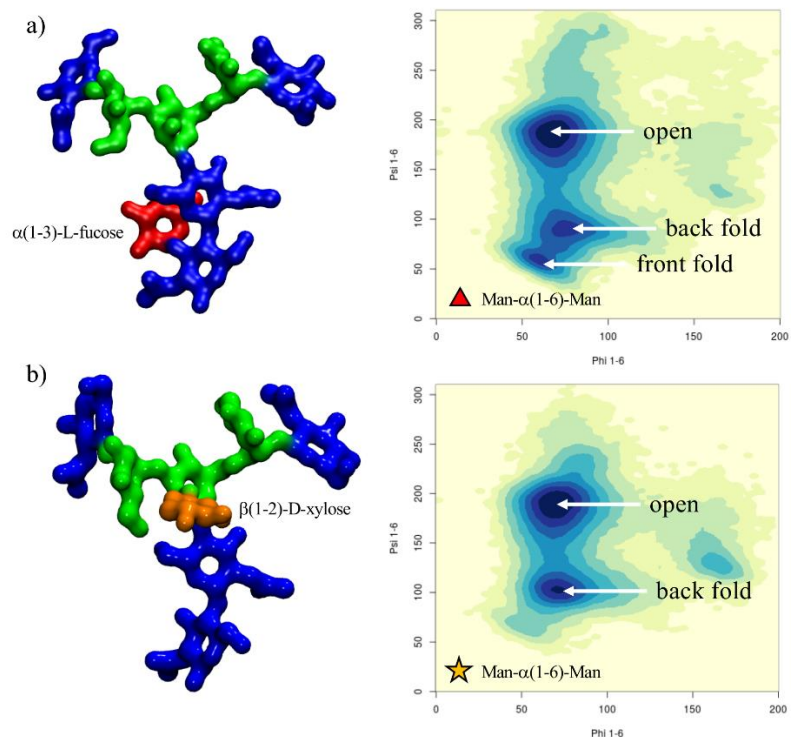
All starting structures were generated with the GLYCAM Carbohydrate Builder (<http://www.glycam.org>). For each sequence we selected the complete set of torsion angle values obtained by variation of the 1-6 dihedrals, namely the three *gg*, *gt* and *tg* conformations for each 1-6 torsion. The topology file for each structure was obtained using *tleap*<sup>31</sup>, with parameters from the GLYCAM06-j1<sup>32</sup> for the carbohydrate atoms and with TIP3P for water molecules<sup>33</sup>. All calculations were run with the AMBER18 software package<sup>31</sup> on NVIDIA Tesla V100 16GB PCIe (Volta architecture) GPUs installed on the HPC infrastructure *kay* at the Irish Centre for High-End Computing (ICHEC). Separate production steps of 500 ns each were run for each rotamer (starting system) and convergence was assessed based on conformational and clustering analysis, see **Appendix II** for all relevant tables. Simulations were extended, if the sampling was not deemed sufficient, i.e. in case standard deviation values measured were significantly larger than  $15^\circ$  for each cluster in each trajectory. All trajectories were processed using *cpptraj*<sup>31</sup> and visually analysed with the Visual Molecular Dynamics (VMD) software package<sup>34</sup>. Backbone Root Mean Square Deviation (RMSD) and torsion angles values were measured using VMD. A density-based clustering method was used to calculate the populations of occupied conformations for each torsion angle in a trajectory and heat maps for each

dihedral were generated with a kernel density estimate (KDE) function. Statistical and clustering analysis was done with the R package and data were plotted with RStudio ([www.rstudio.com](http://www.rstudio.com)). Further details on the simulation set-up and running protocol are included in **Appendix II**.

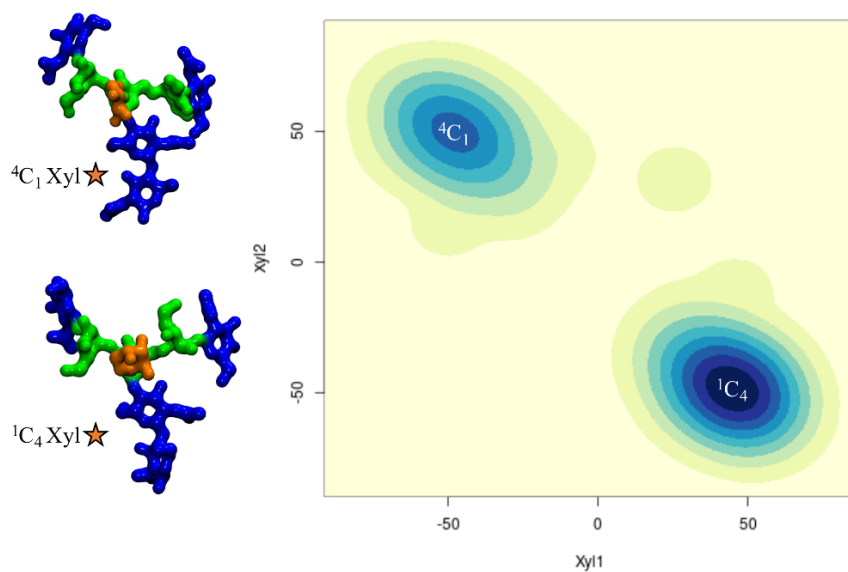
### 4.3 Results

#### **Core $\alpha(1-3)$ fucose in plant N-glycans.**

One distinctive feature of plants N-glycans is the occurrence of core fucosylation in  $\alpha(1-3)$ , rather than  $\alpha(1-6)$ -Fuc, normally found in mammalian N-glycans<sup>23,24</sup>. To understand the effects on the 3D structure of this modification, we have considered two biantennary systems, one terminating with  $\beta(1-2)$ -GlcNAc on both arms (*ngf*) and the other with terminal  $\beta(1-3)$ -Gal on both arms (*gf*), shown in **Figure 4.1**. In both glycoforms core  $\alpha(1-3)$ -Fuc occupies a stable position, with one single conformer populated (100%), see **Tables S.1 and S.2 in Appendix II**. This conformation is supported by a stacking interaction between the core  $\alpha(1-3)$ -Fuc and  $\beta(1-4)$  GlcNAc of the chitobiose in a “closed” conformation, which resembles the stable conformation of LeX<sup>35</sup>. This spatial arrangement imposes a 20° rotation of the GlcNAc- $\beta(1-4)$ -GlcNAc linkage, see **Tables S.1 and S.2 in Appendix II**, relative to the  $\alpha(1-6)$  core fucosylated or non-fucosylated chitobiose<sup>24</sup>, where the average psi value is -127.8° (14.8)<sup>24</sup>, but doesn’t affect the structure of the linkage to the central mannose. As shown by the low standard deviation values and by the lack of multiple minima (clusters), the N-glycan core remains relatively rigid throughout the trajectories. The slight torsion of the GlcNAc- $\beta(1-4)$ -GlcNAc linkage imposed by the  $\alpha(1-3)$ -Fuc has a dramatic effect on the conformational dynamics of the (1-6) arm, which is found predominantly in an outstretched (66%, cluster 1) conformation, rather than folded over (34%, clusters 1 and 2), see **Table S.1 in Appendix II**. The addition of a terminal  $\beta(1-3)$ -Gal in the *gf* N-glycan pushes the equilibrium towards an outstretched (1-6) arm even further, with the open conformation populated at 72%, see **Table S.2 in Appendix II**. Interestingly, in the case of  $\alpha(1-6)$  core fucosylated N-glycans, and with double fucosylation as discussed later on, the equilibrium of the (1-6) arm was the exact opposite, with a predominance of the folded conformation, especially in the presence of terminal  $\beta(1-4)$  Gal<sup>24</sup>. To note, the folded (1-6) arm conformation can be either a ‘front fold’, see **Figure 4.2 panel A**, where the torsion around the  $\alpha(1-6)$  linkage brings the arm towards the reader, or a ‘back fold’ where the (1-6) arm interacts with the  $\alpha(1-3)$ -Fuc, away from the reader. As shown in **Tables S.1 and S.2 in Appendix II**, the equilibrium of the (1-3) arm is not affected by core  $\alpha(1-3)$ -Fuc.



**Figure 4.2.** A representative structure of the non-galactosylated N-glycan with  $\alpha(1-3)$ -linked core fucose (ngf) is shown in **panel a)**, with on the right-hand side the corresponding heat map showing the conformations accessible to the (1-6) arm in terms of the phi/psi torsion angles. A representative structure of the non-galactosylated N-glycan with  $\beta(1-2)$ -linked xylose (ngx) is shown in **panel b)**, with on the right-hand side a heat map showing the conformations accessible to the (1-6) arm in terms of the phi/psi torsion angles. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>) and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).



**Figure 4.3.**  $\beta$ -D-xylose ring pucker analysis over 3  $\mu$ s of cumulative MD sampling of the *ngx* N-glycan. The two snapshots on the right-hand side are representative *ngx* conformations corresponding to the two different ring puckers. The Xyl1 and Xyl2 axis labels refer to the torsion angles C1C2C3C4 and C2C3C4C5, respectively. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>) and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

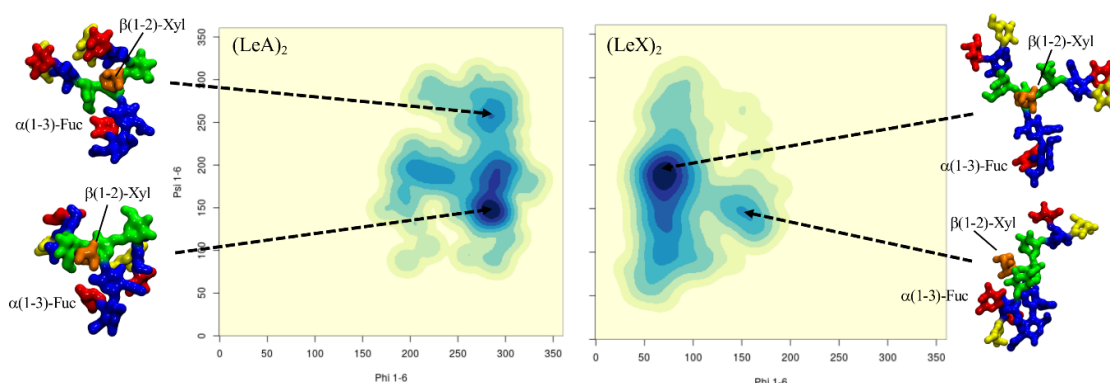
### **$\beta$ (1-2) xylose in plant N-glycans.**

Because the  $\beta$ (1-2)-Xyl sits in front of the two arms, it greatly affects their dynamics. Because of steric hindrance, the (1-3) arm is much more rigid relative to non-xylosylated species, see **Table S.3** in **Appendix II**, losing its “two conformer” dynamics characteristic of the biantennary mammalian N-glycans<sup>24</sup>, also retained in the plant N-glycans with only  $\alpha$ (1-3)-Fuc discussed above, see also **Tables S.1** and **S.2** in **Appendix II**. In regards to the (1-6) arm, as shown in **Figure 4.2 panel B**, the presence of  $\beta$ (1-2)-Xyl has a very similar effect as the  $\alpha$ (1-3)-Fuc, pushing the equilibrium towards an open conformation. To note, in the presence  $\beta$ (1-2)-Xyl, the (1-6) arm cannot fold over the chitobiose core in a ‘front fold’ either, because of steric hindrance. Also, similarly to the  $\alpha$ (1-3) fucosylated glycans, the stability of the open structure is slightly increased when the arm is further functionalized with terminal  $\beta$ (1-3)-Gal, see **Table S.4** in **Appendix II**. As an additional interesting feature, through the cumulative 3  $\mu$ s MD sampling, the xylose ring repeatedly inverts its conformation from the all equatorial  ${}^4C_1$  chair, to the  ${}^1C_4$  chair, where all hydroxyl groups are axial, see **Figure 4.3**. This transition may be energetically facilitated by the hydrogen bonding interaction xylose is able to form when in a  ${}^1C_4$  chair with the  $\alpha$ (1-6)-Man, which may compensate for the steric compression, making the  ${}^1C_4$  chair the highest populated conformer at 76% within an N-glycan scaffold. Both experimental and *ab-initio* theoretical studies<sup>36-38</sup> have shown that  ${}^1C_4$  chair is energetically accessible in isolated  $\beta$ -D-Xyl at room temperature in different dielectric conditions.

### **Core $\alpha$ (1-3) fucose and $\beta$ (1-2) xylose in plant N-glycans.**

The presence of both  $\alpha$ (1-3)-Fuc and  $\beta$ (1-2)-Xyl brings in the characteristic features highlighted earlier in the analysis of the structures with either  $\alpha$ (1-3)-Fuc or  $\beta$ (1-2)-Xyl. Indeed, we see here again the 20° rotation of the chitobiose GlcNAc- $\beta$ (1-4)-GlcNAc *psi* angle caused by the stacking of the  $\alpha$ (1-3)-Fuc to the chitobiose  $\beta$ (1-4)-GlcNAc and the conformational restraints imposed by the  $\beta$ (1-2)-Xyl on the (1-3) arm, see **Table S.5** in **Appendix II**. We also observed that both  $\alpha$ (1-3)-Fuc and  $\beta$ (1-2)-Xyl push the (1-6) arm equilibrium towards an open conformation, which is also the case when both are present in the *ngfx* N-glycan and to an even higher degree, i.e. 87%, in the *gfx* N-glycan, when both arms are

functionalized with terminal  $\beta(1-3)$ -Gal, see **Table S.6** in **Appendix II**. One feature specific to the *ngfx* N-glycan is the higher flexibility of the core Man- $\beta(1-4)$ -GlcNAc linkage, which allows for the rotation of the trimannose group relative to the chitobiose core. This conformation was accessible, but only populated around 2% when either  $\beta(1-2)$ -Xyl or  $\alpha(1-3)$ -Fuc are present, see **Tables S.1** to **S.4** in **Appendix II**. When both fucose and xylose are present, the population of the rotated trimannose reaches above 20%, see **Table S.5** in **Appendix II**, which can be considered a synergistic effect as this conformation is stabilized by a hydrogen bonding network involving the core fucose, the GlcNAc on the (1-6) arm and the xylose, as shown in **Figure S.1** in **Appendix II**. Such folding event has been observed as a stable conformation in two independent simulations. To note, the functionalization of the arms to include terminal  $\beta(1-3)$ -Gal reduces the occurrence of this event down to around 5%, see **Table S.6** in **Appendix II**.



**Figure 4.4.** Comparison of the different conformational equilibria of the (1-6) arm in a core  $\alpha(1-3)$ -Fuc  $\beta(1-2)$ -Xyl1 A2 N-glycan with terminal LeA and LeX groups on the left- and right-hand side, respectively. Representative structures from 1.5  $\mu$ s MD sampling of each system are shown to illustrate the conformations corresponding to the different minima. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>) and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

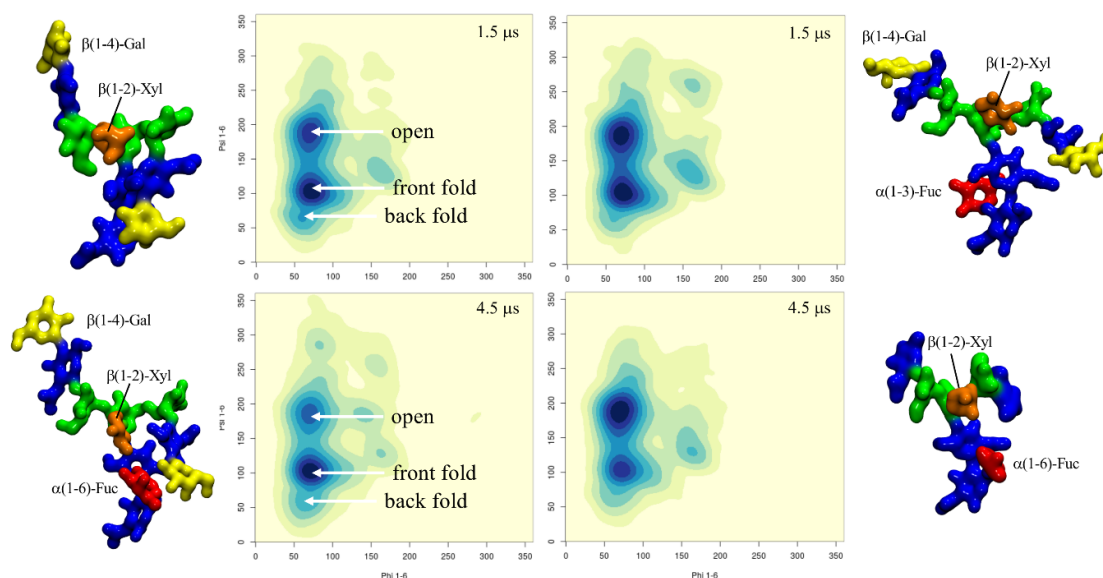
### Terminal LeA and LeX motifs in plant N-glycans.

To understand how an increased complexity on the arms would affect the dynamics of the  $\alpha(1-3)$  fucosylated and  $\beta(1-2)$  xylosylated N-glycans, we considered the functionalization with terminal LeA antigens present in plants N-glycans<sup>26</sup> and with LeX for comparison. As expected<sup>35</sup> the LeA and LeX structures are quite rigid, see **Tables S.7** and **S.15** in **Appendix II**, and remain in what is known as the “closed” conformation throughout the 1.5  $\mu$ s cumulative sampling time for each system. One interesting point is that the branching introduced by functionalizing the terminal GlcNAc residues with  $\alpha(1-4)$ -Fuc and  $\beta(1-3)$ -Gal, i.e. LeA, promotes the interaction between the two arms, which is not observed when the arms are linear, neither here for plants N-glycans, nor for mammalian IgG-type complex biantennary N-glycans<sup>24</sup>. The interaction between the arms is promoted by the ability to form complex hydrogen bonding networks, which in this specific case, may also involve the central xylose. As outcomes of the complex interaction the branched arms can establish, the equilibrium of the (1-6) arm is restrained in conformations previously not significantly populated, see **Figure 4.4** and **Table S.7** in **Appendix II**, and the GlcNAc- $\beta(1-2)$ -Man linkage in both arms is remarkably flexible, which is also not observed when the arms are not branched. Although not natural in plants, to check the corresponding symmetry, we built a core  $\alpha(1-3)$ -Fuc and  $\beta(1-2)$ -Xyl N-glycan with terminal LeX on both arms, a feature actually found in schistosome N-glycosylation<sup>30</sup>. Remarkably, as shown in **Figure 4.4** and **Table S.15** in **Appendix II**, within this framework the dynamics of the (1-6) arm is completely different. Contrary to

the N-glycan with terminal LeA groups, the two arms with LeX are not interacting and the (1-6) arm is predominantly (90%) in an extended (open) conformation, while the closed conformation, which accounts for the remaining 10% is achieved through a rotation around the core Man- $\beta$ (1-4)-GlcNAc. The lack of interaction between the arms is due to the inability to establish the same stable hydrogen bonding network due to the non-complementary position of the deoxy-C6 of the fucose in LeX relative to LeA.

### Hybrid N-glycans.

To understand how characteristic plant N-glycan motifs can affect the structure of mammalian N-glycoforms, we have designed and analysed the dynamics of a set of hybrid systems. In particular, we were interested in the effect of the addition of  $\beta$ (1-2)-Xyl and  $\alpha$ (1-3)-Fuc to (F)A2G2 N-glycans scaffolds in terms of potential alteration of the (1-6) arm dynamics.



**Figure 4.5.** Conformational analysis of the (1-6) arm in four hybrid N-glycoforms,  $\beta$ (1-2)-xylosylated A2G2 (top-left),  $\beta$ (1-2)-xylosylated FA2G2 (bottom-left),  $\beta$ (1-2)-xylosylated  $\alpha$ (1-3)-core fucosylated A2G2 (top-right) and  $\beta$ (1-2)-xylosylated FA2 (bottom-right). The predominant conformations are indicated in the top- and bottom-left heat maps for simplicity. The simulation time relative to each system is indicated in the top right corner of each heat map. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>) and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

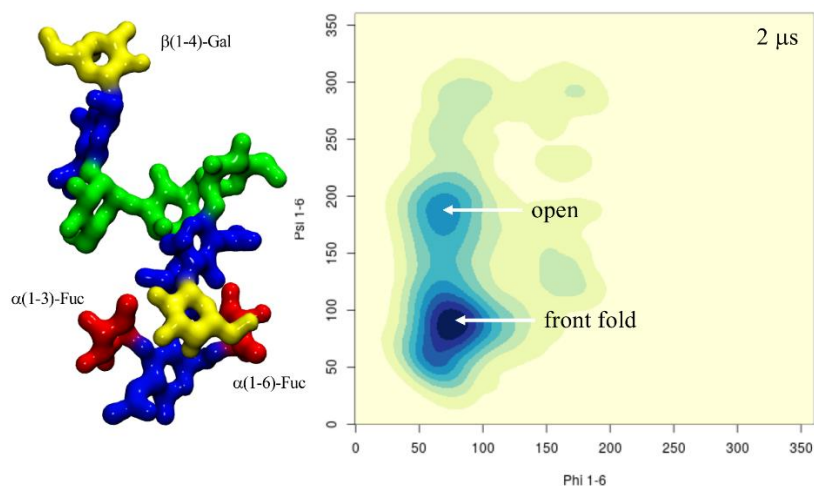
### **$\beta(1-2)$ -xylosylated mammalian N-glycans.**

Unlike the case of plants N-glycans, the presence of  $\beta(1-2)$ -Xyl hinders but does not completely prevent the (1-6) arm from folding over when the terminal galactose is  $\beta(1-4)$ -linked, as folding over the chitobiose can be stabilized by stacking, see **Figure 4.5** and **Table S.8** in **Appendix II**. The folded conformation with a median  $\psi$  value of  $103.5^\circ (\pm 11.3)$  is  $20^\circ$  from the average value of  $82.9^\circ$  calculated for the non-xylosylated (mammalian) counterpart<sup>24</sup>, so slightly distorted, and its population reduced from 74% to 57%. Nevertheless, the closed conformation is still the predominant form, even with  $\beta(1-2)$ -Xyl. The presence of  $\alpha(1-6)$ -linked core fucose to create a  $\beta(1-2)$ -xylosylated FA2G2, which is actually a type of N-glycosylation found in schistosoma<sup>30</sup>, brings in yet another change. As shown in **Figure 4.5** and **Table S.9** in **Appendix II**,  $\alpha(1-6)$ -Fuc and  $\beta(1-2)$ -Xyl are in an optimal conformation to support the closed (folded) (1-6) arm, by stacking of the terminal galactose by fucose and hydrogen bonding by xylose. Within this context the closed (1-6) arm is the highest populated conformer at 70.0% over 4.5  $\mu$ s of cumulative sampling of this system. To note that the conformation of the  $\alpha(1-6)$ -linked core fucose is the same as the one seen in mammalian N-glycans<sup>24</sup>, which on its own we have seen is not enough to affect the (1-6) arm equilibrium, see **Table S.9** in **Appendix II**. The interaction of the  $\alpha(1-6)$ -Fuc with the terminal  $\beta(1-4)$ -Gal is essential to promote the closed conformation of the (1-6) arm as demonstrated by the results obtained for the xylosylated FA2 systems, which recovers a conformational propensity similar to the non-fucosylated, xylosylated A2G2, see **Figure 4.5** and **Tables S.8** and **S.10** in **Appendix II**.

**$\alpha(1-3)$ -fucosylated mammalian N-glycans.** Because of its orientation tucked “behind” the chitobiose core defined in the context of plants N-glycans earlier, the effect of core  $\alpha(1-3)$ -Fuc on the (1-6) arm equilibrium within an A2G2-xylosylated scaffold is not as significant as  $\alpha(1-6)$ -Fuc. As shown in **Figure 4.5** and **Table S.11** in **Appendix II**, this lack of direct effect is demonstrated by the recovery of the same equilibrium as the non-fucosylated A2G2-xylosylated system. The dynamics of the chitobiose core is very similar to the one determined for the corresponding plant N-glycan. To analyse the effect of core  $\alpha(1-3)$  fucosylation without  $\beta(1-2)$ -Xyl, we have looked at two A2G2 hybrid systems, one with only  $\alpha(1-3)$ -linked fucose and one with both core  $\alpha(1-3)$ - and  $\alpha(1-6)$ -linked fucose, a characteristic “double-fucose” glycosylation found in worm and fly cells<sup>30</sup>. As shown in **Table S.12** in **Appendix II** unlike in plants N-glycans, the  $\alpha(1-3)$ -Fuc alone does not affect the A2G2 (1-6) arm equilibrium<sup>24</sup>, as the folding of the (1-6) arm with terminal  $\beta(1-4)$ -Gal is not obstructed by the rotation of the chitobiose core imposed by the  $\alpha(1-3)$ -Fuc position. When both  $\alpha(1-3)$ - and  $\alpha(1-6)$ -linked fucoses are present the (1-6) arm with terminal  $\beta(1-4)$ -Gal is predominantly folded (closed) at 85%, see **Figure 4.6** and **Table S.13** in **Appendix II**, which is higher than in the absence of  $\alpha(1-3)$ -Fuc<sup>24</sup>. Indeed, the latter can actively



contribute in stabilizing the interaction with the terminal  $\beta(1-4)$ -Gal of the folded (1-6) arm. We also observed interesting events, one representing 10% of 2  $\mu$ s as indicated by the values of the GlcNAc- $\beta(1-4)$ -GlcNAc torsion, where the GlcNAc is stacked in between the two fucose residues and another one, contributing to 18% of the simulation time, 14% when the system is also xylosylated, in which the GlcNAc ring transitions from  ${}^4C_1$  to  ${}^1C_4$  allowing the two fucose to stack, see **Tables S.13** and **S.14** and **Figure S.2** in **Appendix II**.






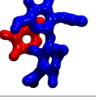

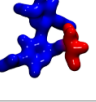

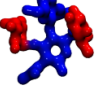

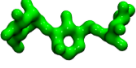

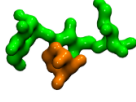

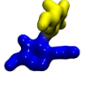

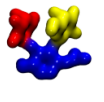

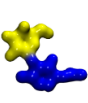

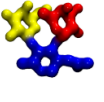
**Figure 4.6.** Conformational equilibrium of the (1-6) arm in terms of phi/psi torsion angle values for the  $\alpha(1-3)$ -fucosylated FA2G2 N-glycoform. The structure with the folded (1-6) arm where the terminal  $\beta(1-4)$ -Gal interacts with both fucose residues is shown on the left-hand side. The N-glycans structures are shown with the (1-3) and (1-6) arms on the left and on the right, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD (<https://www.ks.uiuc.edu/Research/vmd/>) and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

#### 4.4 Discussion

Differences and similarities in N-glycans sequences are highly cell-specific as well as important indicators of health and disease states<sup>1,39</sup>. Exogenous N-glycans motifs can be quite subtle, yet trigger profound differences in terms of molecular recognition<sup>19, 27</sup> and dangerous immunogenic responses<sup>20-22</sup>. In this work we have analysed the effects on the N-glycans structure and dynamics of two motifs in particular, namely  $\beta(1-2)$ -Xyl and core  $\alpha(1-3)$ -Fuc, common in plants<sup>23</sup> and invertebrates<sup>30</sup>, but completely absent in mammalian N-glycans. Within the context of plant-type N-glycans, which have a terminal  $\beta(1-3)$ -Gal, rather than  $\beta(1-4)$ -Gal, both  $\beta(1-2)$ -Xyl and  $\alpha(1-3)$ -Fuc contribute independently in promoting an outstretched (open) conformation of the (1-6) arm because of steric hindrance of the xylose and of the rotation forced upon the chitobiose core by the  $\alpha(1-3)$ -Fuc. The latter is not an

obstruction for the folding of a  $\beta(1-4)$ -Gal terminated (1-6) arm, as we have seen in the hybrid N-glycans constructs. Therefore, in  $\beta(1-2)$  xylosylated N-glycans terminating with  $\beta(1-3)$ -Gal, both arms should be more available for recognition, binding and further functionalization<sup>30</sup>, unlike in mammalian N-glycans where the  $\beta(1-4)$ -Gal determines a prevalently closed and inaccessible (1-6) arm<sup>24,27</sup>. Also, the analysis of the structure and dynamics of the LeA terminating plant N-glycans showed that the specific branching and spatial orientation of the motif allowed for a stable interaction between the arms, which is not observed in complex N-glycans with a linear functionalization of the arms<sup>24</sup>. Notably, the same hydrogen bonding network between the arms cannot be established when the same N-glycan terminates with LeX, because of the non-complementary position of the  $\alpha(1-3)$ -Fuc deoxy-C<sub>6</sub>.

The analysis of all these different complex N-glycoforms clearly shows that every modification, addition or removal of a specific motif, can greatly affect the 3D architecture of the N-glycan, thus its accessibility and complementarity to a receptor. However, these effects are rather complex to understand or to predict, if we think of the N-glycans 3D structure in terms of sequence of monosaccharides, a view that stems from the way we think about proteins. Our results show that the main effect of all functionalizations is actually local. For example, the core  $\alpha(1-3)$ -Fuc forces a rotation of the chitobiose, a degree of freedom very lowly populated otherwise; meanwhile,  $\beta(1-2)$ -Xyl restricts the flexibility of the trimannose core and occupies its centre. Within this framework, the 3D structural and dynamics features of the N-glycoforms can be rationalized by discretizing their architecture in terms 3D units, or “glycoblocks”, that group monosaccharides and their linkages within their immediate spatial vicinity, e.g. the core  $\alpha(1-3)$ -Fuc and the chitobiose which structure it has modified. A list of the glycoblocks that we have identified with the corresponding descriptors of their 3D features are listed in **Figure 4.7**. The whole N-glycan 3D architecture, in terms of the structures accessible and their conformational propensity, can be then described through the combination of these glycoblocks, together with the knowledge of their dynamic properties and flexibility. Also, consideration of these glycoblocks as spatial units can be useful to understand recognition by lectins and antibodies, which is often affected primarily by the targeted monosaccharide’s immediate vicinity and by its accessibility within a specific glycoform. For example, if we consider the 3D structure of the  $\beta(1-2)$ -Xyl Man3 glycoblock vs. the Man3 without Xyl, we can understand how the  $\beta(1-2)$ -Xyl position within that unit negates binding to DC-SIGN lectins<sup>19</sup>, see **Figure S.3 panels a and b in Appendix II**. Additionally, we can see that the slight rotation on the chitobiose imposed by the core  $\alpha(1-3)$ -Fuc does not prevent recognition and binding, see **Figure S.3 panel c. in Appendix II**.

Glycoblock	Unit 3D structure	Dominant motif descriptor	Characteristic glycosidic linkage
		Very stable conformation, with a low degree of flexibility around the equilibrium structure	<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b> phi = -78.7 (11.1)/100    psi = -130.8 (15.7)/99
		The $\alpha$ (1-3)-Fuc stacking interaction imposes a 20° rotation around the chitobiose glycosidic linkage and further stabilizes the chitobiose structure	<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b> phi = -72.1 (8.3)/100    psi = -107.1 (7.6)/100
		The longer glycosidic linkage provides the $\alpha$ (1-6)-Fuc with more degrees of freedom and does not alter the chitobiose structure	<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b> phi = -77.8 (11.2)/100    psi = -127.8 (14.8)/100
		This motif is structurally similar to the one with the $\alpha$ (1-3)-Fuc, which acts as the dominant restraint. The presence of the $\alpha$ (1-6)-Fuc increases flexibility	<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b> phi = -73.9 (9.5)/100    psi = -106.4 (14.1)/90
		Joint with differently flexible arms. The $\alpha$ (1-6) torsion has two dominant, different conformations accessible that can open and close the (1-6) arm. Populations depend on the attached glycoblock units	<b>Man <math>\alpha</math>(1-6) Man</b> (with terminal $\beta$ (1-4)-Gal) <b>open</b> phi = 76.3 (15.0)/100    psi = -185.1 (22.1)/23 <b>closed</b> phi = 76.3 (15.0)/100    psi = 85.3 (15.8)/76
		The addition of the $\beta$ (1-2)-Xyl further restricts the (1-3) arm and hinders the closing forward of the (1-6) arm, which is only favored in (F)A2G2 hybrids	<b>Man <math>\alpha</math>(1-6) Man</b> (with terminal $\beta$ (1-3)-Gal) <b>open</b> phi = 70.5 (10.4)/100    psi = -173.6 (16.3)/70 <b>closed</b> phi = 70.5 (10.4)/100    psi = 103.8 (12.7)/26
		This unit is rather rigid, while the orientation of the terminal Gal in $\beta$ (1-3/4) is determinant for the interaction with the chitobiose in the closed (1-6) arm. These units in the two arms do not interact	<b>Gal <math>\beta</math>(1-3) GlcNAc</b> phi = -72.9 (11.2)/100    psi = -124.0 (18.2)/96
		The Lewis A epitope is stable in its closed conformation and its presence greatly affects the (1-6) arm conformation because branching promotes association of the arms	<b>Man <math>\alpha</math>(1-6) Man</b> (with terminal LeA and $\alpha$ (1-3)-Fuc) <b>open</b> phi = -70.1 (5.7)/100    psi = -177.1 (16.3)/28 <b>closed</b> phi = -70.1 (5.7)/100    psi = 148.2 (7.4)/54
		This unit is rather rigid, while the orientation of the terminal Gal in $\beta$ (1-3/4) is determinant for the interaction with the chitobiose in the closed (1-6) arm. These units in the two arms do not interact	<b>Gal <math>\beta</math>(1-4) GlcNAc</b> phi = -76.2 (15.1)/100    psi = -125.7 (15.4)/97
		The Lewis X epitope is stable in its closed conformation. Contrary to Lewis A its branching does not favor association of the arms because of the relative positions of the deoxy-C <sub>6</sub> .	<b>Man <math>\alpha</math>(1-6) Man</b> (with terminal LeX, core $\alpha$ (1-3)-Fuc and $\beta$ (1-2)-Xyl) <b>open</b> phi = 70.7 (9.2)/90    psi = -173.6 (13.4)/90 <b>closed</b> phi = 152.4 (12.9)/10    psi = 145.6 (12.3)/10

**Figure 4.7.** List of 3D structural units of monosaccharides (glycoblocks) that regulate the 3D architecture and dynamics of complex biantennary N-glycans from plants and invertebrate sources and hybrid mammalian constructs. The SFNG representation of each glycoblock is indicated in the first column from the left, 3D structures from the highest populated conformers are shown in the second column, rendered with VMD (<https://www.ks.uiuc.edu/Research/vmd/>). A brief summary of the conformational features of each glycoblock and the characteristic linkage or its effect on the (1-6) arm conformation are indicated in the last two columns, respectively.

## 4.5 Conclusions

In this work we used extensive sampling through MD simulations to study the effects on the N-glycan architecture of subtle, yet highly consequential modifications, namely core  $\alpha(1-3)$ -Fuc and  $\beta(1-2)$ -Xyl<sup>19</sup>. These are part of standard N-glycoforms found in plants<sup>23</sup> and invertebrates<sup>30</sup>, but immunogenic in humans<sup>21, 22, 26</sup>. Our results show that these modifications can greatly affect the 3D structure of the N-glycan and its structural dynamics, therefore its selective recognition by lectin receptors and antibodies. The atomistic-level of detail information that the MD simulations provide us with, highlights that the effects of different functionalizations, in terms of monosaccharide types and linkages, are primarily local, affecting the immediate spatial vicinity of the monosaccharide within the N-glycan structure. Within this framework, we propose an alternative approach that can help describe and predict the architecture of N-glycans based on the combination of structural 3D units, or glycoblocks. Unlike a description based on monosaccharide sequence and linkages as two separate features, the transition to well-defined and self-contained units, integrating information on both monosaccharides and linkages, can help us rationalize and deconvolute the glycans structural disorder and ultimately understand more clearly the relationships between sequence and structure in complex carbohydrates.

## References

1. A. Varki, Biological roles of glycans. *Glycobiology* **27**, 3-49 (2017).
2. P. Stanley, Taniguchi N, M. Aebi, in *Essentials of Glycobiology, 3rd Edition*. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press 2015-2017, 2017), chap. 9
3. B. A. Cobb, The history of IgG glycosylation and where we are now. *Glycobiology* **30**, 202-213 (2020).
4. C. Ferrara *et al.*, Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fcγ<sub>3</sub> and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669-12674 (2011).
5. C. J. Day *et al.*, Glycan:glycan interactions: High affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells. *Proc Natl Acad Sci U S A* **112**, E7266-7275 (2015).
6. M. Nagae, Y. Yamaguchi, Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* **13**, 8398-8429 (2012).
7. L. F. Zacchi, B. L. Schulz, N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj J* **33**, 359-376 (2016).
8. M. L. A. De Leoz *et al.*, NIST Interlaboratory Study on Glycosylation Analysis of Monoclonal Antibodies: Comparison of Results from Diverse Analytical Methods. *Mol Cell Proteomics* **19**, 11-30 (2020).
9. Y. Mimura *et al.*, Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* **9**, 47-62 (2018).
10. M. Thaysen-Andersen, N. H. Packer, Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim Biophys Acta* **1844**, 1437-1452 (2014).
11. D. Alocci *et al.*, GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical. *J Proteome Res* **18**, 664-677 (2019).
12. F. Lisacek *et al.*, Databases and Associated Tools for Glycomics and Glycoproteomics. *Methods Mol Biol* **1503**, 235-264 (2017).
13. J. Mariethoz *et al.*, Glycomics@ExPASy: Bridging the Gap. *Molecular & Cellular Proteomics* **17**, 2164-2176 (2018).
14. K. Aoki-Kinoshita *et al.*, GlyTouCan 1.0--The international glycan structure repository. *Nucleic Acids Res* **44**, D1237-1242 (2016).
15. M. A. Rojas-Macias *et al.*, Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nat Commun* **10**, 3275 (2019).

16. J. D. Durrant *et al.*, Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent Sci* **6**, 189-196 (2020).
17. I. Yu *et al.*, Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* **5**, (2016).
18. K. Lindorff-Larsen, P. Maragakis, S. Piana, D. E. Shaw, Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J Phys Chem B* **120**, 8313-8320 (2016).
19. K. Brzezicka *et al.*, Synthesis and microarray-assisted binding studies of core xylose and fucose containing N-glycans. *ACS Chem Biol* **10**, 1290-1302 (2015).
20. I. B. Wilson, J. E. Harthill, N. P. Mullin, D. A. Ashford, F. Altmann, Core alpha1,3-fucose is a key part of the epitope recognized by antibodies reacting against plant N-linked oligosaccharides and is present in a wide variety of plant extracts. *Glycobiology* **8**, 651-661 (1998).
21. R. van Ree *et al.*, Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **275**, 11451-11458 (2000).
22. M. Bardor *et al.*, Immunoreactivity in mammals of two typical plant glyco-epitopes, core alpha(1,3)-fucose and core xylose. *Glycobiology* **13**, 427-434 (2003).
23. R. Strasser, Plant protein glycosylation. *Glycobiology* **26**, 926-939 (2016).
24. A. M. Harbison, L. P. Brosnan, K. Fenlon, E. Fadda, Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94-103 (2019).
25. A. Varki *et al.* (2015).
26. A. C. Fitchette-Lainé *et al.*, N-glycans harboring the Lewis a epitope are expressed at the surface of plant cells. *Plant J* **12**, 1411-1417 (1997).
27. B. Echeverria *et al.*, Chemoenzymatic Synthesis of N-glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-type Lectin Receptors. *ACS Chem Biol* **13**, 2269-2279 (2018).
28. A. W. Barb, E. K. Brady, J. H. Prestegard, Branch-specific sialylation of IgG-Fc glycans by ST6Gal-I. *Biochemistry* **48**, 9705-9707 (2009).
29. U. Moginger *et al.*, Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Frontiers in Oncology* **8**, (2018).
30. K. Paschinger, I. B. H. Wilson, Comparisons of N-glycans across invertebrate phyla. *Parasitology* **146**, 1733-1742 (2019).
31. D. Case *et al.* (University of California, San Francisco, 2018).
32. K. N. Kirschner *et al.*, GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622-655 (2008).

33. W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, M. Klein, Comparison of simple potential functions for simulations of liquid water. *Journal of Chemical Physics* **79**, 926-935 (1983).
34. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).
35. I. Alibay, K. Burusco, N. Bruce, R. Bryce, Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *Journal of Physical Chemistry B* **122**, 2462-2474 (2018).
36. H. B. Mayes, L. J. Broadbelt, G. T. Beckham, 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* **136**, 1008-1022 (2014).
37. J. Rönnols, S. Manner, A. Siegbahn, U. Ellervik, G. Widmalm, Exploration of conformational flexibility and hydrogen bonding of xylosides in different solvents, as a model system for enzyme active site interactions. *Org Biomol Chem* **11**, 5465-5472 (2013).
38. J. Iglesias-Fernández, L. Raich, A. Ardèvol, C. Rovira, The complete conformational free energy landscape of  $\beta$ -xylose reveals a two-fold catalytic itinerary for  $\beta$ -xylanases. *Chem Sci* **6**, 1167-1177 (2015).
39. C. Reily, T. J. Stewart, M. B. Renfrow, J. Novak, Glycosylation in health and disease. *Nat Rev Nephrol* **15**, 346-366 (2019).

## Chapter 5: An Atomistic Perspective on Antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 N-glycans from enhanced sampling molecular dynamics

**Paper citation:** Harbison, A. & Fadda, E. An atomistic perspective on antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics. *Glycobiology* 30, 407–414 (2020).

### 5.1 Introduction

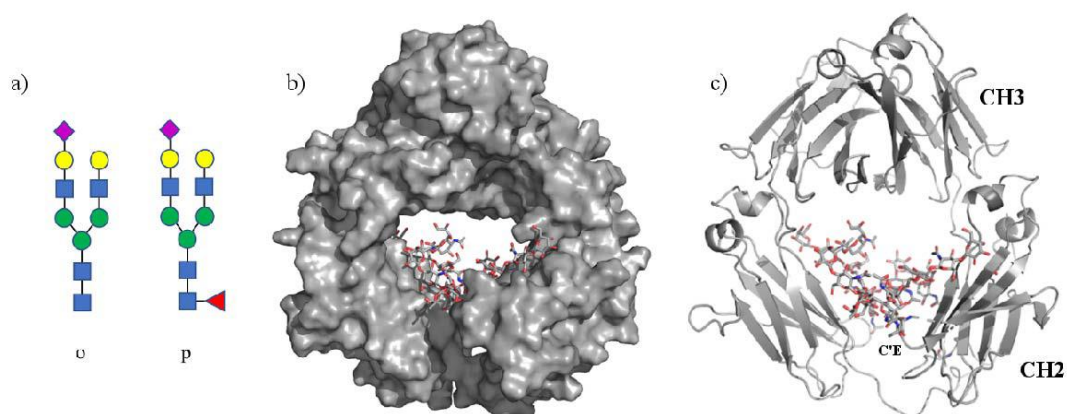
Immunoglobulins type G (IgGs) are the most abundant antibodies in human serum<sup>1</sup>. Their ability to trigger an effective immune response is dependent on their interaction with the fragment crystallizable (Fc)  $\gamma$  receptors (Fc $\gamma$ Rs) bound to the outer membrane of immune system effector cells<sup>2</sup>. The interaction between IgGs and Fc $\gamma$ Rs type III (Fc $\gamma$ RIII or CD16) specifically triggers an antibody-dependent cell-based cytotoxicity (ADCC) response<sup>3</sup> that leads to the destruction of a targeted cell. Because ADCC is considered as the main antibody-based mechanism against tumour cells, strategies aimed at regulating ADCC are highly sought after, specifically within the framework of the antibody engineering of cancer therapeutics<sup>4</sup>. The IgGs/Fc $\gamma$ Rs binding specificity and affinity depend not only on the specific amino acids in the IgG1 CH2 region in direct contact with the Fc $\gamma$ R<sup>5</sup>, but also on the type of glycosylation on both the IgG1 Fc and the Fc $\gamma$ R<sup>6-8</sup>. More specifically, both CH2 domains of human IgGs are glycosylated at Asn 297 with complex biantennary N-glycans, 96% of which are core-fucosylated and 60% terminating with one or two galactose residues<sup>9</sup>. Around 20% of the Fc N-glycans are sialylated<sup>9</sup> with one terminal sialic acid preferentially on the  $\alpha$ (1-3) arm<sup>10</sup>, and di-sialylation only occurring 4% of the time<sup>11</sup>. Decreasing IgG1 Fc galactosylation levels have been linked to autoimmune conditions, such as rheumatoid arthritis<sup>12</sup>, and both galactosylation and especially sialylation are known to trigger an anti-inflammatory response<sup>13</sup>. The mechanism(s) linking the Fc N-glycans sequence to IgG1 function are unclear. Nevertheless, it is very likely that the molecular basis for these effects resides in how the Fc N-glycoforms modulate the IgG/Fc $\gamma$ Rs recognition and binding affinity.

The crystal structures of the complex between an IgG1 and Fc $\gamma$ RIII give important insight into this matter, showing that the Fc and Fc $\gamma$ RIII N-glycans form an intricate network of interactions upon



binding<sup>14, 15</sup>. These carbohydrate-carbohydrate contacts are important in regulating the binding affinity of the complex<sup>6, 14-16</sup>. Indeed, the comparison between two crystal structures of the IgG1/Fc $\gamma$ RIII complex, one where the IgG1 N-glycan is core-fucosylated and the other where it is not<sup>14</sup>, shows that core-fucose hinders the carbohydrate-carbohydrate interaction, displacing the Fc N-glycans by 2.6 Å in comparison to the structure with the non-fucosylated Fc-linked N-glycan<sup>14, 15</sup>. Because of this steric hindrance, a less effective binding network can be formed between the two glycans, with a consequent reduction of the complex binding affinity. Although the effectiveness of the glycans interaction is most likely a contributing factor to the reduction of the ADCC by core-fucosylation, carbohydrate-carbohydrate binding affinities are known to be very weak<sup>17, 18</sup>, thus a slight change in the enthalpic contribution due to looser N-glycans contacts is rather difficult to reconcile with the 100-fold ADCC reduction observed in the presence of a core-fucosylated Fc N-glycan<sup>19-22</sup>. Furthermore, it does not explain how sialylation can decrease ADCC but only in the context of core-fucosylation<sup>21</sup>. Meanwhile, it is important to note that sialylation does not change the affinity of core-fucosylated Fc for CD16s and CD32s and only increases the affinity of a-fucosylated Fc slightly<sup>32</sup>. Also, carbohydrate-carbohydrate interactions reach significant binding strengths only when large scale cooperativity is involved<sup>23</sup>.

Therefore, to gain further insight in the role of the Fc N-glycans sequence, structure and dynamics in the IgGs function, here we present the results of a molecular dynamics simulation study of IgG1 Fc domains with specific N-glycans, shown in **Figure 5.1**, known to be significantly populated in human IgG1s<sup>9</sup>. Extensive sampling through temperature replica exchange molecular dynamics (REMD)<sup>24</sup> was chosen as the method to explore exhaustively the potentially rugged conformational space. The N-glycoforms were chosen specifically to address the three following points, 1) to investigate a potential link between the structure and dynamics of core-fucosylated N-glycans to the dramatic ADCC reduction, 2) to determine how the preferential conformation and dynamics of a sialylated  $\alpha(1-3)$  arm may affect ADCC in context of core-fucosylation<sup>21</sup> and finally 3) to analyse the dynamics of the galactosylated  $\alpha(1-6)$  arm in Fc-linked relative to the free N-glycans.



**Figure 5.1.** The two different types of biantennary complex N-glycans we have considered in this study are shown in panel a). For simplicity, we named the core-fucosylated N-glycan sugar *p* and the non-fucosylated N-glycan sugar *o*. A representative structure from one of the REMD simulations of the IgG1 Fc with the linked N-glycans is shown in panel b) with the protein represented as solvent accessible surface area and in panel c) with cartoon rendering, where the CH3 and CH2 domains are labelled, together with the CH2 C'E loop carrying the N297-linked N-glycan.

The simulations results show that core-fucosylation affects dramatically the dynamics of the  $\alpha(1-3)$  arm of the opposed N-glycan and as a consequence of steric hindrance it significantly enhances the dynamics of the Fc domain. Also, structural alignment of our systems to available crystallographic data show that the core-fucose position within the Fc core obstructs the entry of N162 N-glycosylated Fc $\gamma$ Rs. This suggests that binding a core-fucosylated IgG1 involves a more complex conformational displacement of the CH2 domain. Within the framework of this mechanism, not only the CH2 domain needs to “open” in order to accommodate the incoming Fc $\gamma$ R, but also that the core fucosylated N-glycan has to undergo a more complex conformational change, to create the space for binding and to allow proper contact with the the Fc $\gamma$ R N162 glycan. This argument is consistent with a higher energetic cost and a lower binding affinity. Furthermore, we find that at room temperature (300 K) the terminal galactose in the  $\alpha(1-6)$  arm is heavily restrained to the CH2 domain of the protein, which promotes an outstretched conformation as the only significantly populated conformer. In the following sections we will present and discuss these results in detail within the framework of the available experimental data and of the known evidence of the N-glycosylation dependence of IgG1 effector functions.

## 5.2 Computational Methods

Protein, counterions and carbohydrate atoms were represented with the ff14SB<sup>25</sup> and GLYCAM06j-1<sup>26</sup> forcefields, respectively, while the TIP3P model<sup>27</sup> was used to represent the solvent. The total electrostatic charge of the system was neutralized by adding Cl<sup>-</sup> ions. All simulations were carried out using NAMD version 2.31b1<sup>28</sup>. For the temperature Replica Exchange Molecular Dynamics (REMD)<sup>24</sup>, 90 replicas were generated in the temperature range between 300 and 500 K, corresponding to an exchange probability of 0.13. The simulations were carried out on six IgG1 Fc models in total, one with two core-fucosylated N-glycans (*pp*), namely sugar *p* shown in **Figure 5.1**, one with two non-fucosylated N-glycans (*oo*) and another with one core-fucosylated and one non-fucosylated N-glycan (*op*). Because the molecular crowding within the Fc core could limit the conformational space even within an enhanced sampling scheme, in order to explore the possibility of the folding (closing) of the galactosylated  $\alpha(1-6)$  arm we described in previous work for the corresponding unlinked glycans<sup>29</sup>, we built all starting structures with  $\alpha(1-6)$  arms both in open and closed conformations, see **Figure S.1** in

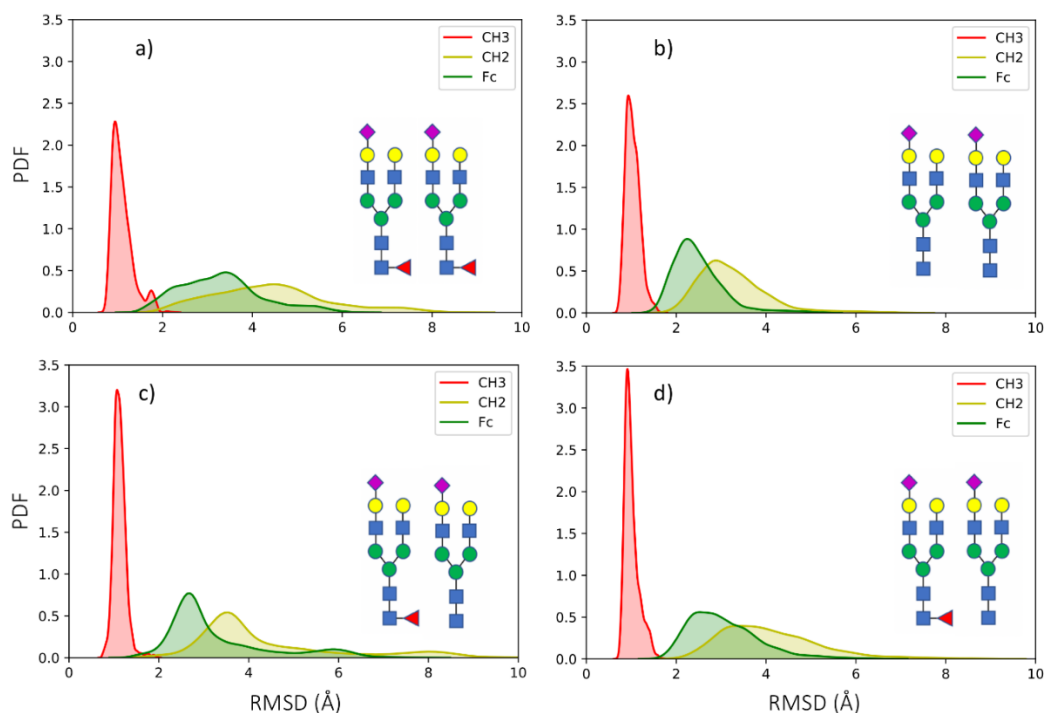
**Appendix III.** The data analysis was done on ensembles collected from 13 and 11 ns simulations per replica. Further information on the simulation set-up and running protocol is included in **Appendix III.**

### 5.3 Results

The results are presented in separate sections for clarity. Unless stated otherwise, all results refer to simulations started with the N-glycans  $\alpha(1-6)$  arm in the open (outstretched) conformation.

#### **Protein and Fc-linked N-Glycans Dynamics.**

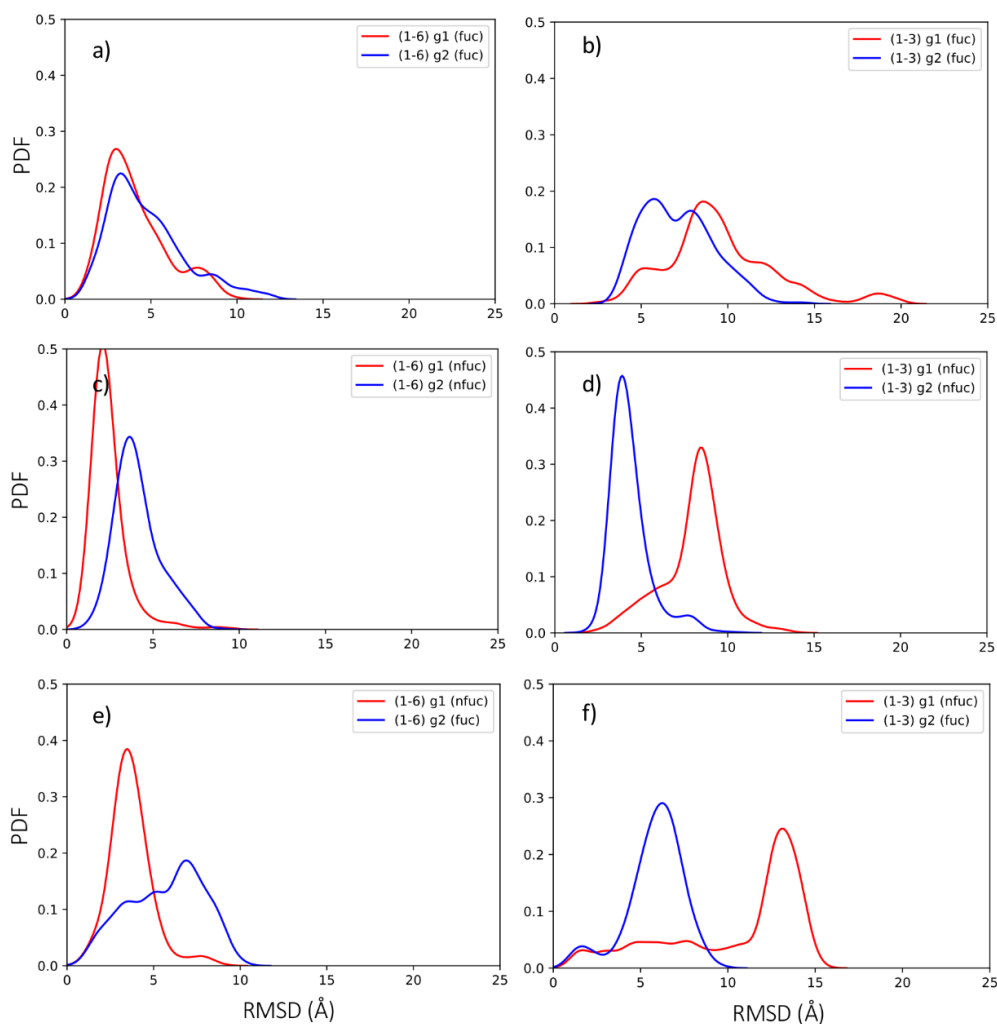
The IgG1 Fc domains flexibility and its overall dynamics has been evaluated in terms of backbone RMSD values, calculated from the simulations relative to the 1.9 Å resolution crystal structure (PDBid 4DZ8) of a N-glycosylated and core-fucosylated IgG1 Fc<sup>30</sup>. The structural alignment was done on the two symmetric CH3 domains alone for all simulations as it is the most stable part of the IgG1 Fc. Results are shown in **Figure 5.2** in terms of kernel density estimation (KDE) plots of the protein backbone RMSD values obtained from simulations of an IgG1 Fc linked to two core-fucosylated sugar *p* on both sides (*pp*), see **Figure 5.2 panel A**, an IgG1 Fc linked to two non-fucosylated sugar *o* on both sides (*oo*), see **Figure 5.2 panel B**, and to an IgG1 Fc with one sugar *o* and one sugar *p* on each side (*op*), see **Figure 5.2 panels C and D**. The average backbone RMSD values for all the systems studied are shown in **Table S.1** in **Appendix III**. The results indicate that core-fucosylation significantly enhances the protein dynamics, with a maximum effect when both Fc N-glycans are fucosylated, see **Figure 5.2 panel A**. As it will be discussed in detail further below, this enhanced dynamic is a consequence of the steric interaction between the core-fucose of one N-glycan and the  $\alpha(1-3)$  arm of the symmetrically opposed N-glycan. The KDE distributions also indicate that the protein dynamics is primarily determined by the CH2 domain, which is linked to the more rigid CH3 domain by an unstructured loop, see **Figure 5.1 panel c**. that allows for a flexible architecture. The stability of the CH3 domain is not affected by the presence of core-fucose, nor by the conformational dynamics of the arms. The results obtained for the *op* IgG1 Fc show that the presence of even one core-fucose increases the dynamics of the CH2 domain, even if to a lesser degree than two core-fucosylated N-glycans. The CH2 linked to the non-fucosylated sugar *o*, see **Figure 5.2 panel C**, appears slightly more dynamic, as a consequence of the steric hindrance between sugar *o*'s  $\alpha(1-3)$  arm and the core-fucose on the facing sugar *p*.



**Figure 5.2.** Kernel density estimation (KDE) plots are shown here to represent the probability density distribution of the protein backbone RMSD values. These are calculated throughout the simulations started from open conformations of the  $\alpha(1-6)$  arms, in panel a) on an IgG1 Fc with two core-fucosylated sugars  $p$ , in panel b) with two non-fucosylated sugar  $o$ , and with one core-fucosylated sugar  $p$  and one non-fucosylated sugar  $o$  in panels c) and d), where the two CH3-CH2 domains are represented separately for the non-fucosylated and for the core fucosylated Fc-linked CH2 in panels c) and d), respectively. The probability density function (PDF) is on the y axis and the root mean square deviation (RMSD) values on the x axis. The plots are. Based on 6500 points for panel a), and 5500 points for panels b), c) and d).

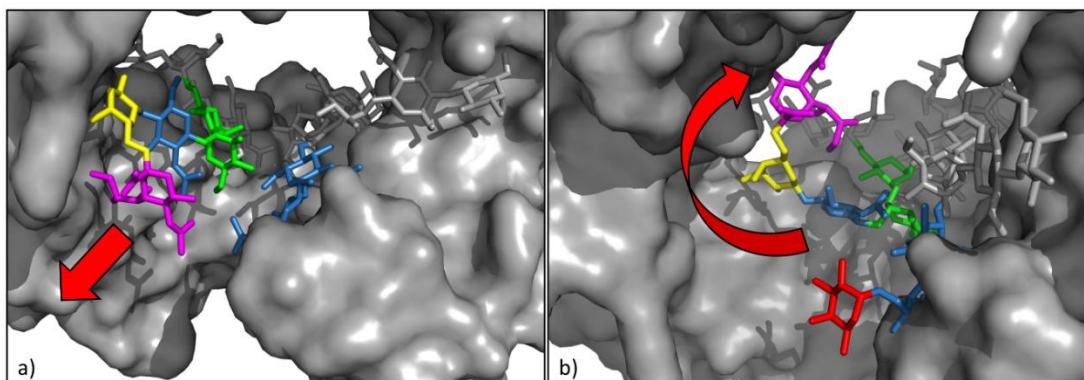
In regards to the dynamics  $\alpha(1-6)$  and  $\alpha(1-3)$  arms, the RMSD values distributions are shown in **Figure 5.3** for the  $pp$ ,  $oo$  and  $op$  IgG1 Fcs. The distributions show that the presence of core-fucosylation enhances the dynamics of the whole N-glycans, but especially of the  $\alpha(1-3)$  arms. This enhanced flexibility is due to the steric hindrance between the core-fucose of one N-glycan and the extended sialylated  $\alpha(1-3)$  arm of the other. This clash pushes the  $\alpha(1-3)$  arm to interconvert between an extended conformation, which requires the opening of the CH2 domains in order to fit, and a bent conformation, where the sialic acid points towards the core of the Fc instead of towards the bulk water, see **Figure 5.4**. The extended  $\alpha(1-3)$  arm is the preferential conformation seen during the simulation of the  $oo$  system and corresponds to the maxima in **Figure 5.3 panel D**, where none of the glycans are core-fucosylated. As a note, the shift in the RMSD values observed for glycan 1 (g1) in **Figure 5.3 panel D** (red line) depends on the fact that the RMSD values are calculated relative to the starting conformation, which in this specific case was a low populated one where the  $\alpha(1-3)$  arm is slightly bent interacting with the opposite N-glycan's chitobiose core. The results shown in **Figure 5.3 panels E and F**, obtained

for the mixed *op* IgG1 Fc system are quite interesting in that they show that the  $\alpha(1-3)$  arm of the non-fucosylated glycan g2 is more dynamic than the  $\alpha(1-3)$  arm of the fucosylated g1, as it does interact with the g1 core-fucose. Also, the dynamics of  $\alpha(1-6)$  arm of the core-fucosylated glycan g1 is directly affected by this interaction. The corresponding average RMSD values calculated for the N-glycans heavy atoms are shown in **Table S.2** in **Appendix III**.



**Figure 5.3.** Kernel density estimation (KDE) plots are shown here to represent of the RMSD values distributions calculated for the  $\alpha(1-6)$  and  $\alpha(1-3)$  arms (heavy atoms) during the simulations of the IgG1 Fc with two fucosylated N-glycans (pp) in panels a) and b), and with two non-fucosylated N-glycans (oo) in panels c) and d) and with one fucosylated and one non-fucosylated N-glycans (op) in panels e) and f). The red and blue lines refer to the two different glycans g1 and g2, respectively.

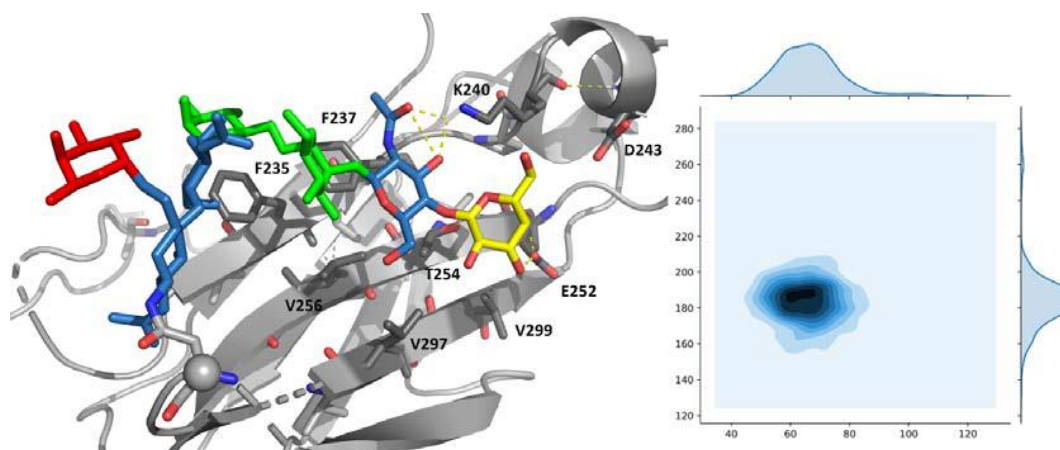
The folding backwards of the  $\alpha(1-3)$  arm towards the inside of the Fc core promoted by the core-fucose of the opposite glycan does not involve changes from the equilibrium torsion angle values we have determined for the unlinked (free) N-glycans in earlier work<sup>29</sup>. As shown in **Table S.3** and **Figure S.2** in **Appendix III** the relative populations and values of the torsion angles for the  $\alpha(1-3)$  arm are quite similar to the ones we determined in solution<sup>29</sup>. The only significant difference is the increase of the population of the  $\alpha(1-3)$   $\psi = 96^\circ$  (15) torsion to 55%, corresponding to the value in solution  $\psi = 102^\circ$  (14) at 39%<sup>29</sup>, which corresponds to the bent conformation, see **Table S.3** in **Appendix III**.



**Figure 5.4.** Different orientations of the sialylated  $\alpha(1-3)$  arm in the absence and presence of core-fucose on the opposite N-glycan. In panel a) the snapshot from the *oo* IgG1 Fc simulation shows the  $\alpha(1-3)$  arm outstretched towards the solvent, meanwhile in panel b) the snapshot from the *pp* IgG1 Fc simulation shows the  $\alpha(1-3)$  arm folded over the Fc core obstructed by the core-fucose of the opposite N-glycan. The important residues of the N-glycans are highlighted according to the SNFG nomenclature, the protein is represented through a solvent accessible surface in grey. Rendering was done with *pyMol*.

#### N-Glycans interactions within the Fc core.

In all simulation started from an open conformation, the  $\alpha(1-6)$  arm remains outstretched over the CH2  $\beta$  sheet and restrained to it through an extensive and organized architecture of interactions with both hydrophobic and hydrophilic residues. We did not observe any unbinding events at room temperature (300 K). As shown in **Figure 5.5**, residues Lys 240 can interact with the carbonyl oxygen of the  $\alpha(1-6)$  GlcNAc, while Glu 252, Asp 243 and Thr 254 are all engaging with the  $\alpha(1-6)$  terminal Gal. We have observed throughout all simulations that Lys 240 and Asp 243 intermittently engage in a salt-bridge interaction, partially releasing the  $\alpha(1-6)$  arm. Residues Phe 237, Val 256, Val 297 and Val 299 line-up to form a hydrophobic patch that supports by stacking the  $\alpha(1-6)$  arm movement across the CH2 domain. This outstretched conformation is the only one populated in all simulations started from an open  $\alpha(1-6)$  arm conformation, as shown by the heat map in **Figure 5.5**. Because of the complexity of the system, as explained in the Computational Method section all simulations were also started from a closed  $\alpha(1-6)$  arm conformation, see **Figure S.1** in **Appendix III**, where the  $\alpha(1-6)$  arm is folded over the chitobiose core as seen for the corresponding N-glycans in solution<sup>29</sup>. Because of the molecular crowding and the limited space available due to the partial collapse of the Fc core, in none of the simulations started from a closed  $\alpha(1-6)$  arm conformation the N-glycan opens.



**Figure 5.5.** On the left-hand side, a representative image of the highest populated conformer observed for the  $\alpha(1-6)$  arm in the simulations of both core-fucosylated (sugar *p*, represented here) and also of non-fucosylated sugar *o*. The  $\alpha(1-3)$  arm and the rest of the protein are not represented for clarity and the residues are coloured according to the SNFG nomenclature. On the right-hand side, the corresponding phi/psi Ramachandran plot of the  $\alpha(1-6)$  torsion calculated from the REMD simulation of the *pp* IgG1 Fc system, showing that only the extended conformer is populated at 300 K.

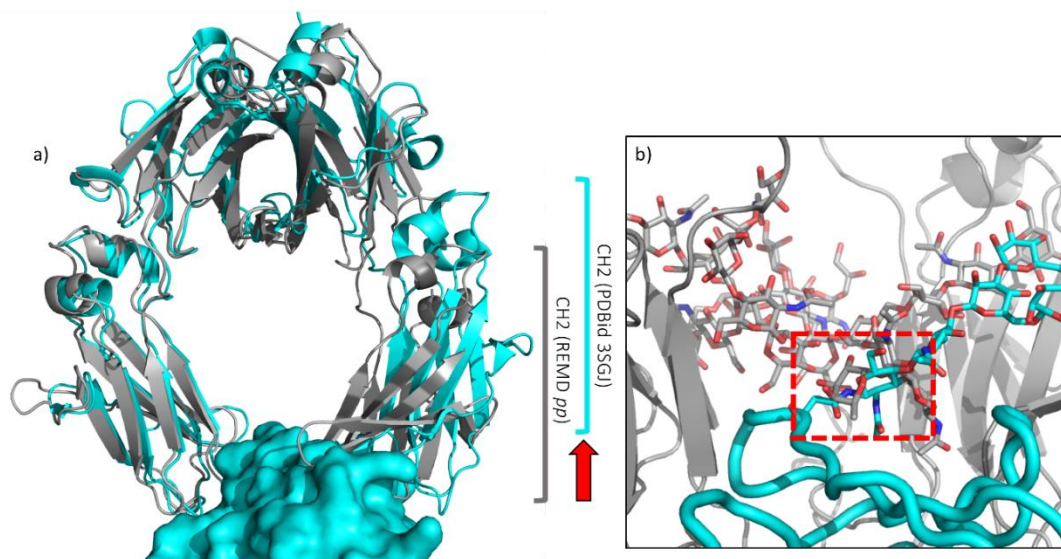
The two N-glycans interact quite extensively with each other through transient and interchanging hydrogen bonds within a network that involves primarily the trimannose residues, while the arms are not involved. A representative structure from the simulation of the *oo* IgG1 Fc is shown in **Figure S.3** in **Appendix III**, where is also evident the narrow space between the  $\alpha(1-3)$  arm of one N-glycan and the CH2-linked GlcNAc of the other, which leaves very little room for the  $\alpha(1-6)$  fucose.

### Core-fucosylation hinders Fc $\gamma$ R access to the binding site.

To evaluate a potential direct effect of the core-fucose in the binding of the Fc $\gamma$ Rs we performed different structural alignments of representative conformations obtained throughout our simulations to crystal structures of IgGs Fcs in complex with Fc $\gamma$ Rs, one with N-glycosylated Fc $\gamma$ RIII at Asn 162 with PDBids 3SGJ and 3SGK<sup>14</sup> and one where the Fc $\gamma$ RIII is non-glycosylated with PDBid 1E4K<sup>31</sup>. The structural alignments were done on the set of CH2 and CH3 domains on the left-hand side of the **Figure 5.6 panel A**, with a corresponding backbone RMSD value of 1.9 Å based on 196 atoms. As shown in **Figure 5.6 panel A**, the alignment to the 3GSJ shows that the accommodation of the Fc $\gamma$ RIII requires a significant displacement of the opposite CH2 domain relative to the unbound IgG1 Fc conformation. This is regardless of the core-fucosylation state of the Fc N-glycan linked to that specific CH2 domain. The obstructing effect of the core-fucose is quite evident when we align the C'E loops, i.e. from residue 293 to 300, to identify the relative position of the Fc N-glycan relative to the Fc $\gamma$ RIII residues and to the N-glycan at Asn 162 on the Fc $\gamma$ RIII when sufficient space has been created to accommodate the



Fc $\gamma$ RIII due to the CH2 displacement. As shown in **Figure 5.6 panel B**, the position of the core-fucose does indeed obstruct the access of the Fc $\gamma$ RIII due to a steric clash with the N-glycan at Asn 162. The higher energetic cost of moving the fucosylated N-glycan in addition to the CH2 domain displacement is consistent with a lower binding affinity of the complex. The alignment of our core-fucosylated *pp* IgG Fc to the structure of a complex with Fc $\gamma$ RIII (PDBid 1E4K)<sup>31</sup> shows that in the absence of the N-glycan at Asn 162 the fucose is not hindering the binding, see **Figure S.4 in Appendix III**. Nevertheless, a lower binding affinity may be expected because of the missing interactions of the IgG Fc with the Fc $\gamma$ RIII Asn 162 N-glycan<sup>6</sup>.



**Figure 5.6** Structural alignments of a representative structure from the REMD simulation of the *pp* IgG1 Fc (grey) with the IgG Fc in complex with the Fc $\gamma$ RIII PDBid 3SGJ (cyan). In panel a) the Fc $\gamma$ RIII is shown as a solvent accessible surface while the IgG Fc are represented with cartoon rendering. The N-glycans are omitted for clarity purposes. The CH2 domain shift is represented through a red arrow, while the brackets span the length of the CH2 domain. In panel b) the Fc $\gamma$ RIII backbone is represented in cyan as tube while the glycan as sticks. The structure from *pp* IgG1 Fc MD aligned through the C'E loop to the PDBid 3SGJ is represented in grey, with the Fc N-glycan also in sticks. The clash between the core fucose and the N162 N-glycan on the Fc $\gamma$ RIII is highlighted within a red-dashed square.

## 5.4 Discussion

In this work we have determined the conformational propensities of two specific N-glycans, shown in **Figure 5.1**, linked to the IgG1 Fc at Asn 297 to identify how their sequence and structure modulates the dynamics of the system and in turn how it affects the Fc $\gamma$ Rs molecular recognition and binding affinity. We designed the study specifically to address the still open question on how core-fucosylation of Fc-linked N-glycans reduces drastically the IgGs ADCC especially in the context of sialylation<sup>21</sup>. The comparison between core-fucosylated and non-fucosylated systems shows that core-fucose affects the position of the sialylated  $\alpha(1-3)$  arm of the opposite N-glycan because of steric hindrance. As a consequence, the  $\alpha(1-3)$  arm becomes increasingly dynamic, switching between an outstretched and a bent conformation where the sialic acid is directed towards the Fc core. The  $\alpha(1-3)$  arm outstretched conformation is found to be the highest populated in non-fucosylated systems. Accommodating an outstretched  $\alpha(1-3)$  arm beside the core fucose translates into a widening of the Fc core, thus it has a significant effect in enhancing the dynamics of the Fc, in particular at the level of the CH2 domains, which are quite flexible due to the Fc architecture. The higher level of conformational disorder of the protein and the seclusion of the sialic acid from the interaction with the incoming Fc $\gamma$ Rs could both be determinants in decreasing the levels of molecular recognition and ultimately in weakening the binding affinity. Furthermore, structural alignments of representative conformers from our simulations with different crystal structures of IgG1 Fcs in complex with Fc $\gamma$ RIII<sub>s</sub><sup>14,31</sup> show that core-fucose obstructs the entry of a Asn 162 N-glycosylated Fc $\gamma$ RIII, posing a further burden in terms of the energy required to displace it in addition to the displacement of the CH2 domain. This information derived from the structural alignments of N-glycosylated IgG1 Fcs is quite helpful in that it provides for the first time a view of how the correct structure and dynamics of the Fc-linked N-glycans work within the framework of the IgG1 Fc/Fc $\gamma$ RIII<sub>s</sub> complex. Indeed, because of their intrinsic dynamics, in most crystal structures the N-glycans are either invisible or fitted to very unusual (and debatable) conformations.

In previous work<sup>29</sup> we have determined the intrinsic dynamics and conformational propensities of all N-glycans significantly populated in IgG1s when unlinked from the IgG, i.e. free in solution. We found that the galactosylation of the  $\alpha(1-6)$  arm results in a dramatic change in the N-glycan preferential conformation, where the  $\alpha(1-6)$  arm is folded over the chitobiose core, instead of being outstretched<sup>29</sup>. Such compact structure is consistent with a more difficult recognition of the  $\alpha(1-6)$  in unbound N-glycans from lectins<sup>34</sup> and from sialyltransferases<sup>10</sup>, and it can also possibly explain the interdependence of the N-glycans arms functionalization process<sup>35</sup>. Notably, the  $\alpha(1-6)$  arm folding is independent of the presence of a terminal  $\alpha(2-6)$  sialic acid, of the sequence on the  $\alpha(1-3)$  arm and of core-fucosylation<sup>29</sup>. In an Fc-linked N-glycan the terminal galactose on the  $\alpha(1-6)$  arm is firmly restrained

to the CH2 domain through a network of hydrogen bonds and hydrophobic interaction and we observed no  $\alpha(1-6)$  unbinding nor folding events at 300 K. This is somewhat in disagreement with NMR data that suggest that the Fc-linked  $\alpha(1-6)$  arm has a structural behaviour in between a restrained and a free (unlinked) N-glycan at 15 °C and at room temperatures<sup>36</sup>. Nevertheless, in agreement with the same NMR study above we observe that the free N-glycan behaviour mentioned in the study, which we assume to be the form unrestrained from the CH2 domain, does increase with temperature, see **Figure S.5 in Appendix III**. Indeed, as the temperature raises the  $\alpha(1-6)$  arm becomes progressively more dynamic and at the extreme of 500 K is completely unrestrained. Therefore, it is possible that the strength of the hydrogen bonding interactions delivered by the specific force field representation we have used is too high at 300 K relative to the experimental conditions. It is also to note that the experimental conditions, in terms of complexity of the solution and relative concentrations are quite different from the simulation conditions. Nevertheless, the strength of the non-bonded interactions gets progressively balanced as the temperature is increased through the replicas, providing the correct trend. Additionally, based on these data and analysis the reasons why the galactosylated  $\alpha(1-6)$  arm is more difficult to functionalize both in free and in Fc-linked N

glycans are actually different. In the former case a compact folded structure preclude access relative to the more accessible and outstretched  $\alpha(1-3)$  arm, while in the case of an Fc-linked

N-glycan the strong interaction of the terminal galactose with the CH2 domain segregates the  $\alpha(1-6)$  arm from the solvent, thus precluding accessibility.

## 5.5 Conclusions

In this work we used temperature REMD to assess the role of core-fucosylation of the Fc-linked N-glycans in the IgG1 structure and interaction with Fc $\gamma$ Rs. The results indicate a significant enhancement of the dynamics of the protein and of the N-glycans intrinsic structural disorder. Additionally, we suggest a mechanistic pathway for the binding of Fc $\gamma$ Rs where core-fucose functions as a “door-stop” to the access of Asn 162 N-glycosylated Fc $\gamma$ Rs. The atomistic level, dynamic information in this work provides for the first time to our knowledge a clear understanding of the effect of core fucose in the IgG1 structure and dynamics and a working framework for the rational design of IgG1 N-glycoforms that can enhance ADCC.

## References

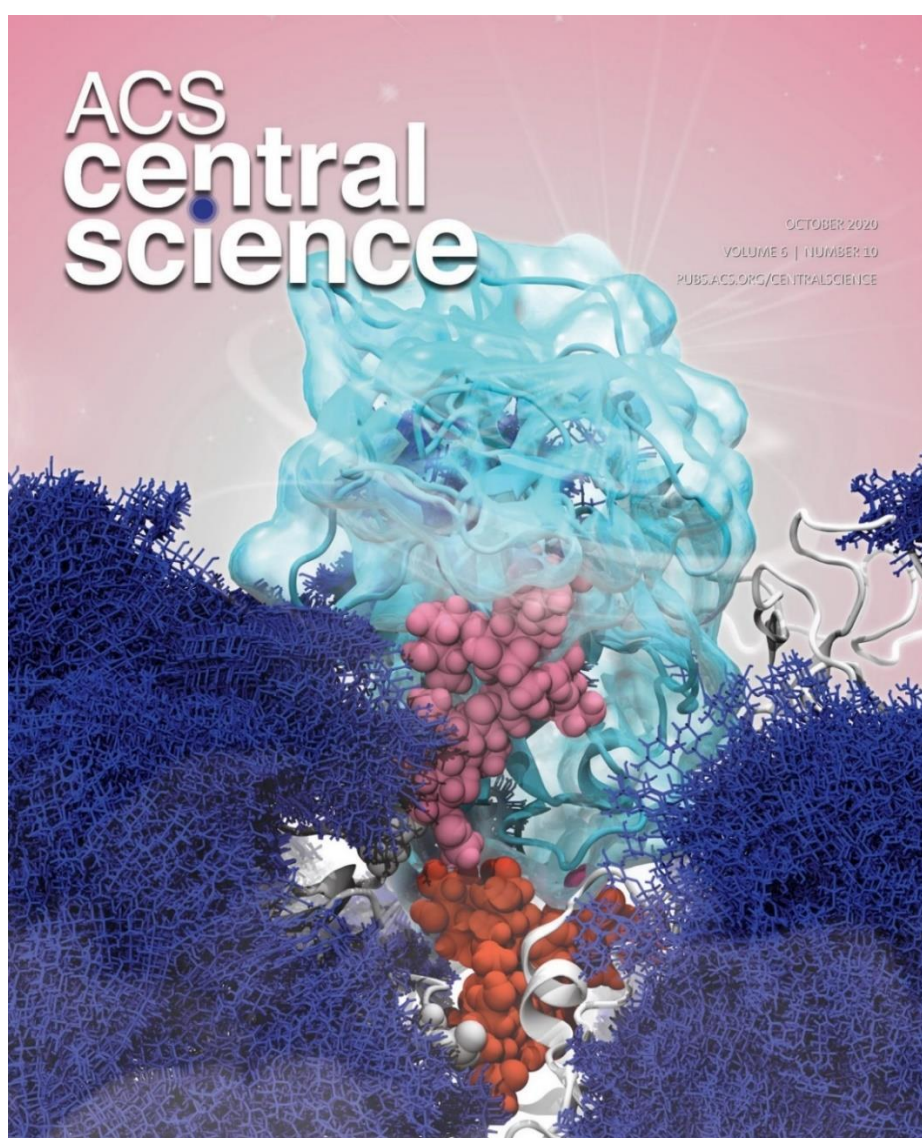
1. G. Vidarsson, G. Dekkers, T. Rispen, IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* **5**, 520 (2014).
2. F. Nimmerjahn, J. V. Ravetch, Divergent immunoglobulin g subclass activity through selective Fc receptor binding. *Science* **310**, 1510-1512 (2005).
3. S. Battella, M. C. Cox, A. Santoni, G. Palmieri, Natural killer (NK) cells and anti-tumor therapeutic mAb: unexplored interactions. *J Leukoc Biol* **99**, 87-96 (2016).
4. Y. Mimura *et al.*, Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* **9**, 47-62 (2018).
5. J. T. Roberts, A. W. Barb, A single amino acid distorts the Fc  $\gamma$  receptor IIIb/CD16b structure upon binding immunoglobulin G1 and reduces affinity relative to CD16a. *J Biol Chem* **293**, 19899-19908 (2018).
6. G. P. Subedi, A. W. Barb, CD16a with oligomannose-type. *J Biol Chem* **293**, 16842-16850 (2018).
7. J. M. Hayes *et al.*, Glycosylation and Fc receptors. *Curr Top Microbiol Immunol* **382**, 165-199 (2014).
8. J. M. Hayes *et al.*, Identification of Fc Gamma Receptor Glycoforms That Produce Differential Binding Kinetics for Rituximab. *Mol Cell Proteomics* **16**, 1770-1788 (2017).
9. M. Pucić *et al.*, High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* **10**, M111.010090 (2011).
10. A. W. Barb, E. K. Brady, J. H. Prestegard, Branch-specific sialylation of IgG-Fc glycans by ST6Gal-I. *Biochemistry* **48**, 9705-9707 (2009).
11. J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd, R. A. Dwek, The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol* **25**, 21-50 (2007).
12. I. Gudelj, G. Lauc, M. Pezer, Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol* **333**, 65-79 (2018).
13. J. D. Pagan, M. Kitaoka, R. M. Anthony, Engineered Sialylation of Pathogenic Antibodies In Vivo Attenuates Autoimmune Disease. *Cell* **172**, 564-577.e513 (2018).
14. C. Ferrara *et al.*, Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc $\gamma$ RIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **108**, 12669-12674 (2011).
15. Y. Sakae *et al.*, Conformational effects of N-glycan core fucosylation of immunoglobulin G Fc region on its interaction with Fc gamma receptor IIIa. *Scientific Reports* **7**, (2017).

16. C. Ferrara, F. Stuart, P. Sondermann, P. Brünker, P. Umaña, The carbohydrate at Fcγ3b1 Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. *J Biol Chem* **281**, 5032-5036 (2006).
17. C. H. Lai *et al.*, Analysis of Carbohydrate-Carbohydrate Interactions Using Sugar-Functionalized Silicon Nanoparticles for Cell Imaging. *Nano Lett* **16**, 807-811 (2016).
18. D. Spillmann, M. M. Burger, Carbohydrate-carbohydrate interactions in adhesion. *J Cell Biochem* **61**, 562-568 (1996).
19. R. L. Shields *et al.*, Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcγ3b1 and antibody-dependent cellular toxicity. *J Biol Chem* **277**, 26733-26740 (2002).
20. S. Iida *et al.*, Nonfucosylated therapeutic IgG1 antibody can evade the inhibitory effect of serum immunoglobulin G on antibody-dependent cellular cytotoxicity through its high binding to Fcγ3b1. *Clin Cancer Res* **12**, 2879-2887 (2006).
21. T. Li *et al.*, Modulating IgG effector function by Fc glycan engineering. *Proc Natl Acad Sci U S A* **114**, 3485-3490 (2017).
22. M. Kiyoshi *et al.*, Structural basis for binding of human IgG1 to its high-affinity human receptor FcγRI. *Nat Commun* **6**, 6866 (2015).
23. C. J. Day *et al.*, Glycan:glycan interactions: High affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells. *Proc Natl Acad Sci U S A* **112**, E7266-7275 (2015).
24. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141-151 (1999).
25. J. A. Maier *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713 (2015).
26. K. N. Kirschner *et al.*, GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622-655 (2008).
27. W. JORGENSEN, J. CHANDRASEKHAR, J. MADURA, R. IMPEY, M. KLEIN, COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. *Journal of Chemical Physics* **79**, 926-935 (1983).
28. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
29. A. M. Harbison, L. P. Brosnan, K. Fenlon, E. Fadda, Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94-103 (2019).
30. P. Strop *et al.*, Generating bispecific human IgG1 and IgG2 antibodies from any antibody pair. *J Mol Biol* **420**, 204-219 (2012).

31. P. Sondermann, R. Huber, V. Oosthuizen, U. Jacob, The 3.2-Å crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature* **406**, 267-273 (2000).
32. D. J. Falconer, G. P. Subedi, A. M. Marcella, A. W. Barb, Antibody Fucosylation Lowers the FcγRIIIa/CD16a Affinity by Limiting the Conformations Sampled by the N162-Glycan. *ACS Chem Biol* **13**, 2179-2189 (2018).
33. M. Frank, R. C. Walker, W. N. Lanzilotta, J. H. Prestegard, A. W. Barb, Immunoglobulin G1 Fc domain motions: implications for Fc engineering. *J Mol Biol* **426**, 1799-1811 (2014).
34. B. Echeverria *et al.*, Chemoenzymatic Synthesis of N-glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-type Lectin Receptors. *ACS Chem Biol* **13**, 2269-2279 (2018).
35. U. Moginger *et al.*, Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Frontiers in Oncology* **8**, (2018).
36. A. W. Barb, J. H. Prestegard, NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic. *Nat Chem Biol* **7**, 147-153 (2011).

## Chapter 6: All-Atom Molecular Dynamics Simulations of SARS-CoV-2 Spike Protein with glycan variation

**Note:** Part of the work from this chapter was published in the paper: L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, R. E. Amaro, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734 (2020), which was also highlighted on the journal’s back cover.

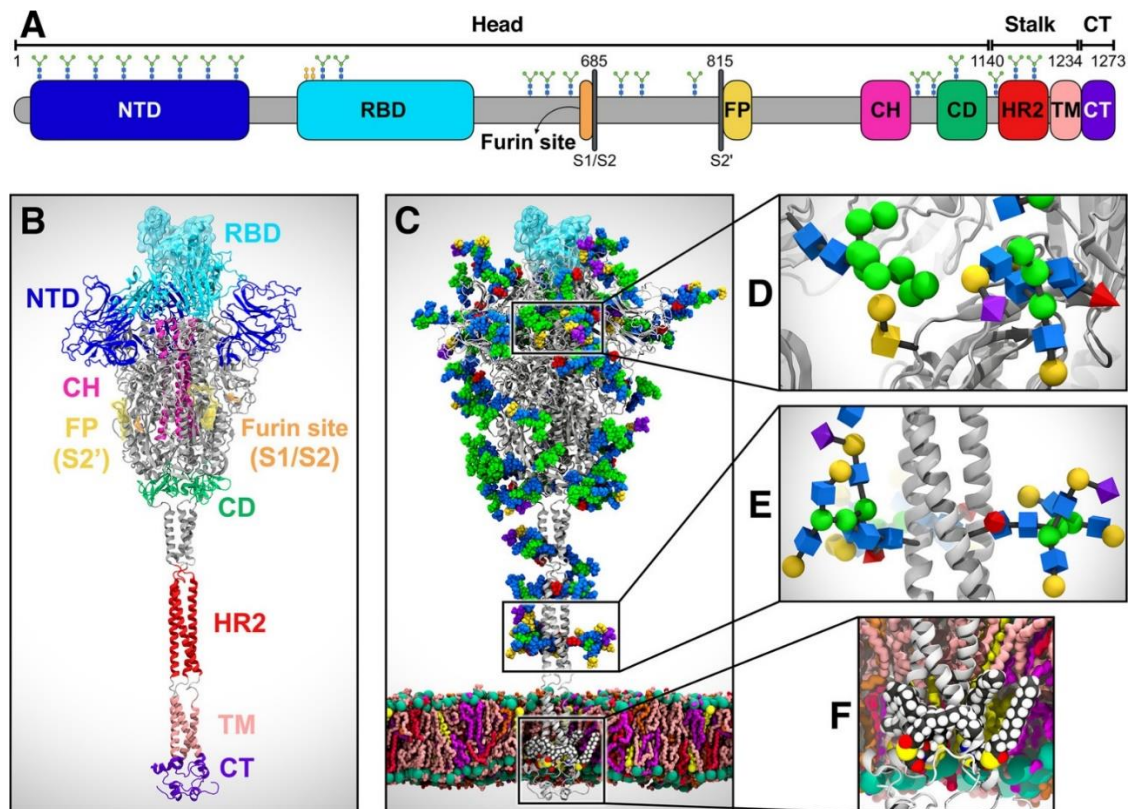


## 6.1 Introduction

Coronaviruses are a family of related RNA viruses that infect mammals and bird species, with four genera: *α-coronavirus*, *β-coronavirus*, *χ-coronavirus* and *δ-coronavirus*. There are varieties of coronaviruses that are attributed to mild “common cold” symptoms in humans<sup>1</sup>, while others pose a more severe threat. In the last two decades, *β*-coronaviruses have been a cause of concern for human health, as there have been epidemics caused by Severe Acute Respiratory Syndrome (SARS-CoV) in 2003<sup>2</sup>, Middle Eastern Respiratory Syndrome (MERS) in 2012<sup>3</sup>, and in 2020, an outbreak of SARS-CoV-2, also referred to as COVID-19 or 2019-nCoV, has caused a global pandemic<sup>4</sup>.

A characteristic feature of the coronavirus architecture is the “spike” (S) protein, anchored to the surface membrane of the virus, used primarily for recognition, binding to a primary receptor which leads to fusion to the host cell<sup>5</sup>. The S protein consists of a single chain precursor that forms a homotrimeric structure once folded<sup>6</sup>. Each S monomer can be organised into two subunits, namely S1 and S2 and in three topological sections, namely head, stalk and tail, see **Figure 6.1.A**.





**Figure 6.1. System overview.** (A) Sequence of the full-length spike (S) protein contains the N-terminal domain (NTD, 16–291), receptor binding domain (RBD, 330–530), furin cleavage site (S1/S2), fusion peptide (FP, 788–806), central helix (CH, 987–1034), connecting domain (CD, 1080–1135), heptad repeat 2 (HR2, 1163–1210) domain, transmembrane domain (TM, 1214–1234), and cytoplasmic tail (CT, 1235–1273). Representative icons for N-glycans (blue and green) and O-glycan (yellow) are also depicted according to their position in the sequence. (B) Assembly of the head, stalk, and CT domains into the full-length model of the S protein (Open system). (C) Fully glycosylated and palmitoylated model of the S protein in the (Open system). (D) Magnified view of S protein head glycosylation, where glycans are depicted using the Symbol Nomenclature for Glycans (SNFG). (E) Magnified view of S protein stalk glycosylation. (F) Magnified view of S protein S-palmitoylation within CT. Image from Casalino *et al*<sup>7</sup>.

Each topological domain can be further subdivided based on function, see **Figure 6.1.B**, with the majority of pre-fusion mechanistic tasks being performed by the protein head subdomain. The most important domains in terms of mechanistic roles in the S protein head are the receptor binding domain (RBD), shown in cyan in **Figure 6.1.B**, responsible for making initial contact with the host cell-bound receptor and the fusion peptide (FP), shown yellow in **Figure 6.1.B**, which penetrates and primes the host cell membrane for fusion<sup>8</sup>. The S protein is also heavily glycosylated, see **Figure 6.1.C**, with 22 N-glycosylation sequons per chain<sup>9,10</sup>. Carrying a thick glycan shield is part of a common strategy used by viruses to evade the host immune system by means of “shielding” and “mimicry”<sup>11</sup>. Glycan shielding by viral glycoproteins uses the coverage of densely glycosylated proteins to mask immunogenic

epitopes, while also hijacking host cell machinery to glycosylate their proteins, a strategy called glycan mimicry, that creates a “wolf in sheep’s clothing” effect<sup>11</sup>. Glycomics data analysis recently revealed that coronaviruses envelope proteins are less shielded than other viral glycoproteins, such as HIV-1 Env, which results in a lower level of underprocessed oligomannose N-glycans and a higher level of complex N-glycans<sup>12</sup>. Moreover, our work has highlighted that uniquely to the SARS-CoV-2, specific N-glycans on the S protein mediate its mechanism of action, and therefore its method of infection<sup>7</sup>. As it will be explained further below in this section, the specific type of glycosylation at N234, N165 and N343 is crucial towards the RBD’s relative orientation and thus to the S protein activity.

Conservation of the architecture of the S protein varies between members of the *Coronaviridae* family and therefore a wide variety of host cell receptors are targeted. Although belonging to the same  $\beta$ -genus as SARS-CoV-2, MERS has a different RBM structure, which recognises dipeptidyl peptidase 4 (DPP4)<sup>13</sup>. SARS-CoV and SARS-CoV-2 however have a 76% sequence identity<sup>10</sup>, both targeting angiotensin-converting enzyme 2 (ACE2)<sup>14,15</sup>. Between SARS-CoV and SARS-CoV-2, 18 of 22 N-glycosylation sites *per* chain are conserved, which suggests that the degree of shielding of the immunogenic epitopes is similar<sup>12</sup>. In terms of sequence mutations, which includes the addition/deletion of glycosylation sites, most of the changes are observed in and around the S1 subunit and especially in the RBD. To this end, the binding affinity of ACE2 by the RBD of SARS-CoV-2 has been reported to be 10- to 20- fold higher than that of the RBD of SARS-CoV<sup>16</sup>, which correlates to a higher rate of infectivity and transmissibility with SARS-CoV-2<sup>17</sup>. In order to understand this higher binding affinity of the RBD of SARS-CoV-2, the role that both the surrounding protein and glycans play in the mechanics and interactions of the RBD needs to be considered.

The S protein head undergoes a complex conformational change, from a “closed” (or “down”) state to an “open” (or “up”) state of the RBDs<sup>10</sup>, see **Figure S.1 in Appendix IV**. Only when the RBD is in an open conformation the receptor binding motif (RBM) can bind the ACE2 receptor in a one-to-one interaction achieved through protein-protein and protein-carbohydrate interactions<sup>18</sup>. The ACE2-RBD binding triggers the concerted actions of S2 rearrangement and S2’ site cleavage by host cell proteases<sup>19</sup>. Uniquely for its lineage, SARS-CoV-2 has a RRAR cleavage site between subunits S1 and S2 specific to furin<sup>20</sup>, which further promotes the dissociation of S1 and S2 domains, which precedes membrane fusion<sup>21</sup>. After membrane fusion, the virion is internalised through endocytosis and the viral mRNA genome is released into the cytoplasm. The host cell ribosome reads and translates the RNA into viral proteins, including the S protein, which is transported through the endoplasmic reticulum (ER) and the Golgi in order to be glycosylated, further assembled and incorporated into the budding virions<sup>22</sup>.

Obtaining atomistic structural data on the full “prefusion” S glycoprotein was achieved through mutations at residues 986 and 987 which stabilize S in its prefusion trimeric form<sup>23</sup>, named SARS-CoV-

2 S-2P variant. This allowed for easier expression and for characterization of the native-like structure by cryogenic electron microscopy (cryo-EM)<sup>16</sup>. Structural studies have revealed both the open and closed conformations<sup>10,16</sup> and glycosylation sites of the recombinant S protein bearing one or even two GlcNAc residues, which can help indicate the orientation of the whole N-glycans relative to the protein surface. This paired with glycomics data obtained on the HEK cells recombinant S-2P protein<sup>9</sup> has given us insight into the potential glycosylation pattern of a fully assembled prefusion trimer. Results from Wantanabe *et al.*<sup>9</sup> are summarised in **Table 6.1** in terms of relative oligomannose abundance. The same work indicates that an overall 22% of glycans are of the oligomannose-type, 5% of the hybrid-type and 66% of the complex-type, with 1% of the sites unoccupied. Since the beginning of 2020, different glycosylation profiles have been published<sup>24–26</sup> showing broader ranges of glycosylation types for each site. These differences have been attributed to whether the recombinant structure was analysed as an intact ectodomain or as separate S1 and S2 subunits, or based on a different interpretation of glycopeptide fragmentation software data following mass spectrometry (MS).

**Table 6.1.** Percentage abundance of oligomannose-type N-glycans at each identified N-glycosylation site of the S protein. Data from Wantanabe *et al.*<sup>9</sup>.

N-glycosylation site	% oligomannose content	N-glycosylation site	% oligomannose content
N17	4	N616	6
N61	69	N657	0
N74	4	N709	94
N122	56	N717	74
N149	8	N801	77
N165	0	N1074	57
N234	97	N1098	9
N282	1	N1134	0
N331	2	N1158	0
N343	2	N1173	1
N603	65	N1194	0

In collaboration with Prof Rommie Amaro's group at UCSD, we set out to explore the glycans coverage of the S protein surface and the specific interactions between glycan and protein with both the open and closed conformations of the RBD, using all-atoms molecular dynamics (MD) simulations on both, a model of the whole S protein embedded in the viral membrane and on its ectodomain alone. In this work we explored the effect of different glycosylation patterns on multiple uncorrelated copies of the simulations run concurrently to explore a wider section of the conformational space. We also considered the effects of the use of different force fields, namely GLYCAM06j and CHARMM36 and of starting cryo-EM structures, namely PDBid 6VYB and 6VSB for the open conformation of the RBD. Our simulations reveal a greater role of specific S1 N-glycans in stabilization of the RBD, as well as

modulating the transition between open and closed states. Following these results, we focused our efforts in assessing the importance of the glycosylation type at N234 and in neighbouring sites on the dynamics of the open conformation of the RBD.

## 6.2 Computational Method

The next subsections detail the setup and running specifications for our MD simulations of the SARS-CoV-2 S protein model, with our collaborative work with Prof. Amaro's group described, followed by the details of our continued simulation of the ectodomain of the S protein, with specific glycan variation at N234.

### **Initial Simulation Set**

The full-length S glycoprotein model, shown in **Figure 6.1.B**, was built in Prof Rommie Amaro's lab under the guidance obtained from data we collected on preliminary simulations of the S ectodomain. The full model was built by homology in both, "open" (PDB 6VSB<sup>16</sup>) and "closed" (PDB 6VXX<sup>10</sup>) conformational states, see **Figure S1** in **Appendix IV** and **Section 1.1** of **Appendix IV**. The protein models were glycosylated at N-/O-glycosylation sites<sup>9,24</sup> and the cysteines in the CT, see **panel F** in **Figure.6.1**, were palmitoylated<sup>27,28</sup>. Finally, the full length structures were embedded into a pre-equilibrated all-atom membrane, in order to mimic the endoplasmic reticulum-Golgi intermediate compartment, see **Table S4** in **Appendix IV**, where the spike is glycosylated and the virus buds. Following this, the system was solvated in an atmosphere of counterions. Six replicas of the open conformation models and three of the closed conformation models were run on the NSF Frontera computing system at the Texas Advanced Computing Center (TACC) with NAMD 2.14,<sup>34</sup> achieving benchmarks of ~60 ns/day on 256 nodes for cumulative extensive sampling of ~4.2 and ~1.7  $\mu$ s, respectively, see **Figure S5**, **Movie S1** in **Appendix IV**. A greater number of replicas of the open conformation systems were run to understand the "meta-stable" state of the open RBD that was captured by the cryo-EM experiments

In our group, we rebuilt the SARS-CoV-2 S protein ectodomain in the open conformation by homology modelling with SWISS MODEL<sup>29</sup> using the cryo-EM structure with PDB ID 6VYB<sup>10</sup> (3.2 Å resolution) and NCBI YP\_009724390.1 as reference sequence. The missing loops in the 6VYB cryo-EM structure were built automatically by SWISS MODEL based on structural libraries of backbone fragments from the PDB with similar sequences. The resulting protein structure exhibits 18 N-glycosylation sites (or sequons) per protomer, for a total of 54 sites per trimer. Glycosylation on these sites was built by

structural alignment of conformationally equilibrated N-glycan structures from our GlycoShape library (under construction), more specifically fucosylated and non-fucosylated biantennary (FA2B/A2B) N-glycans and oligomannose type (Man5 and Man9)<sup>30</sup> to the GlcNAc residues resolved in the cryo-EM structure. Selection of the type of glycan at each specific N-glycosylation site was decided based on the glycosylation profile in Wantanabe *et al.*<sup>9</sup>, shown in **Table 6.1**, which was the only data set available at the time we started the work (March 2020). The variation of glycosylation predicted at in some of the sites was broad. Based on our knowledge of glycan sequence-to-structure relationship and dynamics, we decided to include, whenever likely, more processed glycans, e.g. with galactosylation on the both arms and core-fucosylation, to investigate possible interactions with the protein.

Some of the glycosylation sites reside in disordered loops that needed to be reconstructed, therefore the final 54-glycans model was built in two phases. In the first phase we built models with 46 glycosylation sites and run a 20 ns equilibration to obtain conformations of the rebuilt loops that allowed for the linking of our pre-equilibrated glycan structures. Appropriate loop conformations were then chosen to link the glycans on those sites to complete the glycosylation profile. In this additional set of simulations of the SARS-CoV-2 S protein ectodomains, we considered three slightly different glycosylation profiles, shown in **Table 6.2**, to complement results obtained for the full S model. These three models differed specifically at position N234, occupied either by a Man9 (Man9-N234) or where N234 is mutated into Ala (N234A) leading to glycan deletion, or where the sequon at N234 is not mutated but not glycosylated. We ran two independent trajectories for the 46-glycans model (i.e., Man9-N234) and one for the related N234A mutant. Moreover, we performed one run each for the 54-glycans models, i.e., Man9-N234, the N234A mutant, and the 53-glycan model (non-glycosylated at N234) for a total of 6 independent MD runs, see **Table 6.3**. Notably, also the glycosylation we chose at N165 and N343 differs from the one we selected for the full model of the spike to explore different profile. Also to note, the N165 and N234 sites in the ectodomain models are located in the NTD of chain C (NTD-C), while the N343 is located in the RBD of chain C (RBD-C).

**Table 6.2.** Site specific glycosylation in the two 54-glycans models of SARS-CoV2 S head (~60,000 atoms), i.e. Man9-N234 and N234A. The asterisk on position 234 chain C indicates the critical region where the glycans support the open RBD (chain B in 6VYB) by filling the empty space. We note that the third model is based on these 54-glycans models, but it was not glycosylated at position N234 within chain C.

Chain	Resid	Glycan	Chain	Resid	Glycan	Chain	Resid	Glycan
A	61	Man5	B	61	Man5	C	61	Man5
A	74	FA2G	B	74	FA2G	C	74	FA2G
A	122	Man5	B	122	Man5	C	122	Man5
A	149	FA2G	B	149	FA2G	C	149	FA2G
A	165	Man5	B	165	Man5	C	165	Man5
A	234	Man9	B	234	Man9	C	234*	Man9/ N234A
A	282	FA2G	B	282	FA2G	C	282	FA2G
A	331	FA2G	B	331	FA2G	C	331	Man5
A	343	Man5	B	343	Man5	C	343	Man5
A	603	Man5	B	603	Man5	C	603	Man5
A	616	A2G	B	616	A2G	C	616	A2G
A	657	Man5	B	657	Man5	C	657	FA2G
A	709	Man5	B	709	Man5	C	709	Man5
A	717	Man5	B	717	Man5	C	717	Man5
A	801	Man5	B	801	Man5	C	801	Man5
A	1074	Man5	B	1074	Man5	C	1074	Man5
A	1098	A2G	B	1098	A2G	C	1098	A2G
A	1134	FA2G	B	1134	FA2G	C	1134	FA2G

In the simulations of the S ectodomain models based on PDB 6VYB, the protein and counterions (200 mM) were represented by the AMBER ff14SB parameter set<sup>31</sup>, whereas the glycans by the GLYCAM06j-1 version of the GLYCAM06 force field<sup>32</sup>. Water molecules were represented by the TIP3P model<sup>33</sup>. All simulations were run with v18 of the AMBER software package<sup>34</sup>. The same system preparation and running protocol was used for all MD simulations.

The energy of the S ectodomains models was minimized in two steps of 50,000 cycles of the steepest descent algorithm each. During the first minimization all the heavy atoms were kept harmonically restrained using a potential weight of 5 kcal mol<sup>-1</sup>Å<sup>2</sup>, while the solvent, counterions and hydrogen atoms were left unrestrained. During the second minimization step, only the protein heavy atoms were kept restrained, while the glycans, solvent, counterions and hydrogens were left unrestrained. After energy minimization the system was equilibrated in the NVT ensemble with the same restraints scheme, where heating was performed in two stages over a total time of 1 ns, from 0 to 100 K (stage 1) and then

100 to 300 K (stage 2). During equilibration the SHAKE algorithm was used to constrain all bonds to hydrogen atoms. The Van der Waals were truncated at 11 Å and Particle Mesh Ewald (PME) was used to treat long range electrostatics with B-spline interpolation of order 4. Langevin dynamics with collision frequency of 1.0 ps<sup>-1</sup> was used to control temperature, which a pseudo-random variable seed to ensure there are no synchronization artefacts. Once the system was brought to 300 K an equilibration phase in the NPT ensemble of 1 ns was used to set the pressure to 1 atm. The pressure was held constant with isotropic pressure scaling and a pressure relaxation time of 2.0 ps. At this point all restraints on the protein heavy atoms were removed, allowing the system to evolve for 15 ns of conformational equilibration before production. The total simulation times, including equilibration, are shown in **Table 6.3**.

**Table 6.3.** MD production times (ns) used for the data analysis of the SARS-CoV2 S glycoprotein ectodomain (60,000 atoms). MD1 and MD2 refer to two independent MD trajectories started from different velocities.

Systems	46 glycan MD1	46 glycan MD2	54 glycan
Man9-N234	210	-	120
N234A	240	240	390
N234-nogly	-	-	420

### Construction of the second set of S protein ectodomain models

The first set of simulations<sup>7</sup> highlighted the crucial role of the N-glycan at N234 in filling the void left vacant by the opening of the RBD, therefore as a second stage of the work we wanted to explore if any type of glycosylation would have been capable to fulfil this role. An additional three models of the S ectodomain with 54-glycans. Since the completion of our first phase of simulation, more glycoanalytic studies were performed<sup>24-26</sup>, and we could make a more informed decision on our choice of glycosylation at each N-glycan site, as well as inclusion of O-glycans, based on Shajahan *et al.*<sup>24</sup>, with an O-glycan included at T323. In these models we changed the level of processing of the oligomannose glycan at N234 to understand what specific features of the oligomannose N-glycan is required (if any) to stabilize the RBD in the open conformation. Inclusion of paucimannose (Man3) as one of the variations has not been observed at N234 in any glycan profiling, but we chose to include it as it allows us to understand the basic function of an N-glycan at this glycosylation site and to rationalize the additive features of the antennae of the more elaborate oligomannose forms. The same glycosylation profile was used for each protomer, in order to directly compare behaviour of the N-glycans, with regards to the closed and open conformations of the RBD on each model. The change in glycosylation from set 1 to set 2, including details of each model is shown in **Table 6.4**.

**Table 6.4.** Details of changes of glycosylation type at each available glycosylation site of the SARS-CoV-2 S protein head, applicable to all three chains of the trimer. The asterisk denotes the difference in glycosylation at site N234 for the previous models. The empty spaces indicate no change in glycosylation from the previous models. The dagger indicates the glycan type for N165 and N343 for Chains A and C in the previous models.

PDB	Previous Models	Model 1	Model 2	Model 3
<b>Residue</b>	<b>Glycan Type</b>			
61	Man5			
74	FA2G2			
122	Man5			
149	FA2G2			
165	Man5†	FA2G2	FA2G2	FA2G2
234	Man9*	Man5	Man3	Man9
282	FA2G2			
T323	unoccupied	O-glycan	O-glycan	O-glycan
331	FA2G2			
343	Man5†	FA2G2	FA2G2	FA2G2
603	Man5			
616	A2G2			
657	Man5	FA2G2	FA2G2	FA2G2
709	Man5			
717	Man5			
801	Man5			
1074	Man5			
1098	A2G2			
1134	FA2G2			

This second set of spike ectodomain models were set-up and studied by MD simulations following a same protocol described earlier. The total simulation time for each replica produced to this day is shown in **Table 6.5**, including equilibration times.

**Table 6.5.** MD simulations length ( $\mu$ s) of the second set of SARS-CoV2 S protein ectodomain models.

Systems	Simulation Time ( $\mu$ s)
Model 1 Replica 1 (Man5 R1)	2.01
Model 1 Replica 2 (Man5 R2)	2.01
Model 2 Replica 1 (Man3 R1)	1.95
Model 2 Replica 2 (Man3 R2)	1.89
Model 3 Replica 1 (Man9 R1)	2.01

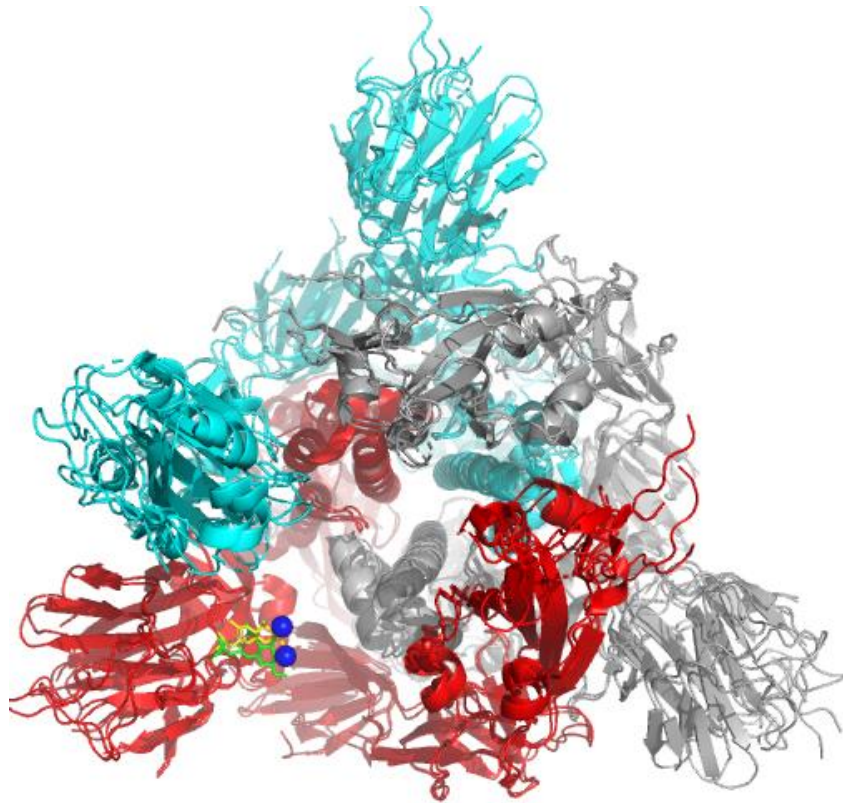


## 6.3 Results

### **Initial MD Simulations of the SARS-CoV-2 S Protein Models in Open and Closed Conformations**

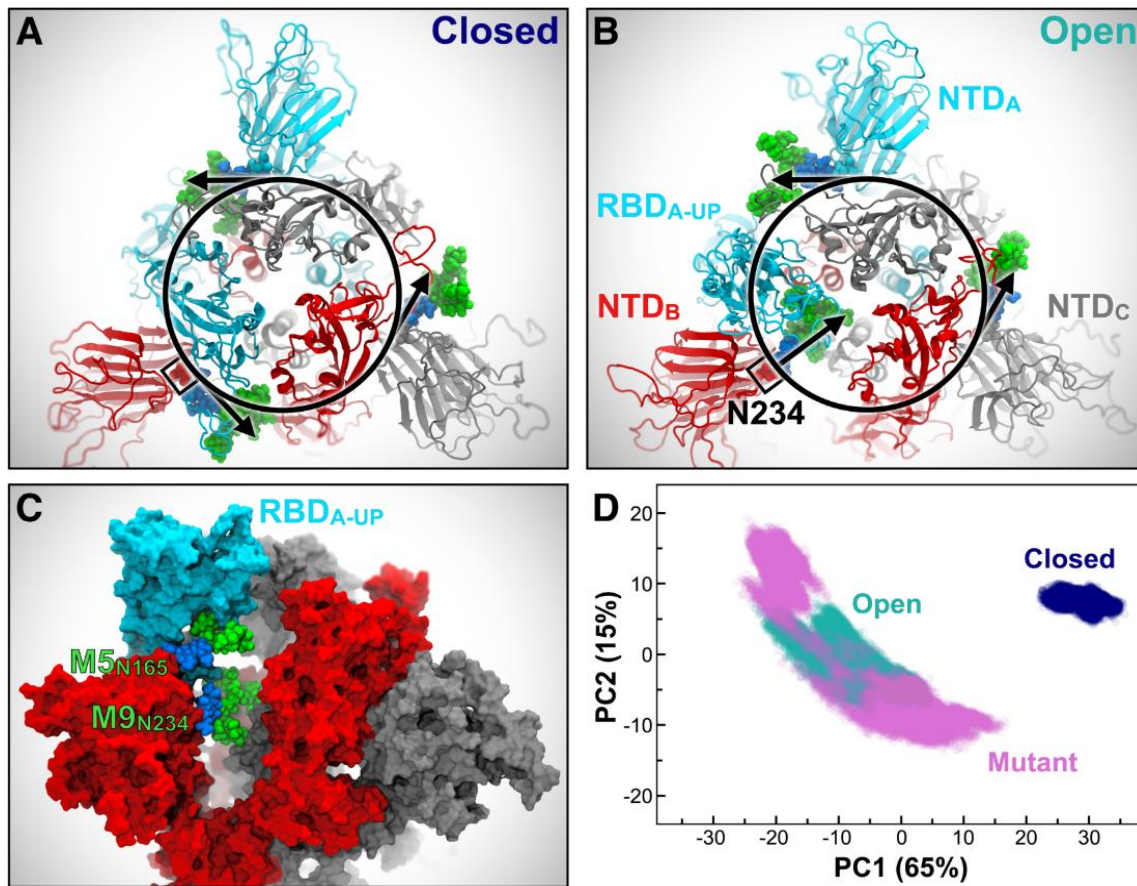
The overall structural stability of the full S protein models was monitored using root-mean-square-deviation (RMSD) values relative to the starting structures, with convergence of the head and stalk domains reached within 400 ns, see **Figure S2 and S3** in **Appendix IV**. The CT domain however fluctuates in RMSD values between replicas, see **Figure S3** in **Appendix IV**, with large deviations due to the section that is exposed to the solvent and not anchored within the lipid bilayer. Root-mean-square-fluctuation (RMSF) analysis performed on the glycans, see **Figure S4** in **Appendix IV**, revealed the more branched, complex N-glycans had the most mobility on average, with the O-glycans at T323 and S325 showing the least flexibility, which may be accounted for due to their short size. The tetra-antennary complex N-glycans located on the stalk domain had the highest fluctuations, suggesting its ability to shield the protein surface.

Based on both variations of the SARS-CoV-2 ectodomain, conducted by the Fadda and Amaro groups, the N-glycans at N165 and N234 were identified as potentially crucial components in mediation of the RBD dynamics. The closed conformation replicate models reflect the results of glycoanalytic data<sup>9</sup> that suggest that the N-glycan at N234 is solvent exposed, orientated away from the core of the S protein (**Figure 6.3**). In both variations of the open conformation models, Man9 at N234 is buried within the cleft left by the RBD in the open structure (**Figure 6.3.B**). Note, in the two different cryo-EM starting structures 6VSB and 6VYB, the open RBDs are named differently, i.e. RBD-A in the full S protein open structure, and RBD-B in the ectodomain open structure, respectively. Interestingly, because of the different orientation of the N-glycan fragments resolved by cryo-EM at N234, the positions of the whole N-glycans during the MD simulations are different (see **Figure 6.2**), as a result of the rigidity of the chitobiose N-glycan core, although they converge to the same final orientation in the core of the trimer at the end of the sampling trajectories.



**Figure 6.2. Alignment of 6VSB and 6VYB open chains.** The glycan fragment at N234 from 6VYB (yellow) and from 6VSB (green) are close in Cartesian space, but the hydroxyl group (O4) of the GlcNAc monosaccharide (blue spheres) are pointed in different directions, which will orientate the superimposed Man9 residues in different starting positions.

More specifically, the simulations of the full S protein models started with the Man9 at N234 inserted into the cleft, while the simulation of the S ectodomains had the Man9 exposed to the solvent in a similar position as the one found in the closed conformation. Interestingly, in the simulations of the ectodomain while the system equilibrated, we observed the Man9 progressively moving inwards to fill the void space left vacant by the RBD opening. This progressive “crawling” takes place through the Man9 forming interactions with the opposing RBD-C residues, and then subsequently forming interactions with protein residues within the core of the spike structure (**Movie S.2** in **Appendix IV**). The N-glycan at N165 in the simulations of the open conformation is found interacting with the RBD-B either above itself or in a position below, between the RBD-B and the NTD-C it is linked to (**Figure 6.3.C**).

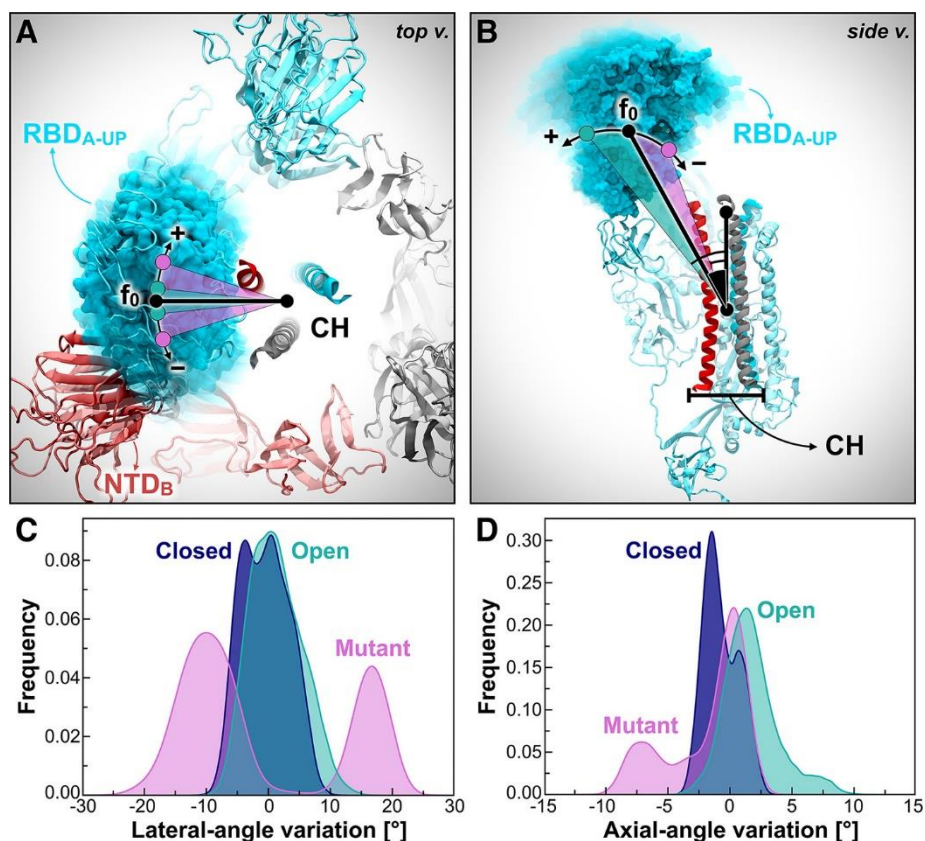


**Figure 6.3. N234A and N165A mutations show increased instability of RBD-A in the “up” state.** (A-B) Top view of the S protein as in the closed (A) and open (B) systems. Protein is represented with cartoons, coloured in cyan, red, and grey for chains A, B and C, respectively. O-mannose N-glycans at position N234 are depicted with VdW spheres, where GlcNAc is colored in blue and Man in green. In closed (A), all the N-glycans at N234 are tangential to a hypothetical circle going through N234. In open (B), the N-glycan at N234 of chain B moves inward, filling in the vacancy under RBD-A in the “up” conformation. (C) Side view of the S protein (surface representation) in open, where the RBD of chain A (RBD-A, cyan) is stabilized by N-glycans at N165 and N234 in the “up” conformation. Same color scheme as panels A and B is applied. (D) PCA plot showing PC1 vs. PC2 of RBD-A (residues 330–530) in closed, open, and mutant in blue, teal, and magenta, respectively. The amount (%) of variance accounted by each PC is shown between parenthesis.

In order to understand the significance of these two glycosylation sites with regards to the RBD dynamics, a set of “mutant” models was produced, by mutating N234A and N165A from the open conformation starting structure. Principal Component Analysis (PCA) was used to compare the conformational landscape of the open, closed and mutant models, with scatter plot projections generated from the first two eigenvectors, the motions with the two largest variances of 63% and 18% (see **Panel D in Figure 6.D**). This plot shows the mutant systems explore a greater conformational space than that of the open or closed conformational systems. This suggests a loss in stability from the absence of the N-glycans at N234 and N165, which suggests that these glycans play an essential role in the structural integrity of the open and closed conformations. The importance of the N-glycan at N234 can also be

observed in a much subtler single-point mutation N234A in the additional set of simulations of the S ectodomain (see **section 2.2** of **Appendix IV**).

The two most prevalent motions identified by PCA were further monitored using angle calculations (details of method in **section 1.1** in **Appendix IV**). The first motion identified (PC1) was the in-plane motion of the RBD along the arc of a hypothetical circle centred on the central helices of the S protein, labelled the “lateral angle” (**Figure 6.4.A**). The second (PC2) was the RBD tilting motion away from or towards the central axis of the central helices, called the “axial angle” (**Figure 6.4.B**).



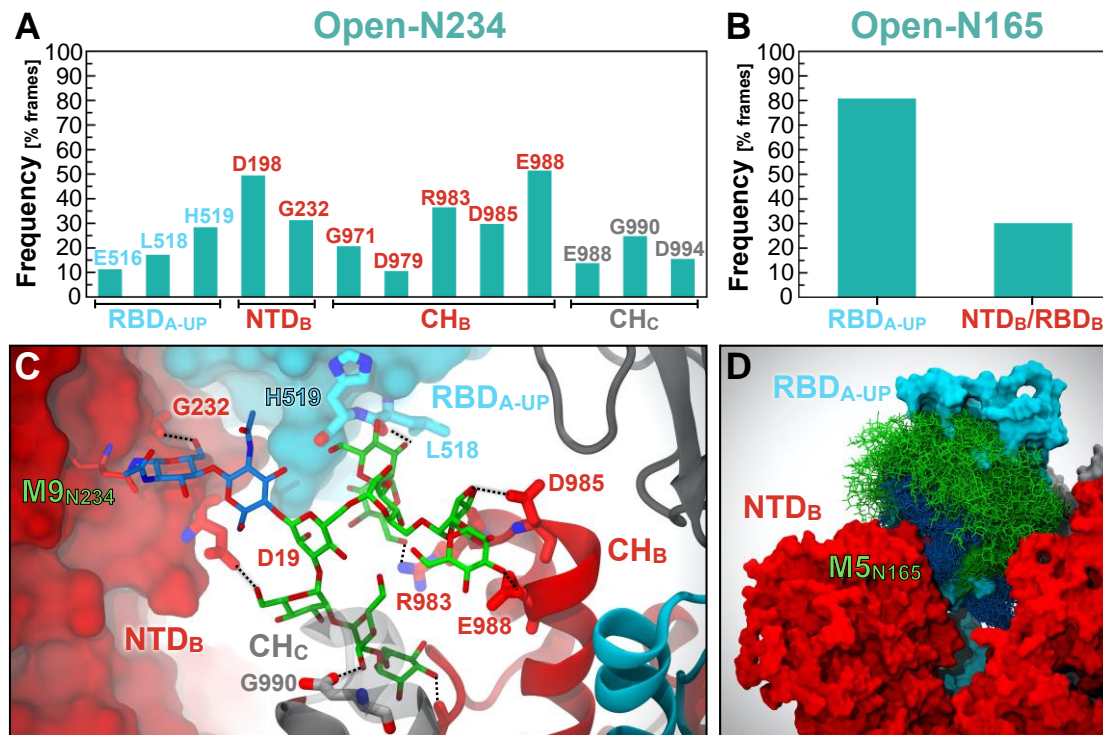
**Figure 6.4. RBD-A lateral and axial angle fluctuations.** (A, B) RBD-A lateral angle (A) and axial angle (B), where chains A, B, and C of the spike are represented as ribbons coloured in cyan, red, and grey, respectively. Positive and negative variations with respect to the initial frame (0) are indicated with the “+” and “-” symbols, respectively. (C, D) Distributions of RBD-A lateral angle (C) and axial angle (D) fluctuations with respect to the initial frame in closed (blue), open (teal), and mutant (magenta). Angle variations were calculated with respect to their value at frame 0. Frequencies have been normalized within the respective data sets.

The RBD lateral angle distributions of the open, closed and mutant systems are shown in **panel C** of **Figure 6.4**. The populations of both the open and mutant distributions agree with the PCA plot, suggesting that the inclusion of N234 and N165 glycans plays a role in stabilising the open conformation of the RBD. The axial-angle analysis (**Figure 6.4.D**) reveals a more negative trend in the mutant systems than in the open systems, indicating that the RBD without

the support of N165 or N234 tends to move closer to the centre of the spike, similar to the closed systems' distribution. In order for the RBM to be accessible to the ACE2 receptor, the RBD must occupy a metastable state that is orientated away from the central axis and exposed to the solvent and our data suggest that the N-glycans at N165 and N234 do help facilitate.

Hydrogen bonding analysis was performed to gain further insight on the interactions between the glycans at N234 and N165, and the protein residues. In the open conformation systems with a glycosylated N234 site, the Man9 glycan occupies a large volume in the void of the structure left vacant by the open RBD (**Figure 6.5.D**). Stable hydrogen bonds were observed between the glycan and H519 of the lower RBD, D198 of the NTD and R983, D985 and E988 of the central helix in the trimer, all of which were present in over 40% of the 4.2 $\mu$ s simulation (**Figure 6.5.A, Movie S.3 in Appendix IV**). In particular, we observed that the interaction with H519 at the base of the RBD contributed to support the RBD in the "up" or open conformation. The glycan at N165 is more solvent exposed than the one at N234, yet it still interacts with the RBD and with the RBM, with a range of different hydrogen bonding interactions observed in different replicas (**panels B and D of Figure 6.5, Figure S8 and Figure S9 of Appendix IV**).

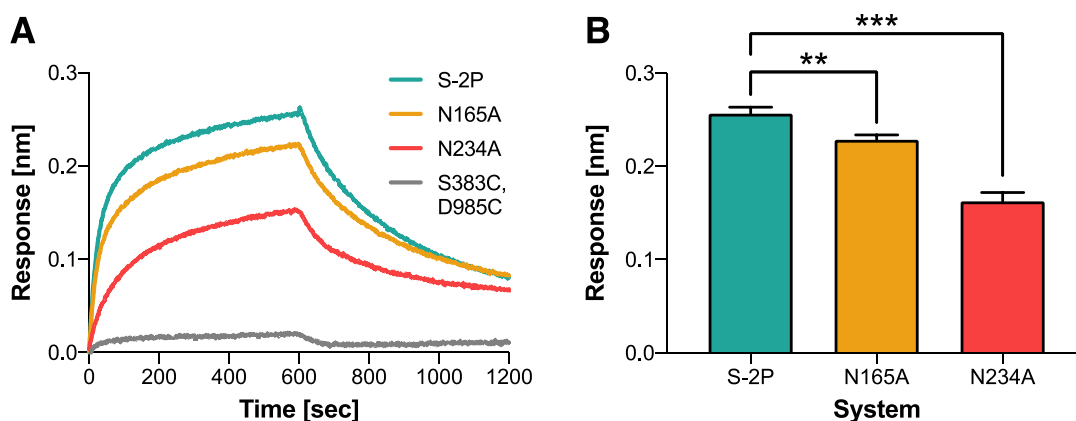




**Figure 6.5. Hydrogen bond interactions of N-glycans at N234 and N165.** The main hydrogen bond interactions of N-glycans at N234 (A) and N165 (B) within the Open system are shown as occupancy across all replicas (% frames). (C) A snapshot capturing Man<sub>9</sub> glycan (licorice representation) at N234 establishing multiple hydrogen bonds with S protein residues (licorice representation) belonging to RBD-A (cyan surface), NTD-B (red surface), CH-B (red cartoons), and CH-C (grey cartoons). GlcNAc and Man carbons are colored in blue and green, respectively. (D) Molecular representation of Man<sub>5</sub> glycan at N165 interacting with RBD-A. Multiple (1000) equally interspersed configurations (selected across all replicas) of the glycan at N165 are simultaneously shown. The glycan is represented as coloured liquorices (GlcNAc in blue, Man in green), whereas RBD-A and NTD-B are represented as cyan and red surfaces, respectively.

### Confirmation of the importance of N234 and N165 glycans by Biolayer Interferometry

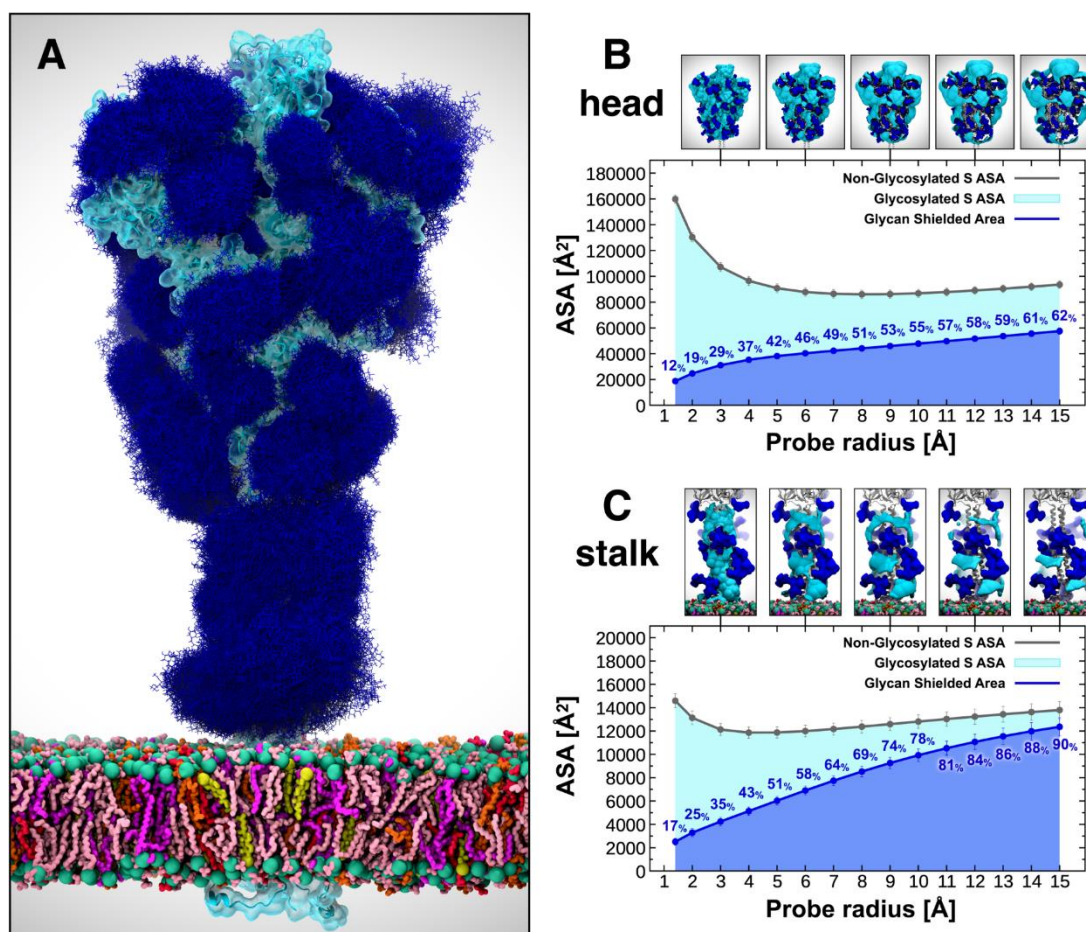
In conjunction with our computational work, Prof Jason McLellan’s lab (University of Texas at Austin) performed biolayer interferometry to measure the binding affinity of mutant S proteins to ACE2 proposed based on our simulations, to quantify experimentally the role of the N-glycans at N234 and N165 play in mediating the open and closed states of the RBD. Prof McLellan and his team expressed two mutants, namely N234A and N165A. The binding responses obtained for N165A and N234A are shown in **Figure 6.6**. Results indicate that ACE2 binding is reduced in both cases relative to the native-like S-2P variant, with N165A being 10% less effective, and N234A being 40% less effective ( $p=0.0051$  and  $p=0.0002$  respectively, Student’s T-test). A negative control spike was engineered with mutations of S383C and D985C to “lock” all three RBDs into closed conformations. The negative control spike shows no binding to ACE2.



**Figure 6.6. N234A and N165A mutations reduce RBD binding to ACE2.** (A) Bi-layer interferometry sensorgrams showing binding of ACE2 to spike variants. (B) Binding responses for bi-layer interferometry measurements of ACE2 binding to spike variants. Data are shown as mean  $\pm$  S.D. from 3 independent experiments. Asterisks represent statistical significance (Student's t test; \* $0.01 < p < 0.05$ , \*\* $0.001 < p < 0.01$ , \*\*\* $0.0001 < p < 0.001$ ).

### Analysis of the glycan shield in the full-length SARS-CoV-2 S Protein

The potential shielding of the protein by the selected glycoprofile was calculated from the S protein's accessible surface area (ASA) in the open conformation, with probe radii ranging from 1.4 to 15 Å, to understand the accessibility with regards to differential molecule size, corresponding to a water molecule to up to the approximate size of an antibody. In **panel A** of **Figure 6.7**, the 3D conformational space of each glycan is rendered using 300 superimposed snapshots from the MD trajectory, representing the dynamics each glycan undergoes within 1  $\mu$ s of simulation and mimic the numbers of conformations a glycan could adopt during antibody/S protein binding. The ASA analysis indicates the head of the protein is less shielded than the stalk (**Figure 6.7, panels B and C**). The glycosylation of the stalk is highly efficient at shielding the protein for the larger values of the probe radius range, with coverage of 90% for antibody-sized molecules, but the shield becomes far less effective when the radius is decreased to the lower end of the scale (1.4-3 Å). In comparison, the head of the protein is equally as shielded to the stalk (20% to 26% respectively) from the smallest sized probe (water molecules), but is far more vulnerable, with only 62% of the surface camouflaged from the larger sized molecules. ASA average and standard deviation values for the head and stalk are shown in Tables **S6** and **S7** of **Appendix IV**.

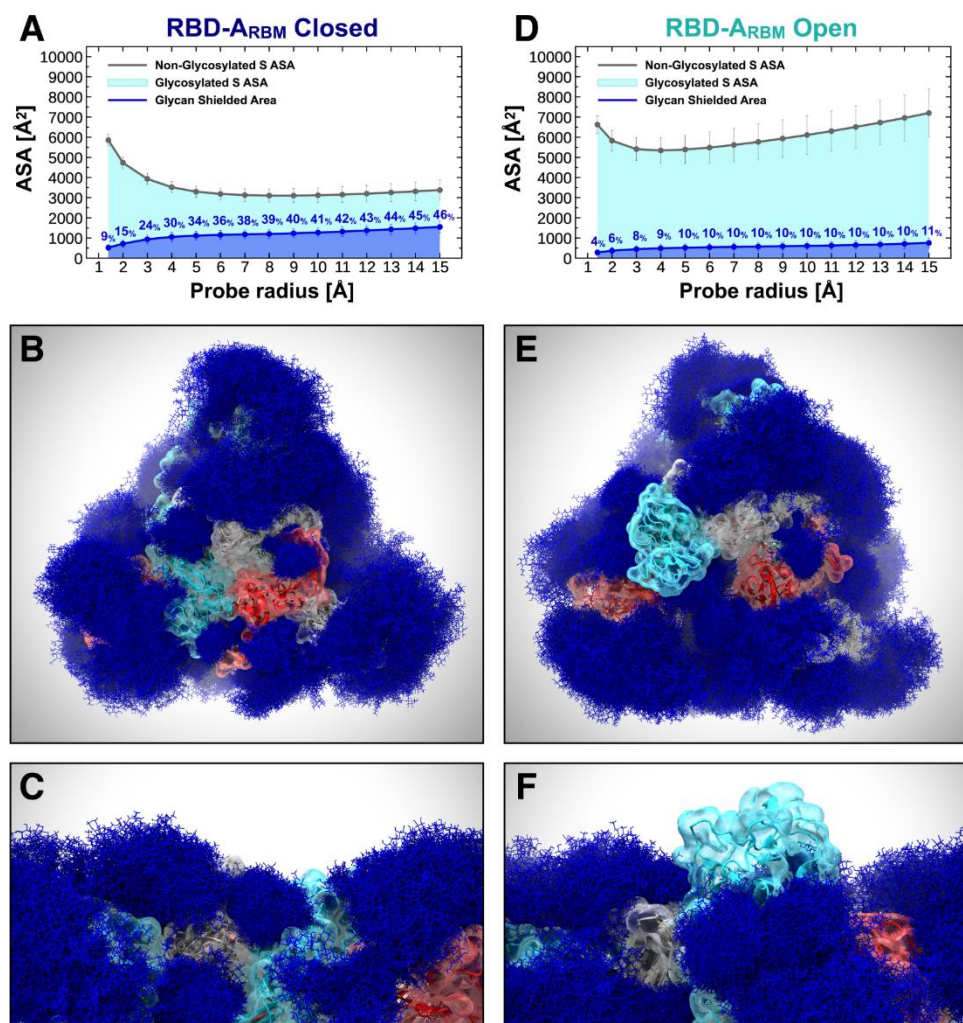


**Figure 6.7. Glycan shield of the SARS-CoV-2 S protein.** (A) Molecular representation of the Open system. Glycans at several frames (namely, 300 frames, one every 30 ns from one replica) are represented with blue lines, whereas the protein is shown with cartoons and highlighted with a cyan transparent surface. Color code used for lipid tails (licorice representation): POPC (pink), POPE (purple), POPI (orange), POPS (red), cholesterol (yellow). P atoms of the lipid heads are shown with green spheres. Cholesterol's O3 atoms are shown with yellow spheres. (B, C) Accessible surface area of the head (B) and stalk (C) and the area shielded by glycans at multiple probe radii from 1.4 (water molecule) to 15 Å (antibody-sized molecule). The values have been calculated and averaged across all replicas of Open and are reported with standard deviation. The area shielded by the glycans is presented in blue (rounded % values are reported), whereas the grey line represents the accessible area of the protein in the absence of glycans. Highlighted in cyan is the area that remains accessible in the presence of glycans, which is also graphically depicted on the structure in the panels located above the plots.

We explored more extensively the glycan coverage around the RBM (residues 400 to 508), in both open and closed systems, using varying probe radii with ASA analysis. In the closed system simulations, the coverage of the RBM is moderate for molecules larger than 3Å; averaging 35% coverage across the range of probe radius sizes (Figure 6.8.A). In the open system however, the glycan coverage remains sparse, averaging to only 9% (Figure 6.8.D). Full data for this analysis can be viewed in Tables S8-S10 in Appendix IV. For the smaller probe radii (1.4-3Å), the glycan shielding is equally ineffective in both systems. The section of the RBD not involved with ACE2 interaction is shielded in both systems (Figure S.7 in Appendix IV). This is due to the presence of the N-glycans of the RBD at N331 and



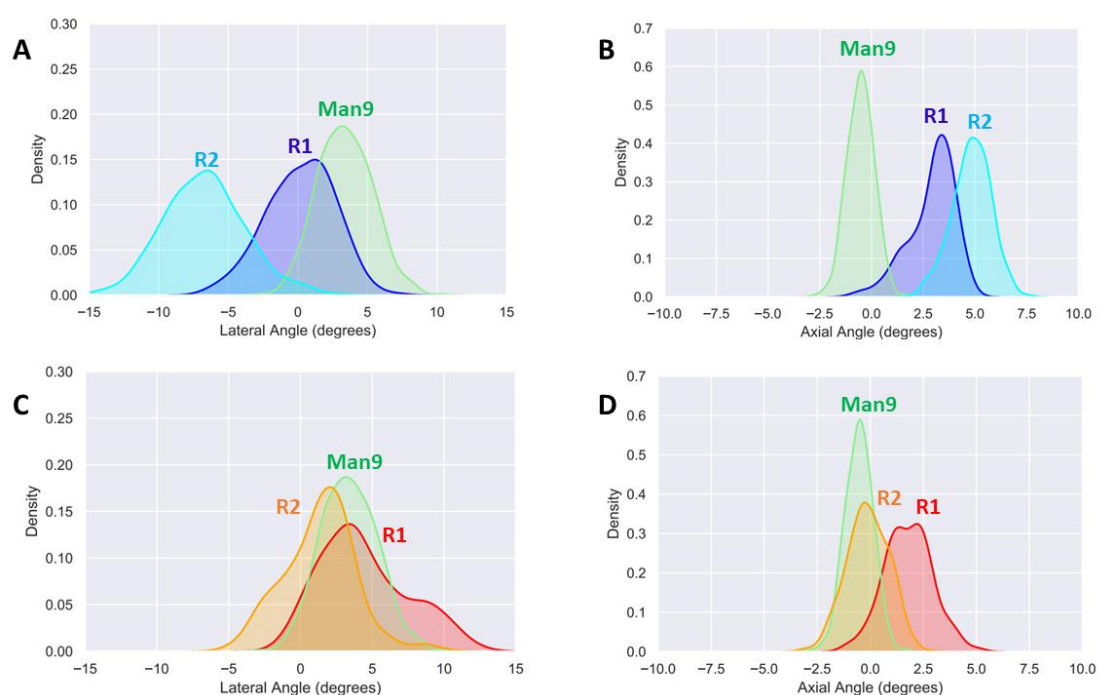
N343, and also owing to the N-glycans of the NTD N165 and N234, that we previously described interacting with the RBD (**Figure S.8 in Appendix IV**).



**Figure 6.8. Glycan shield of the RBD ACE2-interacting region.** The accessible surface area of the RBM-A and the area shielded by neighboring glycans in the Closed (**A**) and Open (**D**) systems are plotted at multiple probe radii from 1.4 (water molecule) to 15 Å. The values have been averaged across replicas and are reported with standard deviation. In blue is the area of the RBM-A covered by the glycans (rounded % values are reported), whereas the gray line is the accessible area in the absence of glycans. Highlighted in cyan is the RBM-A area that remains accessible in the presence of glycans, which is also graphically depicted on the structure in the panels located below the plots. (**B-F**) Molecular representation of Closed and Open systems from top (**B** and **E**, respectively) and side (**C** and **F**, respectively) views. Glycans (blue lines) are represented at several frames equally interspersed along the trajectories (300 frames along 0.55 ns for Closed and 1.0 μs for Open), while the protein is shown with colored cartoons and transparent surface (cyan, red and gray for chains A, B and C, respectively). Importantly, in panel **E** and **F**, RBD within chain A (cyan) is in the ‘up’ conformation and emerges from the glycan shield.

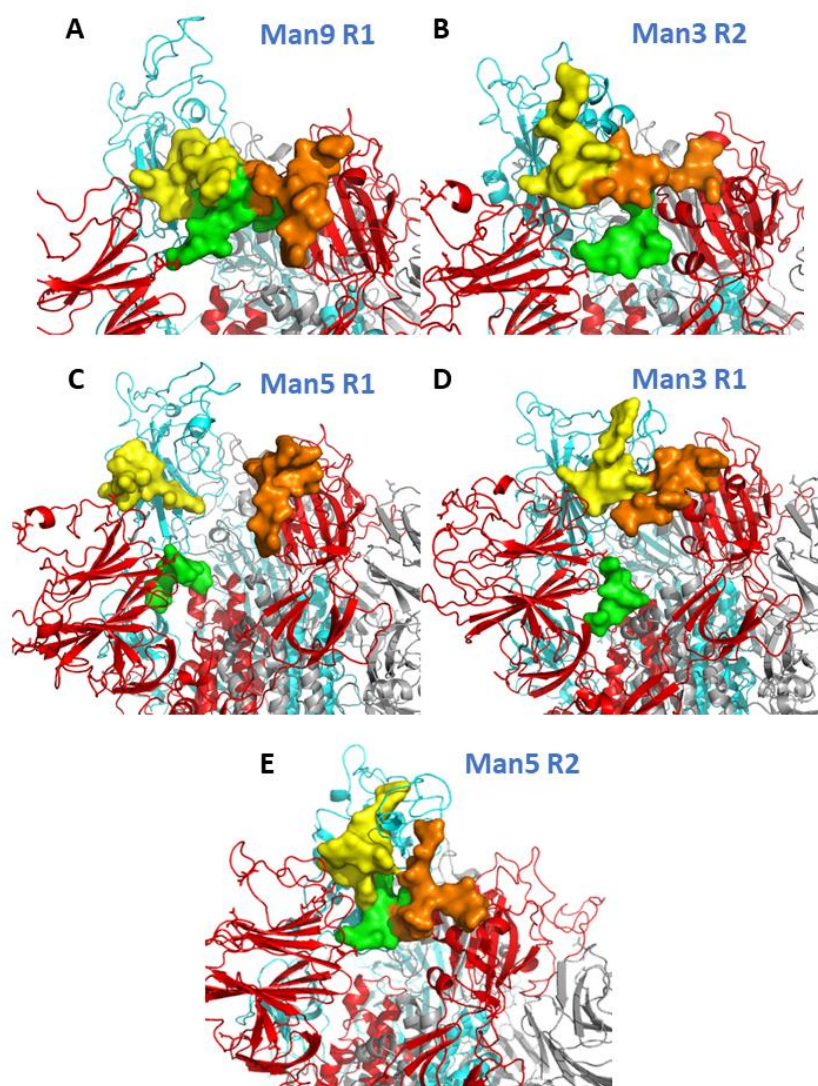
## Further simulation of SARS-CoV-2 ectodomain with varying glycosylation at N234

To determine the impact of the type of glycosylation at N234 in the stability of the open conformation, we ran set of five uncorrelated conventional MD trajectories of the S ectodomain. Data from the simulations were collected after 300 ns of conformational equilibration. This rather long time interval was deemed necessary to allow the system to adjust to the presence of smaller glycans at N234, such as Man5 and Man3, which are not characteristic of the stable S-2P prefusion trimer. Conformational equilibration was estimated in terms of RMSD values of the protein backbone with the cryo-EM structure used as a reference. Lateral angle and axial angle calculations were performed using the same method reported above for the first set of simulations. The resultant distributions of the lateral and angle calculations over the equilibrated period of the simulations can be seen in **Figure 6.9**. In comparison to the open, closed and mutant systems of the initial set, the overall differences between distributions is subtler. The Man9 model used for comparison to the HEK trimer open system distributions (**Figure 6.4, panels C and D**). The smaller range of motions measured can be attributed to an averaging over six replicas for the open system in the previous work<sup>7</sup> versus the single replica of the Man9 model used here. For the two replicas R1 and R2 of Man5 (**panels A and B, Figure 6.9**), their distributions have the same frequency and range of motion, with Man5 R2 deviating more dramatically from its starting conformation. This deviation suggests that a different conformational space was explored, even with the same glycosylation pattern, and this depends on the different interactions formed by the Man5 at N234, and by the N165 and N343 complex N-glycans. The two replicas, R1 and R2, of the Man3 models have slightly different distributions (**panels C and D, Figure 6.9**). The lateral and axial motions of R2 of Man3 are more similar to the Man9 model than to the Man3 R1, which suggests a more stabilised interaction occurring with the RBD in R1, compared to R2.



**Figure 6.9. RBD-B lateral and axial angle fluctuations.** Panels **A** and **B** represent the lateral and angle calculations for replicas R1 (blue) and R2 (cyan) of the model with Man5 at N234. Panels **C** and **D** show the replicas R1 (red) and R2 (orange) of the model with Man3 at N234. The distributions from the Man9 model are included in each graph (green), for reference, as it displays the most stabilised distributions amongst the five separate simulations. Positive and negative variations with respect to the initial frame (0) are indicated with the “+” and “-” symbols, respectively.

We examined the interactions of N234 in each system through hydrogen bonding analysis to gain more insight on the difference in RBD motions between replicas. For the Man9 model, an intricate hydrogen bonding network is formed between the Man9 and the biantennary complex N-glycans at N165 and N343, which has a stabilising effect on the RBD (see **panels A and B in Figure 6.10**). A similar hydrogen bonding network is observed in Man3 R1, which explains the results obtained in the analysis of the lateral and axial angle distributions. While Man3 is substantially smaller than Man9 and so cannot occupy the void created by the open conformation of the RBD, our simulations show that it is indeed long enough to form a stable interaction with the RBD-C at residue N370. This interaction is present for 49% of the production length. Man3 is also found to interact with the glycan at N343. In summary, in both simulations of Man5 and Man3 at N234, a complex hydrogen bonding network involving the glycans at N343 and N234, coupled with the glycan at N165 interacting with either N234 and N343 glycans and with the RBD residues produces a stabilising effect similar to what seen previously with Man9 at N234. Furthermore, the diversity of the RBD orientation sampled in presence of these glycans at N234 suggest a higher dynamics of the RBD that can occupy an ensemble of metastable states between the open and closed conformations identified with a very large oligomannose such as Man9, characteristic of the highly stable S-2P prefusion trimer. Relative populations of these states are modulated by carbohydrate-carbohydrate and carbohydrate-protein interactions and therefore depend on the type of glycosylation at these sites.



**Figure 6.10. Interactions of N165, N234 and N343 glycans.** Representative snapshots of the N-glycans at N165 (yellow), N343 (orange) and N234 (green) interacting with the RBD-B (cyan), or the NTD-C and RBD-C (red), and/or each other. These snapshots capture the more stabilised h-bonding interactions of the terminal glycan of the antennae in each glycan.

Indeed, in Man5 R1 and Man3 R2, there is no interaction between the three glycans at NN234, N165 and N343. The Man3 N234 glycan in R1 explores a completely new region, inaccessible to the larger mannose structures, where it enters the cleft between the NTD-C and RBD-C, interacting with the top of the central helix. With the lack of any direct or indirect support of the RBD from Man3 in this trajectory, RBD-B appears to be less ‘anchored’, thus more dynamic. In Man5 R1, the Man9 at N234 and the complex N-glycan at N165 resume supportive roles to the RBD, propping it up. Unlike Man9, Man5 is shorter and it is not able to reach the residues of the trimer’s core helices. Moreover, Man9 because of its size also forms more stable hydrogen bonds with the NTD-C and with the lower hinge section of RBD-B, which helps to prop up the RBD, or somewhat restrict lateral motion in the direction of RBD-C.

## 6.4 Discussion

Despite the differences in the systems sizes, setups, original cryo-EM structure (6VYB vs. 6VSB), force field parameter sets (AMBER vs. CHARMM), MD software packages (Amber vs. NAMD), in the running protocols and details in the models' glycosylation, the two sets of initial simulations performed by the Fadda and Amaro group, respectively, converge in describing a very uniform picture of the role of the glycan shield in the activation of the SARS-CoV2 S protein. Through PCA, angle and hydrogen bonding analysis, and confirmed by the results of the bilayer interferometry experiments, we show that the absence of glycosylation at position N234 and at both positions N234 and N165 affects the conformational plasticity of the open RBD, allowing it to explore a larger conformational freedom, which indicates a degree of instability, supported by its lower degree of binding the ACE2 receptor. The equilibrium of the open and closed conformations of the RBD can be affected by relatively small modifications, like single point mutations within and around the RBD<sup>35</sup>, or also by varying pH conditions<sup>36</sup>. Similar to the RBD of the MERS S protein<sup>37</sup>, the SARS-CoV-2 open conformation is metastable within its conformational ensemble. We identified two glycosylation sites, N165 and N234, where the glycans can actively interact with the open RBD and “lock” it into an open position. While full conformational shift from the open to the closed conformation cannot be observed within this timescale, or possibly with conventional MD, we can still investigate the relative stability of the open conformation, based on its dependency on the surrounding glycosylation. This reveals a possible vulnerability of the SARS-CoV-2 S protein if the open conformation destabilised, thus decreasing its ability of binding to the ACE2 receptor. Through bilayer interferometry experiments on both N234A and N165A mutations of the S-2P protein, we have seen an agreement with the simulation dynamics, that both N-glycans play a role in stabilisation. The results suggest that the presence of N234 may have a greater effect on the RBD binding affinity, and therefore RBD stability than N165, which is more solvent exposed. The hydrogen bonding analysis however does support the role of the glycan at N165 as a stalwart support, interacting with the open RBD steadily throughout the simulations.

Our second set of simulations helped us to assess the importance of the N-glycan architecture at N234. With these simulations we were able to identify multiple uncorrelated structures that allow us to sample various degrees of stabilisation of the open conformation, determined by the accessible interactions of the oligomannose at N234, in combination with N165 and N343 glycosylated with complex N-glycans. In the original work<sup>7</sup> the glycan at N343 was not specifically investigated as a contributing factor in the stability of the RBD open conformation as in none of those trajectories it was found to interact with the open RBD or with the glycans at N234 and N165. Yet, in our original simulations of the S ectodomain, N165 and N343 were glycosylated with Man5 N-glycans, and thus their “reach”, in terms of the length

of their antennae/arms, was limited. In our second set of simulations, we clearly see the role of N343 is more crucial, as it appears to contribute to a more intricate hydrogen bonding network with N165 and N234, that stabilises the open RBD more effectively. In all replicas of this second set of S protein dynamics, N165 is always in contact with the RBD, reinforcing the idea that N165 is essential for the mediation of a stabilised open RBD. As for N234, the Man3 variants suggest that N-glycosylation of any kind at N234 is fundamental, as the difference in the mutant systems behaviour versus the Man3 models is stark. While Man3 may not be able to stabilise the RBD as effectively as Man5 or Man9 can in this position, so long as it can form an interaction with the opposing RBD, bridging the opening of the void left by the open RBD, it can still form stabilising interactions, suggesting that any glycan at N234 is better than no glycan, provided that the positions at N165 and N343 are fully glycosylated. Interestingly, every oligomannose type substituted at N234 is capable of forming an interaction with N370 of the opposite RBD-C. An important mutation found in SARS-CoV-2 S relative to the SARS-CoV S causes the loss of glycosylation at N370, part of a sequon in SARS-CoV, which suggests that a glycan at that location may interfere with the stabilisation abilities of the N234 glycan, potentially making SARS-CoV S protein less stable in the open conformation than SARS-CoV-2. Further investigation into the behaviour and possible function of the occupied N370 glycosylation site is currently ongoing in our group.

The SARS-CoV-2 unique and extensive glycosylation pattern may be a strength and a weakness for the virus, as we propose that the glycan coverage not only can shield immunogenic epitopes, but plays a functional role in the mechanics of RBD opening and closing, which can be targeted and utilised for vaccine design<sup>27,38,39</sup>. Considering our analysis of the ASA of the S protein, the RBM and RBD are obvious candidates for targeting, as the RBM is concealed and well camouflaged in the closed conformation, but is exposed in the open conformation. This agrees with structural data that reports the open conformation of the RBD is a requirement<sup>40</sup> for neutralization by host antibodies. Several antibodies targeting the S protein have been identified (see **Table S11 in Appendix IV**<sup>38,41-50</sup>). The RBD is the major target amongst these antibodies, with few selecting other antigenic regions in the NTD or CD (**Figure S9 in Appendix IV**). The S protein dynamics can be manipulated through introducing mutations to destabilise the open RBD conformation (N234A and N165A) or changes in glycosylation at N234, N165 or N343, reducing the binding affinity of the RBDs, but leaving the RBM accessible for antibody targeting.



## 6.5 Conclusions

Through this collaborative work with the Amaro group at UCSD and Mc Lellan group at UTexas based on extensive atomistic MD simulation of different models of the SARS-CoV-2 S protein supported by biolayer interferometry experiments, we were able to describe a new and unique role of the viral glycan shield. Indeed, we showed how the time-averaged glycan shield hides a vast amount of the S protein surface area from recognition by the immune system and how this shielding effects varies depending on the protein's conformational state, i.e. between open and closed states. Additionally, the glycans around the RBD-, namely at positions N234, N165, and N343 interact directly (through glycan-protein interactions) or indirectly (through glycan-glycan interactions) with the open RBD, modulating its dynamics and stability in the open conformation. The mutant variations (N234A and N165A) reveal the dependency the open conformation has on these glycans, demonstrated by and the markedly reduced binding affinity to ACE2 without them. With this knowledge, there is potential for taking advantage of this dependence upon glycosylation as a vulnerability of the open RBD state to target the S protein in therapeutic strategies.

## References

1. Liu, D. X., Liang, J. Q. & Fung, T. S. B. T.-R. M. in L. S. Human Coronavirus-229E, -OC43, -NL63, and -HKU1. in (Elsevier, 2020). doi:<https://doi.org/10.1016/B978-0-12-809633-8.21501-X>.
2. Peiris, J. S. M., Guan, Y. & Yuen, K. Y. Severe acute respiratory syndrome. *Nat. Med.* **10**, S88–S97 (2004).
3. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
4. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
5. Heald-Sargent, T. & Gallagher, T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* **4**, 557–580 (2012).
6. Delmas, B. & Laude, H. Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J. Virol.* **64**, 5367 LP – 5375 (1990).
7. Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).
8. Apellániz, B., Huarte, N., Largo, E. & Nieva, J. L. The three lives of viral fusion peptides. *Chem. Phys. Lipids* **181**, 40–55 (2014).
9. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (80-. )*. **369**, 330 LP – 333 (2020).
10. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).
11. Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta - Gen. Subj.* **1863**, 1480–1497 (2019).
12. Watanabe, Y. *et al.* Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *bioRxiv* 2020.02.20.957472 (2020) doi:10.1101/2020.02.20.957472.
13. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol.* **3**, 237–261 (2016).



14. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* (80-. ). **309**, 1864 LP – 1868 (2005).
15. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
16. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* (80-. ). **367**, 1260 LP – 1263 (2020).
17. Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, e238–e244 (2020).
18. Zhao, P. *et al.* Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **28**, 586-601.e6 (2020).
19. Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci.* **114**, 11157 LP – 11162 (2017).
20. Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K. & Whittaker, G. R. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J. Mol. Biol.* **432**, 3309–3325 (2020).
21. Bestle, D. *et al.* TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. alliance* **3**, e202000786 (2020).
22. Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450 (2009).
23. Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci.* **114**, E7348–E7357 (2017).
24. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **30**, 981–988 (2020).
25. Zhang, Y. *et al.* Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. *bioRxiv* 2020.03.28.013276 (2020) doi:10.1101/2020.03.28.013276.
26. Sanda, M., Morrison, L. & Goldman, R. N and O glycosylation of the SARS-CoV-2 spike protein. *bioRxiv* 2020.07.05.187344 (2020) doi:10.1101/2020.07.05.187344.
27. Shulla, A. & Gallagher, T. Role of spike protein endodomains in regulating coronavirus entry. *J. Biol. Chem.* **284**, 32725–32734 (2009).
28. Thorp, E. B., Boscarino, J. A., Logan, H. L., Goletz, J. T. & Gallagher, T. M. Palmitoylations

- on Murine Coronavirus Spike Proteins Are Essential for Virion Assembly and Infectivity. *J. Virol.* **80**, 1280 LP – 1289 (2006).
29. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
  30. Harbison, A. M., Brosnan, L. P., Fenlon, K. & Fadda, E. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **29**, 94–103 (2019).
  31. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).
  32. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **29**, 622–655 (2008).
  33. Jorgensen, W. L. & Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **19**, 1179–1186 (1998).
  34. Case DA, V. B., Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE III, Darden TA, R.E.; Duke HG, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A.; Lee TS, S. L., Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitbe, X. . K. W. a. P. A. AMBER 18. (2018).
  35. Henderson, R. *et al.* Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **27**, 925–933 (2020).
  36. Urbanowicz, R. A. *et al.* Antigenicity and Immunogenicity of Differentially Glycosylated Hepatitis C Virus E2 Envelope Proteins Expressed in Mammalian and Insect Cells. *J. Virol.* **93**, e01403-18 (2019).
  37. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (80-. ).* **367**, 1444 LP – 1448 (2020).
  38. Zost, S. J. *et al.* Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**, 443–449 (2020).
  39. Jiang, S., Hillyer, C. & Du, L. Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. *Trends Immunol.* **41**, 355–359 (2020).
  40. Gui, M. *et al.* Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017).

41. Huang, Y., Yang, C., Xu, X., Xu, W. & Liu, S. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
42. Ejemel, M. *et al.* IgA MAb blocks SARS-CoV-2 Spike-ACE2 interaction providing mucosal immunity. *bioRxiv* 2020.05.15.096719 (2020) doi:10.1101/2020.05.15.096719.
43. Huo, J. *et al.* Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host Microbe* **28**, 445-454.e6 (2020).
44. Wu, Y. *et al.* A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science (80-. )*. **368**, 1274 LP – 1278 (2020).
45. Pinto, D. *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
46. Wang, C. *et al.* A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **11**, 2251 (2020).
47. Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science (80-. )*. **369**, 650 LP – 655 (2020).
48. Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science (80-. )*. **369**, 643 LP – 650 (2020).
49. Tai, W., Zhang, X., He, Y., Jiang, S. & Du, L. Identification of SARS-CoV RBD-targeting monoclonal antibodies with cross-reactive or neutralizing activity against SARS-CoV-2. *Antiviral Res.* **179**, 104820 (2020).
50. Wrapp, D. *et al.* Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell* **181**, 1004-1015.e15 (2020).

## Chapter 7: Conclusions

In my PhD research presented in this thesis, I explored the sequence-to-structure-to-function relationships in complex N-glycans, by means of HPC molecular simulations by molecular dynamics. Within this framework, all-atom simulations can be considered loosely as a form of computational microscopy, where we can observe the structural evolution of biomolecules at the molecular scale. Using conventional and enhanced sampling schemes, I provided the necessary atomistic level of detail to define how sequence and branching determines the 3D structural propensity of a given N-glycan.

In **Chapter 3**, starting from the smallest disaccharide motif that occurs in every N-glycan, i.e. the core GlcNAc- $\beta$ 4-GlcNAc chitobiose, I was able to build and analyse progressively larger complex N-glycans and to determine the role each specific monosaccharide-linkage addition has in affecting the overall structure. To achieve this goal, I performed virtually exhaustive sampling experiments, considering all possible rotamers for each unique glycan sequence in separately run simulations. This extensive sampling scheme revealed that N-glycans occupy distinct conformations of varying population at room temperature. This result debunks the idea that N-glycans are unstructured, and allows us to extract the most populated conformations from the “blur” of the conformational ensembles observed at experimental timescales, such as high microsecond to milliseconds for NMR and much higher for X-ray crystallography data collection.

The significance of the N-glycan sequence-to-structure relationship was further explored in **Chapter 4**, where I described the effects of modifications of the N-glycans’ sequence, found in plant and/or invertebrates and non-natural alterations to mammalian sequences, to the N-glycans’ structure and dynamics. We found that by changing the linkage, e.g. from  $\alpha$ 1-6 to  $\alpha$ 1-3 linked core-fucose, or by adding one monosaccharide, like  $\beta$ 1-2 xylose, we can significantly alter the conformational space occupied by the adjacent monosaccharides, and thus influence the overall conformational landscape of the N-glycan. By quantifying the effect of each specific modification through conformational analysis, we were able to discretize the intrinsic 3D architecture of N-glycans in terms of specific groupings of monosaccharides, which we named “glycoblocks”. Contrary to the standard “monosaccharide sequence” viewpoint, this approach allows us to define more clearly 3D structure dependencies. As a step further in this direction, we are planning to build a structural database of glycoblock units for docking and molecular recognition studies, we named *GlycoShape*.

The general validity of this glycoblock viewpoint to describe the intrinsic glycan 3D structure and dynamics, became quite useful when we studied different glycoproteins, as discussed in **Chapters 5 and 6**, to understand the functional role of the N-glycans and of specific glycosylation patterns. For example, in **Chapter 5** we have shown how and why the variation in N-glycosylation, more specifically sialylation and core-fucosylation, of the two N-glycans in the IgG1 Fc region modulates the relative

stability and dynamics of the Fc protein, and in turn how it may affect the binding affinity to the Fc  $\gamma$  receptor. Indeed, core-fucosylation of the Fc N-glycans is known to dramatically reduce ADCC, weakening the antibody effector function, thus its efficacy in biopharmaceutical applications.

In **Chapter 6** I described how glycosylation plays a crucial role in viral glycoproteins, and more specifically in the SARS-CoV-2 spike fusion protein. Through extensive sampling, we have shown how the extent and type of glycosylation in these heavily glycosylated proteins are imperative for the virus evasion of the host immune system, through shielding and mimicry. Moreover, as a unique feature of the SARS-CoV, within a collaborative effort we discovered an additional functional role of N-glycans at N234, N165 and N343, whereby the N-glycans at these sites structurally support the receptor binding domain (RBD) open conformation and mutations have shown strongly diminished binding affinity. The extent of the interaction and support provided by the N-glycans at these glycosylation sites is dependent on the N-glycan type and sequence.

As a whole, my PhD work has contributed to the discovery of new structural dependencies of complex N-glycans from their sequence, bringing in a three-dimensional dynamics perspective to our understanding of glycans' molecular recognition and more in general of their function. The discretization of the N-glycan architecture in terms of 3D glycoblocks has allowed us to understand how the addition or removal of single monosaccharides can radically alter the overall structural ensemble, an approach that can be used purposely to devise synthetic strategies to obtain complex glycosylation patterns, e.g. to achieve high or complex degrees of functionalization of the 1-6 arm. Within a glycoprotein context, we were able to explain at the atomistic level of detail for the first time how glycosylation impacts on antibody activity. This insight opens the door to further studies, where custom-designed (non-immunogenic) glycoforms can be tested for activity based on the profile obtained by computational models. Indeed, the effect imparted by specific glycoforms depends exclusively on their 3D structure and dynamics, inherent to their sequence/identity. Within this framework, another aspect that my studies have contributed to highlight is a paradigm shift of how glycan structure is viewed. We found in all our studies that the complex N-glycans' structure is not actively changed by the protein landscape, but as an equal player, it adapts to it, depending on its allowed degrees of freedom. Because of this ability, different glycosylation types can modulate differently the function of a glycoprotein, and we have shown that this is particularly true in the SARS-CoV2 S proteins, where glycans play a functional role. Therefore, we (and others) see that targeting glycosylation is an extremely promising therapeutic strategy against viral infection<sup>1-4</sup>.

## References

1. Doores, K. J. The HIV glycan shield as a target for broadly neutralizing antibodies. *FEBS J.* **282**, 4679–4691 (2015).
2. Williams, S. J. & Goddard-Borger, E. D.  $\alpha$ -glucosidase inhibitors as host-directed antiviral agents with potential for the treatment of COVID-19. *Biochem. Soc. Trans.* **48**, 1287–1295 (2020).
3. Watanabe, Y. *et al.* Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *bioRxiv* 2020.02.20.957472 (2020) doi:10.1101/2020.02.20.957472.
4. Nascimento da Silva, L. C. *et al.* Exploring lectin–glycan interactions to combat COVID-19: Lessons acquired from other enveloped viruses. *Glycobiology* (2020) doi:10.1093/glycob/cwaa099.

## Appendices

**Appendix I:** Supplementary material from the paper: A.M. Harbison, L. P. Brosnan, K. Fenlon, E. Fadda, Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: A molecular dynamics study, *Glycobiology*. 29, 94–103 (2019).

**Appendix II:** Supplementary material from the paper: C. A. Fogarty\*, A. M. Harbison\*, A. R. Dugdale, E. Fadda, How and why plants and human N-glycans are different: Insight from molecular dynamics into the ‘glycoblocks’ architecture of complex carbohydrates, *Beilstein J. Org. Chem.* 16, 2046–2056 (2020)

**Appendix III:** Supplementary material from the paper: A.M. Harbison, E. Fadda, An atomistic perspective on antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics, *Glycobiology*. 30, 407–414 (2020)

**Appendix IV:** Supplementary material from paper: L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, R. E. Amaro, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein, *ACS Cent. Sci.* 6, 1722–1734 (2020)

## Appendix I

### **Computational Methods**

All glycans were built using the Glycam Carbohydrate Building tool (<http://www.glycam.org>). The appropriate rotamers of each glycan were chosen in order to generate a complete set of starting structures for the molecular dynamics simulations. Once generated, the glycan pdb files were prepared using *tleap* (Salomon-Ferrer, R., Case, D.A., et al. 2013). For each starting structure topology and coordinate files were created with the GLYCAM06h-12SB forcefield (Kirschner, K.N., Yongye, A.B., et al. 2008) and TIP3P water model (Jorgensen, W.L. and Jenson, C. 1998) was included to represent the counterions for the sialylated structures.

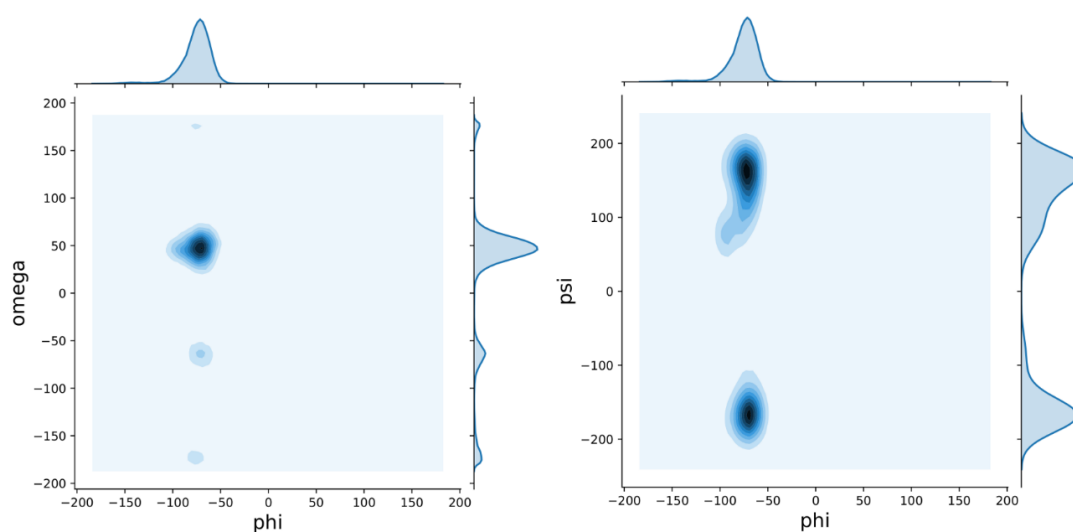
The energy minimisations and MD simulations were carried out using AMBER 12 and AMBER 16 (Salomon-Ferrer, R., Case, D.A., et al. 2013). Each system was initially minimized through 500,000 cycles of steepest descent, with a restraint of 5 kcal/mol\*Å on all heavy atoms. Following minimisation, the system was heated from 0 to 300 K over two stages, with restraints on heavy atoms in place. For the first stage, the system was heated from 0 to 100 K over 500 ps at constant volume. For the second stage the system was heated from 100 to 300 K over 500 ps. All degrees of freedom were equilibrated at 300 K and 1 atm of pressure for 5 ns. A production step followed, with a minimum of 250 ns run for each set of glycans. The overall length of the production run was set in function of the convergence of the system. The whole study entailed a cumulative simulation time in excess of 62  $\mu$ s. Trajectory files were generated for each simulation, and were visualized using VMD (Humphrey, W., Dalke, A., et al. 1996). Conformational analysis of the torsion angles was performed on each trajectory using VMD, Microsoft Excel and seaborn for the contour plots.

trajectory. 2D contour plot were made with seaborn ([seaborn.pydata.org](http://seaborn.pydata.org)) based on 30,000 data points.

As an interesting note, because of the better scaling on our machines of v. 4.6.3 and 5.0.x of GROMACS (GMX) for the calculations on these relatively small systems, we ran some tests on the medium-sized sugar H, see **Figure S1**, and compared the results with the GLYCAM/AMBER set-up. The starting structure and parameter files obtained from the carbohydrate builder on GLYCAM-WEB were converted to GMX format with the AnteChamber PYthon Parser interfacE (*acpype.py*) tool (Sousa da Silva, A.W. and Vranken, W.F. 2012). It is important to note that in all GMX simulations the 1-4 scaling was re-set to "1" as required by the GLYCAM force field (Fadda, E. and Woods, R.J. 2010, Kirschner, K.N., Yongye, A.B., et al. 2008). Equal amount of sampling was done with both GMX and AMBER,



preceded by a very similar set-up and equilibration protocols. Details of the GMX protocol are provided as Supplementary Material. The results indicate large differences in  $\alpha(1-6)$  torsions populations between GMX and GLYCAM/AMBER, shown in **Tables S.1** and **S.2**. The reason for this may be problems in the transfer of torsional parameters from a GLYCAM/AMBER format to the GMX format. More specifically we found that simulations of sugar H with GMX do not reproduce the correct conformer populations, or give energetically disfavoured conformers, such as the *tg* in the core fucose  $\alpha(1-6)$  linkage as the highest populated for sugar H2, see **Table S.2**.



**Figure S. 1** Conformational propensity of the Fuc(12)- $\alpha(1-6)$ -GlcNAc(1) linkage during a conventional (unbiased) 3  $\mu$ s MD

**Table S. 1** Comparison of the average values of the torsion angles for all glycosidic linkages of sugar H1 calculated during 500 ns MD trajectory obtained with Amber 12 and with Gromacs v. 4.6.3. Averages, standard deviations, and population analysis are calculated over 5000 frames.

<i>Sugar H1</i>	<i>GlcNAc(1)-<math>\beta(1-4)</math>-GlcNAc(2)</i>		
	$\phi$	$\psi$	
<b>Amber</b>	-77.1 (9.7) 100	-126.2 (14.5) 100	
<b>GMX</b>	-77.4 (9.9) 100	-126.5 (15.0) 100	
	<i>Fuc-<math>\alpha(1-6)</math>-GlcNAc(1)</i>		
	$\phi$	$\psi$	$\omega$

<b>Amber</b>	-79.6 (17.7) <b>100</b>	-186.1 (19.6) <b>78</b> / 92.1 (19.1) <b>20</b> / -98.6 (17.0) <b>2</b>	44.8 (11.7) <b>96</b> / -169.5 (13.9) <b>3</b> / -45.7 (30.4) <b>1</b>
<b>GMX</b>	-78.6 (18.5) <b>100</b>	-184.4 (20.0) <b>80</b> / 91.3 (19.3) <b>19</b> / -91.8 (19.0) <b>1</b>	45.6 (12.3) <b>98</b> / -171.5 (16.1) <b>1</b> / 54.8 (38.8) <b>1</b>
<i>Man(3)-β(1-4)-GlcNAc(2)</i>			
	$\phi$	$\psi$	
<b>Amber</b>	-77.4 (18.8) <b>100</b>	-123.6 (15.0) <b>100</b>	
<b>GMX</b>	-79.6 (22.2) <b>100</b>	-126.2 (16.8) <b>100</b>	
<i>Man(5)-α(1-6)-Man(3)</i>			
	$\phi$	$\psi$	$\omega$
<b>Amber</b>	74.3 (15.0) <b>100</b>	-178.3 (19.3) <b>53</b> / 87.8 (17.0) <b>46</b> / -100.4 (18.5) <b>1</b>	52.0 (10.4) <b>92</b> / -172.6 (16.0) <b>6</b> / -75.4 (14.9) <b>2</b>
<b>GMX</b>	73.2 (15.9) <b>100</b>	-177.9 (19.1) <b>54</b> / 80.6 (18.4) <b>44</b> / -92.9 (18.1) <b>2</b>	50.5 (10.5) <b>56</b> / -175.8 (16.9) <b>40</b> / -77.0 (19.3) <b>5</b>
<i>GlcNAc(7)-α(1-2)-Man(5)</i>			
	$\phi$	$\psi$	
<b>Amber</b>	-81.5 (15.8) <b>100</b>	162.4 (13.1) <b>91</b> / 106.3 (12.4) <b>9</b>	
<b>GMX</b>	-81.2 (17.1) <b>100</b>	161.5 (15.3) <b>87</b> / 107.5 (12.0) <b>13</b>	
<i>Man(4)-α(1-3)-Man(3)</i>			
	$\phi$	$\psi$	
<b>Amber</b>	72.8 (12.2) <b>100</b>	145.8 (12.6) <b>60</b> / 102.7 (14.4) <b>40</b>	
<b>GMX</b>	73.1 (12.2) <b>100</b>	145.8 (13.6) <b>59</b> / 101.7 (14.8) <b>41</b>	
<i>GlcNAc(6)-α(1-2)-Man(4)</i>			
	$\phi$	$\psi$	
<b>Amber</b>	-80.5 (16.1) <b>100</b>	161.4 (14.1) <b>83</b> / 105.2 (12.9) <b>17</b>	
<b>GMX</b>	-94.3 (17.0) <b>100</b>	161.1 (14.2) <b>80</b> / 104.6 (12.6) <b>20</b>	

**Table S. 2 Comparison of the average values of the torsion angles for all glycosidic linkages of sugar H2 calculated during 500 ns MD trajectory obtained with Amber 12 and with Gromacs v. 4.6.3. Averages, standard deviations, and population analysis are calculated over 5000 frames.**

<i>Sugar H2</i>	<i>GlcNAc(1)-β(1-4)-GlcNAc(2)</i>		
	$\phi$	$\psi$	
<b>Amber</b>	-77.4 (9.8) 100	-126.4 (14.4) 100	
<b>GMX</b>			
	<i>Fuc-α(1-6)-GlcNAc(1)</i>		
	$\phi$	$\psi$	$\omega$
<b>Amber</b>	-79.2 (18.1) 100	-185.2 (19.1) 78/ 92.3 (18.6) 21/ -102.6 (17.7) 1	44.9 (11.7) 95/ -173.7 (13.3) 4/ -38.8 (26.9) 1
<b>GMX</b>	-73.4 (13.4) 100	-183.6 (17.1) 91/ 94.6 (19.9) 4/ -87.6 (18.0) 5	-64.7 (12.1) 76/ 47.4 (12.0) 14 / -166.7 (13.5) 10
	<i>Man(3)-β(1-4)-GlcNAc(2)</i>		
	$\phi$	$\psi$	
<b>Amber</b>	-78.0 (19.6) 100	-124.8 (15.1) 100	
<b>GMX</b>	-79.8 (24.9) 100	-125.6 (15.9) 100	
	<i>Man(5)-α(1-6)-Man(3)</i>		
	$\phi$	$\psi$	$\omega$
<b>Amber</b>	74.3 (15.0) 100	-178.3 (19.3) 53/ 87.8 (17.0) 46/ -100.4 (18.5) 1	51.4 (10.8) 84/ -177.2 (16.4) 12/ -82.1 (20.5) 3
<b>GMX</b>	72.8 (17.1) 100	-179.5 (20.5) 58/ 85.5 (19.0) 39/ -86.8 (18.4) 4	53.2 (10.8) 78/ -169.4 (17.5) 16/ -77.0 (201.2) 6
	<i>GlcNAc(7)-α(1-2)-Man(5)</i>		
	$\phi$	$\psi$	
<b>Amber</b>	-80.8 (15.1) 100	163.2 (16.3) 91/ 106.3 (12.8) 9	
<b>GMX</b>	-76.9 (14.0) 100	158.6 (13.8) 89/ 109.5 (11.4) 11	

	<i>Man(4)-<math>\alpha</math>(1-3)-Man(3)</i>	
	$\phi$	$\psi$
<b>Amber</b>	73.0 (12.3) <b>100</b>	146.1 (12.9) <b>60</b> / 101.9 (14.8) <b>40</b>
<b>GMX</b>	72.9 (11.9) <b>100</b>	145.6 (12.4) <b>59</b> / 102.7 (14.0) <b>41</b>
	<i>GlcNAc(6)-<math>\alpha</math>(1-2)-Man(4)</i>	
	$\phi$	$\psi$
<b>Amber</b>	-81.0 (17.3) <b>100</b>	161.5 (13.6) <b>84</b> / 104.6 (12.3) <b>16</b>
<b>GMX</b>	-77.6 (17.3) <b>100</b>	159.4 (14.4) <b>82</b> / 106.4 (12.8) <b>18</b>

**Table S.3 GlcNAc(2)- $\beta$ (1-4)-GlcNAc(1) linkage in non-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>		<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>A</b>	-78.6	10.2	100		-130.8	15.8	100			0
<b>C</b>	-77.8	10.0	100		-129.5	17.8	100			0
<b>E</b>	-78.3	10.7	100		-130.3	15.5	100			0
<b>G</b>	-78.2	10.9	100		-130.2	15.4	100			0
<b>I</b>	-78.2	10.8	100		-130.2	15.3	100			0
<b>K</b>	-79.0	13.0	100		-130.2	15.7	97	70.1	12.1	3
<b>M</b>	-79.3	12.0	100		-131.5	15.7	100			0
<b>O</b>	-80.3	11.0	100		-133.3	15.0	94	68.3	12.4	4
<b>Q</b>	-78.4	11.3	100		-130.5	15.8	99	70.4	13.6	1
<b>S</b>	-79.3	11.4	100		-131.1	15.4	97	67.1	11.4	3
<b>AVERAGES</b>	<b>-78.7</b>	<b>11.1</b>	<b>100</b>		<b>-130.8</b>	<b>15.7</b>	<b>99</b>	<b>69.0</b>	<b>12.4</b>	<b>1</b>

**Table S.4 GlcNAc(2)- $\beta$ (1-4)-GlcNAc(1) linkage in core-fucosylated glycans. Note: average values do not include sugar D as the dynamic of this linkage is affected by the equilibrium of the GlcNAc(1) ring pucker between  ${}^4C_1$  to  ${}^1C_4$  conformations.**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>		<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (3)</b>	<b>stdev</b>	<b>pop</b>
<b>B</b>	-77.3	9.7	100		-127.2	15.0	100			0			0
<b>D*</b>	-77.9	11.4	100		-129.5	13.8	85	-82.8	11.5	12	69.6	11.8	3
<b>F</b>	-75.6	19.8	100		-126.9	16.2	100			0			0
<b>H</b>	-77.5	10.0	100		-126.6	14.9	100			0			0
<b>J</b>	-77.8	10.6	100		-126.3	15.5	100			0			0
<b>L</b>	-78.7	9.9	100		-128.6	14.7	100			0			0
<b>N</b>	-78.6	10.0	100		-128.0	14.7	100			0			0
<b>P</b>	-78.0	10.0	100		-127.7	14.3	98	70.1	11.9	2			0
<b>R</b>	-77.9	10.7	100		-126.9	14.6	100			0			0
<b>T</b>	-79.0	9.7	100		-129.9	14.0	100			0			0
<b>AVE.</b>	<b>-77.8</b>	<b>11.2</b>	<b>100</b>		<b>-127.8</b>	<b>14.8</b>	<b>100</b>	<b>70.1</b>	<b>11.9</b>	<b>0</b>			<b>0</b>

**Table S.5 Fuc(12)- $\alpha$ (1-6)-GlcNAc(1) linkage. Note: average values do not include sugar D as the dynamic of this linkage is affected by the equilibrium of the GlcNAc(1) ring pucker between  ${}^4C_1$  to  ${}^1C_4$  conformations.**

	phi	stdev	pop	phi(2)	stdev	pop		psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>B</b>	-74.2	12.3	97	-142.8	11.3	3	B	-181.1	18.7	78	109.3	22.5	22			0
<b>D*</b>	-75.9	15.7	100			0	D	-183.0	21.0	79	93.1	18.6	17	-87.8	15.2	4
<b>F</b>	-75	12.8	98	-142	10.3	2	F	-183.1	21.7	75	92.7	18.9	20	-90.2	17	5
<b>H</b>	-78.6	15.8	100			0	H	-185.3	19.4	78	92.4	19.1	20	-99.2	16.3	2
<b>J</b>	-78.5	17.6	100			0	J	-185.5	19.4	79	92	19	19	-96.3	16.6	3
<b>L</b>	-78	16.7	100			0	L	-186.8	19.3	78	93.2	19.1	18	-97.5	16	4
<b>N</b>	-77.8	16.5	100			0	N	-187.5	19.1	81	93.5	19.3	16	-98.6	15.8	2
<b>P</b>	-78.7	17.6	100			0	P	-186.7	19.3	79	92.4	18.6	19	-97.6	17.2	2
<b>R</b>	-78.3	17.5	100			0	R	-184.2	19.6	79	92.5	19	19	-100.8	15.4	2
<b>T</b>	-78.3	16.9	100			0	T	-186.9	19.2	79	92.5	18.9	19	-98	17.5	2
<b>AV.</b>	<b>-77.3</b>	<b>15.9</b>	<b>100</b>	<b>-142.4</b>	<b>10.8</b>	<b>1</b>		<b>-185.2</b>	<b>19.5</b>	<b>78.4</b>	<b>94.5</b>	<b>19.4</b>	<b>19.1</b>	<b>-97.3</b>	<b>16.5</b>	<b>2.4</b>
	omega(1)		stdev	pop	omega(2)		stdev	pop	omega(3)		stdev	pop				
<b>B</b>	66.1		13.5	100				0				0				
<b>D*</b>	46.7		12.0	77	-64.9		12.9	13	-172.2		14.0	11				
<b>F</b>	46.5		11.9	85				0	-166.5		23.8	15				
<b>H</b>	45		11.8	93	-58.4		30.4	2	-169.4		19.1	6				
<b>J</b>	44.9		11.7	91	-67.3		28.3	2	-171.2		15.1	7				
<b>L</b>	43.8		12.6	85	-50.5		25.5	2	-172		14.5	13				
<b>N</b>	43.4		12.9	90	-43.5		29.8	2	-170.9		15	8				
<b>P</b>	44.4		12.3	93	-51.3		33.7	1	-156.9		14.7	6				
<b>R</b>	45.2		11.8	92	-63.5		30.7	1	-170		15.3	7				
<b>T</b>	44.5		12.3	96	-48.9		35.3	1	-170.8		14.8	4				
<b>AV.</b>	<b>47.1</b>		<b>12.3</b>	<b>91.7</b>	<b>-54.8</b>		<b>30.5</b>	<b>1.2</b>	<b>-168.5</b>		<b>16.5</b>	<b>7.3</b>				

**Table S.6 Man(3)- $\beta$ (1-4)-GlcNAc(2) linkage in non-fucosylated glycans**

	phi	stdev	pop	phi(2)	stdev	pop	phi(3)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop
<b>C</b>	-73.5	12.1	95	-156.4	25.5	5			0	-124.8	17.3	82	70.7	13.7	18
<b>E</b>	-76.6	15.3	96	-165.9	16.2	4			0	-125.8	16.2	93	73.1	12.1	7
<b>G</b>	-79	22.4	100			0			0	-124.6	15.5	100			0
<b>I</b>	-83.7	27	100			0			0	-126	15.93	100			0
<b>K</b>	-82.5	25.6	100			0			0	-126	15.2	100			0
<b>M</b>	-82.5	25.6	100			0			0	-126.7	15.4	100			0
<b>O</b>	-82.3	24.4	93			0	56.7	9.1	7	-124.6	14.5	96	71	12.1	4
<b>Q</b>	-77	17.9	100			0			0	-124.6	14.5	96	52.4	9.4	2
<b>S</b>	-79.13	15.07	92	-151.54	12.99	4			0	-129.28	14.3	100			0
<b>AV.</b>	<b>-79.6</b>	<b>20.6</b>	<b>97</b>	<b>-157.9</b>	<b>18.2</b>	<b>1</b>	<b>56.7</b>	<b>9.1</b>	<b>1</b>	<b>-125.4</b>	<b>15.4</b>	<b>96</b>	<b>66.8</b>	<b>11.8</b>	<b>3</b>

**Table S.7 Man(3)- $\beta$ (1-4)-GlcNAc(2) linkage in core-fucosylated glycans**

	phi	stdev	pop	phi(2)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop
<b>D</b>	-75.2	11.7	96	-163.1	26.1	4	-124.9	16.3	96	69.9	14.8	4
<b>F</b>	-76	12.6	95	-166.6	15.4	5	-125.5	15.4	91	70.6	11.3	9
<b>H</b>	-77.6	18.8	100			0	-124.4	15	100			0
<b>J</b>	-79.3	21.8	100			0	-126	15.9	98	71.8	12	2
<b>L</b>	-84.6	27.3	100			0	-125.7	15.1	94	68.3	12.5	6
<b>N</b>	-83	24.8	100			0	-125.3	14.4	93	72	12.5	7
<b>P</b>	-71.5	13.8	87	-152	14.5	13	-126.2	14.5	100			0
<b>R</b>	-78.1	19.6	100			0	-125.6	15.8	96	71.7	12.3	4
<b>T</b>	-77.2	13.2	92	-152.2	16.6	8	-126.5	14.8	99	78.5	15.4	1
<b>AVERAGES</b>	<b>-78.1</b>	<b>18.2</b>	<b>97</b>	<b>-158.5</b>	<b>18.2</b>	<b>3</b>	<b>-125.5</b>	<b>15.2</b>	<b>96</b>	<b>71.8</b>	<b>13.0</b>	<b>4</b>

**Table S.8 Man(4)- $\alpha$ (1-3)-Man(3) linkage in non-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>E</b>	72.7	12.4	100	143.8	13.6	72	102.4	12.2	28.0
<b>G</b>	72.7	11.8	100	145.7	12.6	59	101.9	14.5	41.0
<b>I</b>	73.0	12.1	100	146.0	13.2	58	102.4	14.3	42.0
<b>K</b>	72.5	12.3	100	146.4	17.2	59	103.1	14.6	41.0
<b>M</b>	73.1	12.5	100	146.1	12.7	61	102.1	14.7	39.0
<b>O</b>	73.1	12.3	100	150.0	15.3	65	102.4	14.4	35.0
<b>Q</b>	72.7	12.2	100	145.8	12.7	59	101.7	14.7	40.0
<b>S</b>	72.6	11.3	97	147.9	14.1	58	102.2	14.1	42.0
<b>AVERAGES</b>	<b>72.8</b>	<b>12.1</b>	<b>100</b>	<b>146.5</b>	<b>13.9</b>	<b>61</b>	<b>102.3</b>	<b>14.2</b>	<b>39</b>



**Table S.9 Man(4)- $\alpha$ (1-3)-Man(3) linkage in core-fucosylated glycans**

	<b>phi</b>			<b>phi(2)</b>			<b>phi(3)</b>					
	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>phi(2)</b>	<b>stdev</b>	<b>pop</b>	<b>phi(3)</b>	<b>stdev</b>	<b>pop</b>			
<b>F</b>	72.5	12.1	100			0			0			
<b>H</b>	72.7	11.9	100			0			0			
<b>J</b>	73.1	12.9	100			0			0			
<b>L</b>	72.6	11.9	100			0			0			
<b>N</b>	73.7	13.8	100			0			0			
<b>P</b>	72.4	10.1	93	124.9	21.7	5	164.6	17.9	2			
<b>R</b>	73.7	14.2	100			0			0			
<b>T</b>	72.7	10.3	98	122.4	22.2	2			0			
<b>AVERAGES</b>	<b>72.9</b>	<b>12.2</b>	<b>99</b>	<b>123.7</b>	<b>22.0</b>	<b>1</b>	<b>164.6</b>	<b>17.9</b>	<b>0</b>			
	<b>psi (1)</b>			<b>psi (2)</b>			<b>psi (3)</b>			<b>psi (4)</b>		
	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (3)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (4)</b>	<b>stdev</b>	<b>pop</b>
<b>F</b>	145.5	12.3	65	105.1	13.6	34			0			0
<b>H</b>	145.8	12.6	60	102.3	14.5	40			0			0
<b>J</b>	150.2	21.3	60	102.2	14.6	40			0			0
<b>L</b>	146.4	15.1	60	102.7	14.5	40			0			0
<b>N</b>	151	19.8	56	99	14.5	44			0			0
<b>P</b>	147.4	13.8	55	102.6	15	34	182	17.3	4	-47.4	16.4	2
<b>R</b>	146.1	13	59	101.7	14.9	41			0			0
<b>T</b>	147.4	13.8	64	102.3	14.5	36			0			0
<b>AVERAGES</b>	<b>147.5</b>	<b>15.2</b>	<b>60</b>	<b>102.2</b>	<b>14.5</b>	<b>39</b>	<b>182.0</b>	<b>17.3</b>	<b>1</b>	<b>-47.4</b>	<b>16.4</b>	<b>0</b>

**Table S.10 GlcNAc(6)- $\beta$ (1-2)-Man(4) linkage in non-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>G</b>	-80.7	16.4	100	161.4	13.9	84	105.2	12.9	16
<b>I</b>	-80.9	15.8	100	161.7	16.1	84	105.2	12.4	16
<b>K</b>	-80.4	16	100	162.1	16.3	84	105.1	12.3	16
<b>M</b>	-80.8	16.4	100	162.3	17.1	85	105.4	12.1	15
<b>O</b>	-79.7	15.5	100	161.7	15.7	84	105.6	12.2	16
<b>Q</b>	-81.1	17.6	100	162.3	17.2	83	105.7	12.1	18
<b>S</b>	-83.43	16.21	94	161.09	14.81	86	105.92	12.18	14
<b>AVERAGES</b>	<b>-81.0</b>	<b>16.3</b>	<b>99</b>	<b>161.8</b>	<b>15.9</b>	<b>84</b>	<b>105.4</b>	<b>12.3</b>	<b>16</b>

**Table S.11 GlcNAc(6)- $\beta$ (1-2)-Man(4) linkage in core-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>H</b>	-80.7	16.4	100	161.1	13.9	84	105.2	12.6	16
<b>J</b>	-80.9	16.3	100	162	15.9	84	105.1	12.5	16
<b>L</b>	-80.3	15.7	100	161.7	15.5	84	105.3	12.5	16
<b>N</b>	-81.2	17.1	100	162	16.2	84	105.3	12.4	16
<b>P</b>	-81	15.9	100	162.2	15.9	85	105.9	12.3	15
<b>R</b>	-80.3	17.1	100	162.1	15.6	85	105.7	12.2	15
<b>T</b>	-81.7	17	100	162.4	16.3	85	105.5	12.1	15
<b>AVERAGES</b>	<b>-80.9</b>	<b>16.5</b>	<b>100</b>	<b>161.9</b>	<b>15.6</b>	<b>84</b>	<b>105.4</b>	<b>12.4</b>	<b>16</b>

**Table S.12 Gal(8)- $\beta$ (1-4)-GlcNAc(6) linkage in non-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>I</b>	-74.1	16.9	100	-120.3	15.7	97	71.8	15.3	3
<b>M</b>	-75	16.9	100	-120.3	15.9	100			0
<b>O</b>	-76	18	100	-123.1	16.7	88	76.6	19.4	12
<b>Q</b>	-76.8	18.3	100	-123.7	17.2	98			2
<b>S</b>	-76.4	16.18	100	-122.98	16.3	91	75	18	9
<b>AVERAGES</b>	<b>-75.7</b>	<b>17.3</b>	<b>100</b>	<b>-122.1</b>	<b>16.4</b>	<b>95</b>	<b>74.5</b>	<b>17.6</b>	<b>5</b>

**Table S.13 Gal(8)- $\beta$ (1-4)-GlcNAc(6) linkage in core-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>J</b>	-74.9	16	100	-120.3	15.3	98	73.9	20.4	2
<b>N</b>	-74.8	17	100	-120.5	16.2	97	71.7	17.8	3
<b>P</b>	-76.7	16.7	98	-123.9	16.6	96	79	20.4	4
<b>R</b>	-75.2	18.4	100	-122.9	16.9	89	84.4	19.5	11
<b>T</b>	-74.1	11.6	95	-124	17.3	94	84.5	27.3	6
<b>AVERAGES</b>	<b>-75.1</b>	<b>15.9</b>	<b>99</b>	<b>-122.3</b>	<b>16.5</b>	<b>95</b>	<b>78.7</b>	<b>21.1</b>	<b>5</b>

**Table S.14 Sia(10)- $\alpha$ (2-6)-Gal(8) linkage in non-fucosylated glycans**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>phi(2)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (3)</b>	<b>stdev</b>	<b>pop</b>
<b>O</b>	65.4	11.2	88	-50.3	15.4	12	-182.2	24.3	84	-99.4	16.4	11	105.1	16.6	5
<b>Q</b>	65.2	11.2	88	-50.7	15.9	12	-172.2	21.3	74	-101.1	15.2	23	104.8	15.7	3
<b>S</b>	65.0	11.2	90	-49.2	13.5	10	-182.2	24.5	85	-99.4	16.7	13	107.0	15.9	3
<b>AVE.</b>	<b>65.2</b>	<b>11.2</b>	<b>89</b>	<b>-50.1</b>	<b>14.9</b>	<b>11</b>	<b>-178.9</b>	<b>23.4</b>	<b>81</b>	<b>-100.0</b>	<b>16.1</b>	<b>16</b>	<b>105.6</b>	<b>16.1</b>	<b>4</b>
	<b>omega</b>		<b>stdev</b>	<b>pop</b>	<b>omega(2)</b>		<b>stdev</b>	<b>pop</b>	<b>omega(3)</b>		<b>stdev</b>	<b>pop</b>			
<b>O</b>	-61.5		16.3	51	-166.7		14.2	32	59.4		14.2	14			
<b>Q</b>	-62.6		15.3	65	-165.0		14.8	33	62.1		12.4	2			
<b>S</b>	-63.8		15.7	67	-163.6		14.8	28	58.5		13.8	5			
<b>AVE.</b>	<b>-62.6</b>		<b>15.8</b>	<b>61</b>	<b>-165.1</b>		<b>14.6</b>	<b>31</b>	<b>60.0</b>		<b>13.5</b>	<b>7</b>			

**Table S.15 Sia(10)- $\alpha$ (2-6)-Gal(8) linkage in core-fucosylated glycans**

	phi	stdev	pop	phi(2)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>P</b>	65.2	10.9	91	-50.1	14.1	9	-182.4	24.6	85	-99.7	16.5	11	107.3	16.8	3
<b>R</b>	65.1	11.1	90	-50.4	16.5	11	-182.5	24.4	85	-98.2	16.8	11	106	16.7	4
<b>T</b>	65.1	11.1	90	-50.7	15.1	10	-182.9	24.5	85	-100	16.7	12	100.7	20.1	3
<b>AVERAGES</b>	<b>65.1</b>	<b>11.0</b>	<b>90</b>	<b>-50.4</b>	<b>15.2</b>	<b>10</b>	<b>-182.6</b>	<b>24.5</b>	<b>85</b>	<b>-99.3</b>	<b>16.7</b>	<b>11</b>	<b>104.7</b>	<b>17.9</b>	<b>3</b>
	omega		stdev	pop		omega(2)		stdev	pop		omega(3)		stdev	pop	
<b>P</b>	-62.7		15.5	65		-164.3		14.8	30		57.8		14.3	5	
<b>R</b>	-65.9		15.7	63		-167.3		15.5	33		57		15.2	4	
<b>T</b>	-62.5		15.7	64		-164.9		14.7	30		59.9		13.5	6	
<b>AVERAGES</b>	<b>-63.7</b>		<b>15.6</b>	<b>64</b>		<b>-165.5</b>		<b>15.0</b>	<b>31</b>		<b>58.2</b>		<b>14.3</b>	<b>5</b>	

**Table S.16 Man(5)- $\alpha$ (1-6)-Man(3) linkage in non-fucosylated glycans.**

	phi	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>E</b>	72.2	17.2	100	81.5	17.6	13	-175.3	17.3	84	-87.5	21.2	3
<b>G</b>	72.8	15.5	100	80.9	18.9	44	-177.5	18.8	54	-97.3	17.4	2
<b>I</b>	74.1	17.3	100	82.8	18.5	40	-178.8	18.5	58	-98.6	17.2	2
<b>Q</b>	73.7	16.1	100	83.2	18.4	50	-179.2	19.6	49	-99.1	15.1	2
<b>AVE.</b>	<b>73.2</b>	<b>16.5</b>	<b>100</b>	<b>82.1</b>	<b>18.4</b>	<b>45</b>	<b>-177.7</b>	<b>18.6</b>	<b>54</b>	<b>-95.6</b>	<b>17.7</b>	<b>2</b>
<b>K</b>	74.9	14.2	100	83.2	17.6	68	-178.9	20.4	30	-94.1	18.2	2
<b>M</b>	75.3	14.9	100	83.8	17.8	65	-181.2	20.8	34	-105	13.9	1
<b>O</b>	74.9	15.4	100	80.7	16.6	77	-183	21.8	22	-99.2	13.8	1
<b>S</b>	75.34	15.11	100	83.76	17.36	85	-185.84	24.73	17	-108.97	7.73	0
<b>AVE.</b>	<b>75.1</b>	<b>14.9</b>	<b>100</b>	<b>82.9</b>	<b>17.3</b>	<b>74</b>	<b>-182.2</b>	<b>21.9</b>	<b>26</b>	<b>-101.8</b>	<b>13.4</b>	<b>1</b>
	omega(1)	stdev	pop	omega (2)	stdev	pop	omega(3)	stdev	pop			
<b>E</b>	54.1	10.8	75	-171	16.7	20	-73.4	16.7	5			
<b>G</b>	51.3	11	81	-174.4	16.6	14	-78	19.4	4			
<b>I</b>	51.1	11.3	81	-174.8	16.9	15	-77.2	18	5			
<b>Q</b>	50.6	11	84	-168.9	17.8	11	-74	17.3	5			
<b>AVE.</b>	<b>51.8</b>	<b>11.0</b>	<b>80</b>	<b>-172.3</b>	<b>17.0</b>	<b>15</b>	<b>-75.7</b>	<b>17.9</b>	<b>5</b>			
<b>K</b>	49.8	10.4	85	-169.5	18	11	-78	18.5	4			
<b>M</b>	50.7	10.4	84	-172.4	17	11	-75	15.9	4			
<b>O</b>	50.1	10.4	74	-185.7	12.9	24	-82.4	17	2			
<b>S</b>	48.4	10.3	99	-173.5	18.5	1	-70.5	17.8	0			
<b>AVE.</b>	<b>50.2</b>	<b>10.4</b>	<b>81</b>	<b>-175.9</b>	<b>16.0</b>	<b>15</b>	<b>-78.5</b>	<b>17.1</b>	<b>3</b>			

**Table S.17 Man(5)- $\alpha$ (1-6)-Man(3) linkage in core-fucosylated glycans.**

	phi	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>F</b>	72	16.8	100	83.1	18.2	13	-175.3	16.7	85	-96.3	20.9	2
<b>H</b>	74.3	15.2	100	84.4	17.2	53	-178.8	19.6	46	-75.8	17.4	1
<b>J</b>	74.3	15.8	100	85.8	16.8	49	-179.9	19.3	50	-100.1	17.1	2
<b>R</b>	74.2	15.5	100	87.8	16.4	45	-180.1	19.1	53	-99.8	14.2	2
<b>AVERAGES</b>	<b>73.7</b>	<b>15.8</b>	<b>100</b>	<b>85.3</b>	<b>17.2</b>	<b>49</b>	<b>-178.5</b>	<b>18.7</b>	<b>50</b>	<b>-93.0</b>	<b>17.4</b>	<b>2</b>
<b>L</b>	75.8	14.9	100	84.4	15.5	76	-184.4	21.4	24			0
<b>N</b>	76.3	15.7	100	84.4	16.6	76	-185.2	22.4	23	-98.3	14.8	1
<b>P</b>	76.2	14.4	100	87	15.5	77	-185.8	22.4	22	-87.6	18.3	0
<b>T</b>	76.7	14.9	100	85.2	15.5	76	-184.9	22.2	22	-112.1	7.8	0
<b>AVERAGES</b>	<b>76.3</b>	<b>15.0</b>	<b>100</b>	<b>85.3</b>	<b>15.8</b>	<b>76</b>	<b>-185.1</b>	<b>22.1</b>	<b>23</b>	<b>-99.3</b>	<b>13.6</b>	<b>0</b>
	omega(1)	stdev	pop	omega (2)	stdev	pop	psi (4)	stdev	pop			
<b>F</b>	54.1	11	73	-174	15.8	23	-77.7	19.3	4			
<b>H</b>	51.2	10.7	84	-176.1	16.5	13	-76.6	19.3	3			
<b>J</b>	51.3	10.6	84	-172.9	17.5	12	-79.3	19.7	4			
<b>R</b>	51	10.6	83	-171.2	16.3	12	-76.2	19.9	5			
<b>AVERAGES</b>	<b>51.9</b>	<b>10.7</b>	<b>81</b>	<b>-173.6</b>	<b>16.5</b>	<b>15</b>	<b>-77.5</b>	<b>19.6</b>	<b>4</b>			
<b>L</b>	50.6	9.8	83	-181.2	14.5	16	-74.7	19.5	1			
<b>N</b>	50.5	9.9	85	-175.2	15.2	13	-77	16.4	2			
<b>P</b>	50.5	9.5	94	-170.8	16.3	5	-82	18.8	1			
<b>T</b>	50.2	9.9	90	-163.9	13.6	8	-86.9	17.5	3			
<b>AVERAGES</b>	<b>50.5</b>	<b>9.8</b>	<b>88</b>	<b>-172.8</b>	<b>14.9</b>	<b>11</b>	<b>-80.2</b>	<b>18.1</b>	<b>2</b>			

**Table S.18 GlcNAc(7)- $\beta$ (1-2)-Man(5) linkage in non-fucosylated glycans.**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>G</b>	-80.8	16.4	100	162	14.6	89	106	12.6	11
<b>I</b>	-78.9	22.7	100	160.2	14.6	85	100	15.4	15
<b>K</b>	-82.2	15.3	100	163.4	13.7	93	108.2	11.6	7
<b>M</b>	-75.9	31	100	162.8	13	92	107.6	11.9	8
<b>O</b>	-81.4	14.4	100	160.8	14.2	84	109.2	10.6	16
<b>Q</b>	-78.1	22.5	100	162.2	13	91	105.7	12.6	10
<b>S</b>	-81.39	14.58	100	163.33	11.79	97	109.03	11.76	3
<b>AVERAGES</b>	<b>-79.8</b>	<b>19.6</b>	<b>100</b>	<b>162.1</b>	<b>13.6</b>	<b>90</b>	<b>106.5</b>	<b>12.4</b>	<b>10</b>

**Table S.19 GlcNAc(7)- $\beta$ (1-2)-Man(5) linkage in core-fucosylated glycans.**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>H</b>	-81.2	15.7	100	162.7	14.2	90	106.4	12.2	10
<b>J</b>	-81.1	15.6	100	161.8	13.3	89	104.9	13	11
<b>L</b>	-83.4	14.2	100	160.7	13.4	89	109.5	11.3	11
<b>N</b>	-83.2	14.4	100	161.5	12.7	92	108.6	11.3	8
<b>P</b>	-76.9	14.6	100	161.9	11.9	96	108.8	11.4	4
<b>R</b>	-80.9	15.8	100	162.6	13.8	91	106	12.7	9
<b>T</b>	-83.2	13.9	100	162.8	12.7	94	110.3	10.4	7
<b>AVERAGES</b>	<b>-81.4</b>	<b>14.9</b>	<b>100</b>	<b>162.0</b>	<b>13.1</b>	<b>92</b>	<b>107.8</b>	<b>11.8</b>	<b>9</b>

**Table S.20 Gal(9)- $\beta$ (1-4)-GlcNAc(7) linkage in non-fucosylated glycans.**

	<b>phi</b>	<b>stdev</b>	<b>pop</b>	<b>psi (1)</b>	<b>stdev</b>	<b>pop</b>	<b>psi (2)</b>	<b>stdev</b>	<b>pop</b>
<b>K</b>	-76.6	16	100	-123.3	15.6	88	120.3	7.2	12
<b>M</b>	-76	13.7	100	-126	13.8	100			0
<b>O</b>	-74.5	15.6	100	-124.6	15.7	100			0
<b>S</b>	-77.7	15	95	-128.79	16.35	100			0
<b>AVERAGES</b>	<b>-76.2</b>	<b>15.1</b>	<b>99</b>	<b>-125.7</b>	<b>15.4</b>	<b>97</b>	<b>120.3</b>	<b>7.2</b>	<b>3</b>

**Table S.21 Gal(9)- $\beta$ (1-4)- GlcNAc(7) linkage in core-fucosylated glycans.**

	phi	stdev	pop	phi(2)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop
<b>L</b>	-75.6	15	100			0	-122.9	15.6	99	152.1	10.6	1
<b>N</b>	-75	14.8	100			0	-122.1	15.5	99	154.5	7.6	1
<b>P</b>	-74.7	14.3	100			0	-122.3	15	100			0
<b>T</b>	-73.9	10.8	100	-123.5	19.3	5	-123.2	16.8	96			0
<b>AVERAGES</b>	<b>-74.8</b>	<b>13.7</b>	<b>100</b>	<b>-123.5</b>	<b>19.3</b>	<b>1</b>	<b>-122.6</b>	<b>15.7</b>	<b>99</b>	<b>153.3</b>	<b>9.1</b>	<b>1</b>

**Table S.22 Sia(11)- $\alpha$ (2-6)- Gal(9) linkage in non-fucosylated glycans.**

	phi	stdev	pop	phi(2)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>S</b>	64.54	11.23	90	-50.67	13.68	10	-180.96	25.21	88	-98.61	17.87	9	106.45	15.97	3
<b>AVE.</b>	<b>64.5</b>	<b>11.2</b>	<b>90</b>	<b>-50.7</b>	<b>13.7</b>	<b>10</b>	<b>-181.0</b>	<b>25.2</b>	<b>88</b>	<b>-98.6</b>	<b>17.9</b>	<b>9</b>	<b>106.5</b>	<b>16.0</b>	<b>3</b>
	<b>omega</b>	<b>stdev</b>	<b>pop</b>	<b>omega(2)</b>	<b>stdev</b>	<b>pop</b>	<b>omega(3)</b>	<b>stdev</b>	<b>pop</b>						
<b>S</b>	-62.56	15.61	60	-166.58	14.35	33	58.76						13.61		7
<b>AVE.</b>	<b>-62.6</b>	<b>15.6</b>	<b>60</b>	<b>-166.6</b>	<b>14.4</b>	<b>33</b>	<b>58.8</b>						<b>13.6</b>		<b>7</b>

**Table S.23 Sia(11)- $\alpha$ (2-6)- Gal(9) linkage in core-fucosylated glycans.**

	phi	stdev	pop	phi(2)	stdev	pop	psi (1)	stdev	pop	psi (2)	stdev	pop	psi (3)	stdev	pop
<b>T</b>	64.9	11.8	88	-52.3	19.1	13	-182	25.2	88	-101.3	16.6	9	90.3	16.5	4
<b>AVE</b>	<b>64.9</b>	<b>11.8</b>	<b>88</b>	<b>-52.3</b>	<b>19.1</b>	<b>13</b>	<b>-182.0</b>	<b>25.2</b>	<b>88</b>	<b>-101.3</b>	<b>16.6</b>	<b>9</b>	<b>90.3</b>	<b>16.5</b>	<b>4</b>
	<b>omega</b>	<b>stdev</b>	<b>pop</b>	<b>omega(2)</b>	<b>stdev</b>	<b>pop</b>	<b>omega(3)</b>	<b>stdev</b>	<b>pop</b>						
<b>T</b>	-64.9	16.2	56	-163.5	15.2	34	60.3	14.4	8						
<b>AVE</b>	<b>-64.9</b>	<b>16.2</b>	<b>56</b>	<b>-163.5</b>	<b>15.2</b>	<b>34</b>	<b>60.3</b>	<b>14.4</b>	<b>8</b>						

## References

- Fadda E, Woods RJ. 2010. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today*, 15:596-609.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph*, 14:33-38, 27-38.
- Jorgensen WL, Jenson C. 1998. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry*, 19:1179-1186.
- Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem*, 29:622-655.
- Salomon-Ferrer R, Case DA, Walker RC. 2013. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.*, 3:198-210.
- Sousa da Silva AW, Vranken WF. 2012. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res Notes*, 5:367.



## Appendix II

### 1. Computational Methods

**System preparation.** All N-glycan starting structures for the MD simulations were generated with the GLYCAM Carbohydrate Builder (<http://www.glycam.org>). For each sequence we selected the complete set of rotamers obtained by variation of the 1-6 torsion angles, namely *gg*, *gt* and *tg* conformations for each 1-6 torsion. The topology file corresponding to each structure was obtained using *tLeap*<sup>1</sup>, with parameters from the GLYCAM06-j1<sup>2</sup> for the carbohydrate atoms and with TIP3P for water molecules<sup>3</sup>. Each N-glycan was placed in the centre of a cubic simulation box of 16 Å sides with no counterions to be consistent with the simulations run in earlier work<sup>4</sup>. Long range electrostatic were treated by Particle Mesh Ewald (PME) with cut-off set at 11 Å and a B-spline interpolation for mapping particles to and from the mesh of order of 4. Van der Waals (vdW) interactions were cut-off at 11 Å. The MD trajectories were generated by Langevin dynamics with collision frequency of 1.0 ps<sup>-1</sup>. Pressure was kept constant by isotropic pressure scaling with a pressure relaxation time of 2.0 ps. All calculations were run with the AMBER18 software package<sup>1</sup> on NVIDIA Tesla V100 16GB PCIe (Volta architecture) GPUs installed on the HPC infrastructure *kay* at the Irish Centre for High-End Computing (ICHEC).

**Simulation protocol.** The energy of the hydrated systems was initially minimized through 500,000 cycles of steepest descent, with all heavy atoms restrained with a harmonic potential with a force constant of 5 kcal mol<sup>-1</sup>Å<sup>-2</sup>. After minimization, the system was heated in two stages. During the first stage the temperature was raised from 0 to 100 K over 500 ps at constant volume and in the second stage from 100 K to 300 K over 500 ps at constant pressure. Through the heating process all heavy atoms were kept restrained. After heating phase all restraints were removed and the system was allowed to equilibrate for 5 ns at 300 K and at 1 atm of pressure. Separate production steps of 500 ns each were run for each rotamer (starting system) and convergence was assessed based on conformational and clustering analysis. Simulations were extended, if the sampling was not deemed as fully converged.

**Data analysis.** All trajectories were processed using *cpptraj*<sup>1</sup> and visually analysed with the Visual Molecular Dynamics (VMD) software package<sup>5</sup>. Backbone Root Mean Square

Deviation (RMSD) and torsion angles values were measured using VMD. A density-based clustering method was used to calculate the populations of occupied conformations for each torsion angle in a trajectory and heat maps for each dihedral were generated with a kernel density estimate (KDE) function. Statistical and clustering analysis was done with the R package and data were plotted with RStudio ([www.rstudio.com](http://www.rstudio.com)).

## 2. Supplementary Tables and Figures

**Table S.1.** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles (°) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\alpha$ (1-3) core fucosylated ngf glycan.

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.1 (8.9)	141.1 (6.3)	100
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.1 (8.3)	-107.1 (7.6)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.3 (12.3)	-122.1 (14.9)	90.0
Cluster 2	-166.4 (16.7)	-145.6 (9.8)	8.3
Cluster 3	-64.0 (10.0)	75.3 (9.3)	1.7
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	69.2 (10.2)	-176.6 (17.5)	65.7
Cluster 2	74.2 (9.4)	89.2 (10.8)	25.2
Cluster 3	60.4 (6.4)	60.0 (5.8)	9.1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.8 (14.3)	160.7 (20.6)	96.2
Cluster 2	69.6 (9.3)	154.8 (10.3)	2.5
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.9 (8.5)	140.3 (13.5)	66.8
Cluster 2	70.6 (8.7)	99.2 (9.5)	33.2
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.8 (13.7)	161.5 (13.5)	88.8
Cluster 2	-78.5 (8.0)	109.2 (7.2)	9.7

**Table S.2.** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles (°) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\alpha$ (1-3) core fucosylated gf glycan.

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
--	------------	------------	---------------

Cluster 1	-72.5 (10.7)	125.2 (17.4)	100
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.9 (8.6)	-105.7 (11.7)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-74.0 (12.3)	-123.9 (15.7)	88.5
Cluster 2	-165.7 (14.7)	-145.7 (8.8)	11.5
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	69.3 (10.6)	-177.6 (19.4)	72.5
Cluster 2	71.5 (6.2)	79.2 (7.1)	16.2
Cluster 3	60.7 (5.7)	60.4 (5.1)	1.1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.7 (13.8)	162.5 (12.4)	88.4
Cluster 2	-76.0 (6.2)	113.5 (6.6)	5.9
Cluster 3	-147.6 (9.5)	97.9 (9.4)	1.9
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.9 (11.2)	124.0 (18.2)	96.6
Cluster 2	-82.1 (12.2)	-64.0 (8.4)	3.0
Cluster 3	-150.0 (11.3)	-101.0 (7.3)	1.4
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	71.7 (8.9)	141.5 (14.5)	66.9
Cluster 2	70.4 (8.6)	99.12 (9.5)	33.1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.9 (14.2)	160.9 (14.5)	83.3
Cluster 2	67.2 (9.2)	152.4 (10.4)	8.6
Cluster 3	-77.7 (7.7)	152.4 (6.2)	8.1
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.4 (9.8)	125.0 (15.8)	100

**Table S.3. Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 3  $\mu$ s MD sampling of the  $\alpha$  (1-2) core xylylated ngx glycan.**

<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.1 (10.2)	-129.6 (15.67)	96.3
Cluster 2	-81.8 (9.8)	63.8 (8.11)	3.7
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.5 (12.6)	-123.5 (14.6)	94.4

Cluster 2	-64.2 (6.8)	74.2 (9.7)	5.6
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-80.2 (12.0)	133.2 (17.0)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	69.7 (10.3)	-174.4 (16.8)	77.2
Cluster 2	73.5 (10.1)	106.1 (10.6)	22.8
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-80.9 (15.5)	161.7 (12.0)	90.1
Cluster 2	-73.1 (7.9)	114.2 (7.8)	9.3
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	78.4 (7.3)	114.8 (14.8)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.4 (13.3)	161.5 (12.2)	88.2
Cluster 2	-78.9 (6.4)	109.1 (7.18)	8.2
Cluster 3	-66.9 (9.0)	149.1 (12.3)	3.6

**Table S.4. Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\alpha$  (1-3) core xylylated gx glycan.**

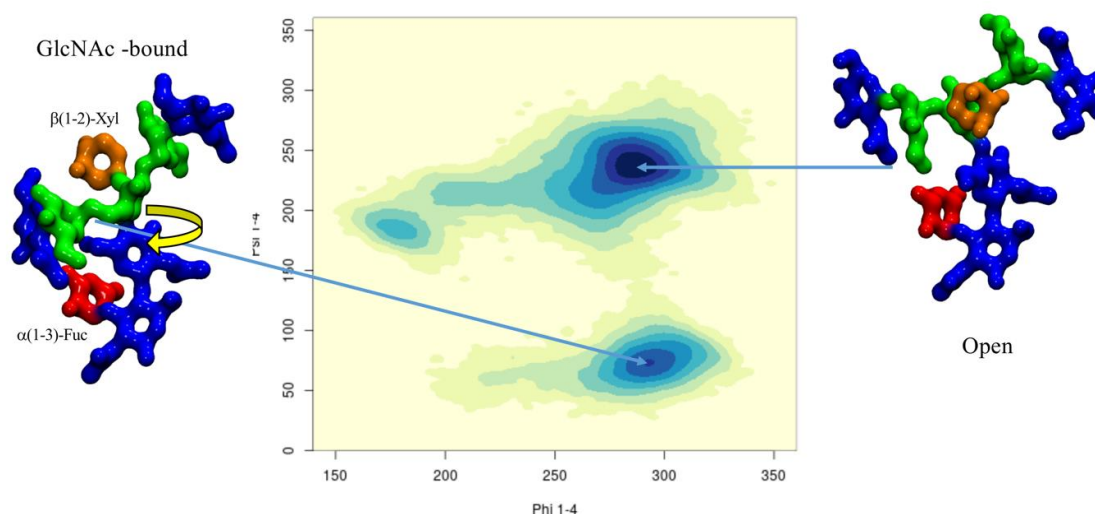
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.0 (10.1)	-130.2(15.7)	97.5
Cluster 2	-84.0 (5.4)	-64.8 (5.6)	2.5
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.2 (13.5)	-124.7 (15.0)	87.6
Cluster 2	-67.9 (14.7)	72.9 (11.7)	11.3
Cluster 3	-178.3 (6.9)	-175.5 (7.3)	1.0
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.2 (9.3)	138.4 (15.4)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.5 (10.4)	-173.6 (19.4)	70.0
Cluster 2	71.8 (9.6)	103.8 (12.9)	25.7
Cluster 3	161.6 (8.0)	132.5 (9.2)	2.3
Cluster 4	79.0 (6.7)	-80.2 (8.05)	1.59
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-82.1 (15.0)	161.3 (12.2)	88.2
Cluster 2	-77.7 (7.47)	112.5 (7.0)	10.0
Cluster 3	-147.3 (9.0)	98.9 (9.4)	1.9

<b>Gal <math>\beta</math>(1-3) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.1 (10.0)	126.1 (15.5)	80.8
Cluster 2	-83.6 (11.4)	-63.0 (7.57)	19.2
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	72.5 (10.7)	125.2 (17.4)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.5 (14.2)	162.4 (13.0)	91.2
Cluster 2	-77.8 (7.4)	111.1 (7.0)	8.8
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.1 (7.2)	125.5 (11.8)	100

**Table S.5.** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\beta$ (1-2) core xylylated and  $\alpha$ (1-3) core fucosylated ngxf glycan.

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.3 (9.0)	140.5 (6.0)	100
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.2 (11.0)	-106.4 (21.0)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.9 (14.4)	-124.2 (15.3)	75.4
Cluster 2	-66.5 (11.3)	73.3 (10.8)	21.1
Cluster 3	179.0 (7.6)	-174.7 (8.2)	2.3
Cluster 4	-151.9 (8.2)	-146.5 (6.7)	1.2
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.8 (10.8)	135.9 (17.4)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.6 (9.0)	175.2 (17.6)	68.5
Cluster 2	71.7 (10.0)	106.5 (12.3)	24.5
Cluster 3	81.6 (8.0)	-75.0 (10.0)	4.4
Cluster 4	158.0 (8.9)	135.9 (10.4)	2.4
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-79.2 (14.4)	158.4 (24.9)	90.1
Cluster 2	-80.0 (6.33)	113.0 (7.3)	9.3
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	69.1 (9.4)	113.8 (16.8)	1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>

Cluster 1	-77.4 (13.3)	161.5 (12.2)	88.2
Cluster 2	-78.9 (6.4)	109.1 (7.18)	8.21
Cluster 3	-66.9 (9.0)	149.1 (12.3)	3.58



**Figure S.1.** Conformational analysis of the (1-6) arm in terms of phi/psi torsion values. Representative structures selected from MD sampling are shown on the left- and right-hand side of the heat map. The GlcNAc-bound conformation is obtained through a torsion of the Man3 “glycoblock” relative to the chitobiose and it is stabilized by hydrogen bonding interactions between the  $\beta(1-2)$ -Xyl and  $\alpha(1-3)$ -Fuc. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

**Table S.6** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\beta(1-2)$  xylylated and  $\alpha(1-3)$  core fucosylated gxf glycan.

<b>Fuc <math>\alpha(1-3)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.1 (9.8)	140.5 (7.1)	88.9
Cluster 2	-156.8 (5.6)	90.7 (6.7)	11.1
<b>GlcNAc <math>\beta(1-4)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.2 (8.5)	-107.0 (8.0)	100
<b>Man <math>\beta(1-4)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.7 (15.1)	-124.5 (17.0)	85.2
Cluster 2	-179.5 (8.5)	-174.7 (9.5)	8.3
Cluster 3	-63.0 (9.9)	-75.6 (10.3)	5.5
Cluster 4	-153.1 (6.2)	-147.4 (5.31)	1

<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.2 (9.9)	135.9 (17.4)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.73 (10.3)	-174.2 (19.0)	87.2
Cluster 2	70.8 (8.1)	101.5 (10.7)	9.3
Cluster 3	161.6 (8.5)	131.9 (10.6)	3.5
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.1 (14.3)	161.9 (12.9)	84.7
Cluster 2	-76.1 (7.6)	111.6 (7.2)	10.5
Cluster 3	-147.1 (10.9)	99.5 (10.1)	3.2
Cluster 4	67.4 (7.4)	152.4 (8.4)	1.6
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.7 (10.4)	125.0 (16.8)	96.7
Cluster 2	-81.6 (7.07)	-64.6 (6.4)	3.3
Cluster 3	-81.6 (7.07)	-64.6 (6.4)	1.3
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.7 (9.1)	114.1 (16.6)	1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.2 (14.7)	162.0 (13.0)	90.1
Cluster 2	-78.1 (7.2)	112.1 (6.8)	9.9
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.9 (6.9)	126.0 (11.9)	100

**Table S.7. Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\alpha$ (1-3) core fucosylated LeA glycan.**

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>sPhi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.7 (9.8)	141.8 (6.6)	82.8
Cluster 2	-156.3 (7.02)	88.8 (17.6)	17.2
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.6 (8.7)	-107.4 (7.4)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.2 (12.0)	-124.5 (14.0)	97.1
Cluster 2	179.7 (6.1)	-178.2 (6.6)	2.9
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.5 (7.4)	148.1 (7.4)	53.6
Cluster 2	-70.1 (5.8)	-177.0 (11.2)	28.2
Cluster 3	-73.2 (6.7)	-101.9 (6.0)	9.4

Cluster 4	-148.0 (6.3)	-165.3 (5.8)	8.8
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-83.9 (10.5)	132.8 (14.0)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-150.0 (9.3)	98.5 (7.9)	40.9
Cluster 2	-91.5 (8.5)	151.6 (8.6)	29.9
Cluster 3	-62.7 (8.5)	161.4 (9.9)	29.2
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.1 (7.6)	131.5 (7.6)	100
<b>Fuc <math>\alpha</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-67.8 (7.5)	-101.26 (7.3)	100
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	61.5 (8.5)	111.0 (15.1)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.9 (16.9)	162.3 (12.3)	57.8
Cluster 2	-79.9 (11.4)	105.1 (12.4)	27.9
Cluster 3	66.1 (10.1)	153.8 (10.8)	14.26
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.6 (7.5)	134.5 (6.8)	100
<b>Fuc <math>\alpha</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-68.6 (8.9)	-100.9 (6.7)	98.4
Cluster 2	-148.9 (8.0)	-150.7 (3.8)	1.6

**Table S.8** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\beta$ (1-2) xylylated mgx glycan. Note: mg refers to the mammalian terminal  $\beta$ (1-4)-Gal.

<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.2 (10.9)	-131.1 (15.8)	97.5
Cluster 2	-79.6 (11.3)	66.6 (11.5)	2.5
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.7 (17.1)	-123.7 (14.7)	91.3
Cluster 2	-68.1 (12.6)	72.1 (11.9)	8.7
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-81.7 (19.1)	133.6 (20.1)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	72.2 (9.4)	103.5 (11.3)	56.9



Cluster 2	69.9 (8.3)	-173.8 (15.7)	39.8
Cluster 3	162.2 (9.1)	131.2 (9.3)	1.7
Cluster 4	58.5 (3.2)	59.98 (4.3)	1.6
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.1 (14.3)	161.9 (12.9)	84.7
Cluster 2	-76.1 (7.6)	111.6 (7.2)	10.5
Cluster 3	-147.1 (10.9)	99.5 (10.1)	3.2
Cluster 4	67.4 (7.4)	152.4 (8.4)	1.6
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.9 (15.8)	-119.9 (16.0)	98.9
Cluster 2	-73.5 (12.6)	-73.24 (12.5)	0.6
Cluster 3	63.6 (10.7)	-117.8 (6.8)	0.8
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.9 (9.7)	114.5 (16.7)	1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-79.0 (13.8)	162.5 (12.2)	87.5
Cluster 2	-80.9 (9.12)	109.4 (9.0)	12.5
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.41 (10.9)	-118.4 (15.1)	100

**Table S.9** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 4.5  $\mu$ s MD sampling of the  $\beta$ (1-2) xylylated and  $\alpha$ (1-6) core fucosylated mgmfx glycan. Note: mg refers to the mammalian terminal  $\beta$ (1-4)-Gal and mf to the mammalian core  $\alpha$ (1-6)-Fuc.

<b>Fuc <math>\alpha</math>(1-6) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-74.5 (9.6)	172.4 (14.5)	92.1
Cluster 2	-95.9 (4.5)	71.78 (6.1)	6.9
Cluster 3	-75.6 (2.5)	1113.8 (2.17)	1.0
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.2 (9.5)	-126.0 (14.3)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-74.3 (15.0)	-122.7 (14.2)	97.1
Cluster 2	-67.4 (10.1)	73.6 (10.7)	2.9
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>

Cluster 1	71.8 (9.4)	103.3 (10.98)	70.0
Cluster 2	70.4 (9.0)	-177.0 (11.2)	26.9
Cluster 3	67.5 (5.24)	-62.5 (5.7)	3.1
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-83.9 (10.5)	132.8 (14.0)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-91.5 (12.5)	159.9 (10.3)	96.3
Cluster 2	-75.5 (3.9)	113.6 (3.8)	2.9
Cluster 3	66.0 (6.8)	154.3 (5.8)	0.8
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-74.9 (13.56)	-122.1 (15.5)	97.5
Cluster 2	-83.6 (18.6)	65.3 (13.5)	2.5
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.7 (9.9)	115.6 (16.9)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-79.1 (13.2)	162.3 (11.4)	90.9
Cluster 2	-77.9 (6.8)	110.4 (7.6)	9.1
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.7 (15.4)	-118.6 (15.7)	97.3
Cluster 2	-74.0 (12.4)	70.12 (13.1)	2.7

**Table S.10 Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 4.5  $\mu$ s MD sampling of the  $\beta$ (1-2) xylylated and  $\alpha$ (1-3) core fucosylated nmgmfx glycan. Note: nmg refers to the absence of mammalian terminal  $\beta$ (1-4)-Gal and mf to the mammalian core  $\alpha$ (1-6)-Fuc.**

<b>Fuc <math>\alpha</math>(1-6) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.0 (9.6)	177.0 (14.8)	93.1
Cluster 2	-95.6 (4.0)	75.0 (5.4)	6.9
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.2 3(9.7)	-126.0 (14.4)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.6 (18.5)	-124.8 (16.9)	91.1
Cluster 2	-66.2 (12.5)	73.0 (11.9)	8.2
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.1 (9.2)	-173.6 (14.6)	71.5
Cluster 2	72.1 (8.3)	104.25 (9.7)	28.5

<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-80.6 (16.5)	135.8 (14.0)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-82.8 (14.94)	161.1 (11.9)	90.6
Cluster 2	-77.8 (7.3)	111.3 (6.8)	7.4
Cluster 3	66.0 (7.3)	152.9 (8.7)	2.0
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.9 (9.4)	114.4 (16.9)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.4 (13.4)	162.6 (11.4)	88.8
Cluster 2	-78.9 (7.2)	110.4 (7.0)	9.0
Cluster 2	66.45 (10.9)	152.6 (10.7)	2.2

**Table S.11 Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\beta$ (1-2) xylylated and  $\alpha$ (1-3) core fucosylated mgpfx glycan. Note: mg refers to the mammalian terminal  $\beta$ (1-4)-Gal and pf to the plant core  $\alpha$ (1-3)-Fuc.**

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.8 (10.7)	141.3 (8.9)	93.8
Cluster 2	-156.8 (7.7)	91.3 (9.3)	6.2
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.5 (8.8)	-107.3 (9.1)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.4 (15.1)	-123.6 (17.3)	87.0
Cluster 2	-68.2 (12.8)	70.9 (11.2)	13.0
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-80.7 (15.3)	134.5 (16.5)	100
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	69.9 (9.0)	-175.5 (15.0)	49.1
Cluster 2	72.4 (9.7)	105.1 (12.5)	46.7
Cluster 3	158.6 (11.8)	135.8 (13.6)	4.2
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-86.2 (15.0)	160.9 (11.4)	96.4
Cluster 2	-78.5 (6.1)	113.5 (4.9)	3.6
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.9 (15.0)	-119.9 (15.6)	97.2
Cluster 2	-74.1 (13.1)	70.1 (13.8)	2.8

<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.8 (9.6)	114.5 (16.7)	1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.8 (13.8)	162.6 (12.6)	88.3
Cluster 2	-80.7 (8.0)	109.5 (7.6)	9.6
Cluster 3	65.2 (9.9)	150.8 (10.8)	2.1
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.5 (11.5)	-118.6 (15.5)	100

**Table S.12 Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 1.5  $\mu$ s MD sampling of the  $\alpha$ (1-3) core fucosylated mgpf glycan. Note: mg refers to the mammalian terminal  $\beta$ (1-4)-Gal and pf to the plant core  $\alpha$ (1-3)-Fuc.**

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-71.1 (10.6)	140.3 (14.8)	98.1
Cluster 2	-156.5 (6.2)	91.4 (9.3)	1.9
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.0 (14.8)	-121.5 (15.7)	95.9
Cluster 2	-80.9 (15.3)	62.5 (13.4)	4.1
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.6 (14.1)	-124.9 (16.1)	77.9
Cluster 2	-153.7 (13.3)	-139.7 (8.5)	12.8
Cluster 3	-71.1 (12.5)	69.6 (11.4)	9.3
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	74.2 (13.0)	86.5 (14.7)	74.8
Cluster 2	70.3 (9.1)	-176.5 (14.7)	25.2
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-80.2 (14.6)	163.2 (12.4)	100
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.0 (14.88)	-121.58 (15.8)	95.9
Cluster 2	-80.9 (15.3)	62.5 (13.4)	4.1
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.95 (9.6)	140.95 (15.2)	73.0
Cluster 1	70.1(8.21)	101.2 (8.8)	27.0
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.5 (13.3)	162.1 (11.7)	91.2
Cluster 2	-77.9 (6.4)	110.7 (6.62)	8.8

<b>Gal <math>\beta</math>(1-4) GlcNAc (1-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.2 (10.9)	-118.4 (15.0)	100

**Table S.13 Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles (°) measured through a cumulative 2  $\mu$ s MD sampling of the  $\alpha$ (1-3) and  $\alpha$ (1-6) core fucosylated mgmfpf glycan. Note: mg refers to the mammalian terminal  $\beta$ (1-4)-Gal, pf to the plant core  $\alpha$ (1-3)-Fuc and mf to the mammalian core  $\alpha$ (1-6)-Fuc.**

<b>Fuc <math>\alpha</math>(1-6) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.0 (9.6)	-179.5 (14.8)	74.3
Cluster 2	-76.2 (4.0)	117.36 (12.1)	12.9
Cluster 3	-144.4 (7.7)	171.0 (5.5)	12.8
<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.2 (10.3)	140.1 (9.7)	88.6
Cluster 2	-157.3 (7.7)	91.4 (8.9)	11.4
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.8 (9.5)	-106.4 (14.1)	91.7
Cluster 2	-154.9 (10.6)	-147.8 (7.4)	8.3
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.1 (10.9)	-120.7 (12.6)	74.8
Cluster 2	-153.0 (13.0)	-139.9 (8.3)	25.2
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	73.5 (11.7)	86.6 (17.32)	85.1
Cluster 2	69.5 (9.9)	-176.3 (15.8)	14.9
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-85.1 (14.9)	161.7 (13.4)	100
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-73.9 (11.5)	-123.5 (16.5)	98.7
Cluster 2	-146.2 (8.5)	-142.1 (6.9)	1.3
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	71.1 (9.3)	140.5 (15.6)	72.9
Cluster 1	69.9 (8.8)	100.63(9.8)	27.1
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.7 (13.8)	161.1 (12.8)	90.4
Cluster 2	-79.1 (7.7)	110.7 (7.2)	9.6
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.4 (10.9)	-118.6 (15.2)	95.0
Cluster 1	-143.2 (10.4)	-144.4 (5.9)	2.9
Cluster 1	-74.2 (9.3)	67.6 (9.4)	2.1

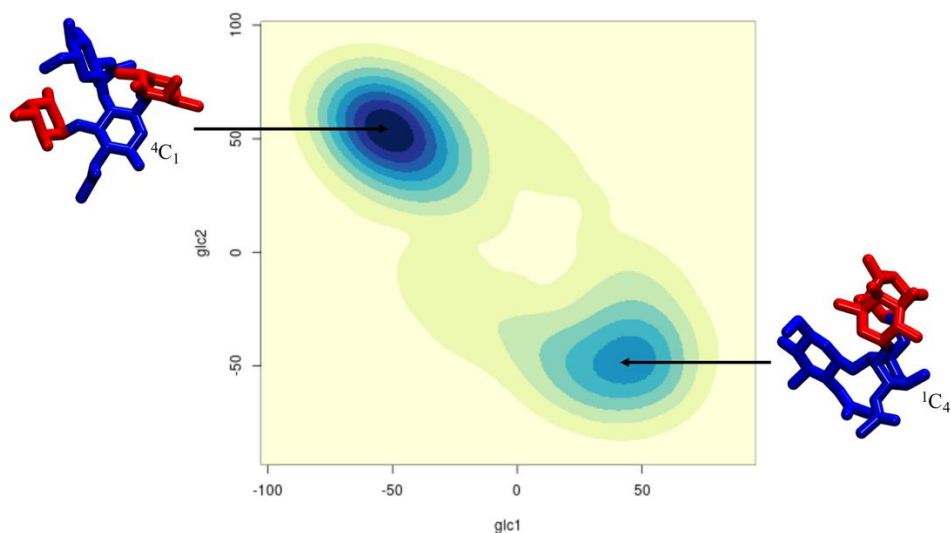


Figure S.2. Conformational analysis N-linked GlcNAc pucker along the 1.5 ms cumulative sampling of the  $\alpha(1-3)$  and  $\alpha(1-6)$  core fucosylated A2G2 (mgmpf) N-glycan. Representative structures of the fucosylated chitobiose selected from MD sampling of the whole N-glycans are shown on the left- and right-hand side of the heat map where the ring pucker is also indicated. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

Table S.14 Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 2  $\mu$ s MD sampling of the  $\beta(1-2)$  xylylated and  $\alpha(1-3)$  and  $\alpha(1-6)$  core fucosylated mgxmpf glycan. Note: mg refers to the mammalian terminal  $\beta(1-4)$ -Gal, pf to the plant core  $\alpha(1-3)$ -Fuc and mf to the mammalian core  $\alpha(1-6)$ -Fuc.

<b>Fuc <math>\alpha(1-6)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.9 (11.25)	175.65 (14.0)	79.3
Cluster 2	-142.7 (9.8)	171.3(6.2)	16.9
Cluster 3	-95.48 (4.4)	71.95(4.8)	3.8
<b>Fuc <math>\alpha(1-3)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.6 (10.3)	141.4 (8.3)	88.8
Cluster 2	-156.0 (8.1)	89.5 (8.7)	11.2
<b>GlcNAc <math>\beta(1-4)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.8 (8.9)	-106.6 (11.0)	96.7
Cluster 2	-82.1 (7.1)	-154.8 (6.8)	3.3
<b>Man <math>\beta(1-4)</math> GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-75.5 (11.0)	-123.5 (12.6)	94.6
Cluster 2	-62.42 (7.9)	72.2 (9.5)	4.4

<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	73.4 (10.5)	102.8 (10.6)	55.3
Cluster 2	70.4(9.4)	-176.8 (16.7)	26.7
Cluster 2	102.5 (9.3)	55.9 (7.7)	17.9
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-91.8 (13.5)	159.4 (11.1)	100
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-74.2 (11.1)	-121.1 (15.3)	99.2
Cluster 2	-146.6 (7.5)	-143.6 (4.9)	0.8
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	68.7 (9.7)	115.4 (16.8)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-79.5 (13.8)	162.3 (12.4)	92.4
Cluster 2	-79.3 (6.9)	110.3 (6.9)	7.6
<b>Gal <math>\beta</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.6 (11.0)	-119.1 (15.2)	93.4
Cluster 1	-143.2 (10.4)	-144.4 (5.9)	4.9
Cluster 1	-74.2 (11.6)	69.7 (12.1)	1.7

**Table S.15** Results of the clustering analysis showing the median and standard deviation values (in parenthesis) for the torsion angles ( $^{\circ}$ ) measured through a cumulative 2  $\mu$ s MD sampling of the  $\beta$ (1-2) xylylated and  $\alpha$ (1-3) fucosylated A2 glycan terminating with LeX on both arms.

<b>Fuc <math>\alpha</math>(1-3) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-70.2 (11.0)	141.9 (8.4)	90.8
Cluster 2	-156.7 (7.6)	90.2 (8.6)	9.2
<b>GlcNAc <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-72.2 (8.6)	-107.5 (7.7)	100
<b>Man <math>\beta</math>(1-4) GlcNAc</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-76.4 (15.7)	-124.6 (16.7)	75.0
Cluster 2	-68.1 (9.7)	71.0 (10.1)	25.0
<b>Man <math>\alpha</math>(1-6) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	70.7 (9.2)	-173.6 (13.4)	89.7
Cluster 2	152.4 (12.9)	145.6 (12.3)	10.3
<b>Xyl <math>\beta</math>(1-2) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-77.6 (9.6)	140.1 (18.1)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>



Cluster 1	-78.2 (14.3)	162.3 (13.1)	91.8
Cluster 2	-76.9 (7.4)	111.1 (6.8)	8.2
<b>Fuc <math>\beta</math>(1-3) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-69.7 (8.9)	142.1 (7.0)	100
<b>Gal <math>\alpha</math>(1-4) GlcNAc (I-6)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-67.7 (8.0)	-107.9(7.2)	100
<b>Man <math>\alpha</math>(1-3) Man</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	61.5 (8.5)	111.0 (15.1)	100
<b>GlcNAc <math>\beta</math>(1-2) Man (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-78.8 (13.9)	162.6 (11.7)	91.6
Cluster 2	-79.1 (6.2)	112.5 (6.2)	8.3
<b>Gal <math>\beta</math>(1-3) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-69.1 (9.45)	142.2 (8.0)	100
<b>Fuc <math>\alpha</math>(1-4) GlcNAc (I-3)</b>	<b>Phi</b>	<b>Psi</b>	<b>Pop(%)</b>
Cluster 1	-67.8 (7.9)	-108.1 (7.4)	94.8
Cluster 2	-153.9 (8.8)	-141.5 (7.7)	5.2

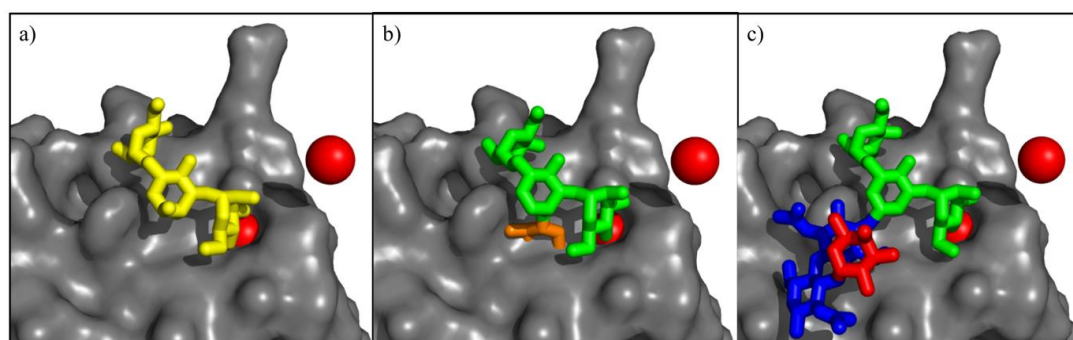


Figure S.3. Structural alignment of different plant N-glycoforms from our MD simulations on to the DC-SIGN/GlcNAc2Man3 complex resolved at 2.5 Å resolution (PDBid 1k9i). Panel a) The water accessible surface of the DC-SIGN (chain C) binding site is shown in grey and the Man3 region of the co-crystallized glycan in 1k9i is rendered as yellow sticks.  $\text{Ca}^{2+}$  ions are shown as red spheres. Panel b) Structural alignment of representative structure from our MD simulation of the ngx plant N-glycan shows that the  $\beta$ (1-6) xylose sterically hinders binding by clashing with the surface of the binding site. Only the Xyl-Man3 glycoblock from the whole N-glycan is represented. Panel c) Structural alignment of representative structure from our MD simulation of the ngf plant N-glycan shows that the  $\alpha$ (1-3) fucose does not hinder recognition or binding by DC-SIGN. Only the Man3 and  $\alpha$ (1-3)-Fuc chitobiose glycoblocks from the whole N-glycan is represented. The monosaccharides colouring, aside from panel a), follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with RStudio ([www.rstudio.com](http://www.rstudio.com)).

## References

1. Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. *AMBER 2018*, University of California, San Francisco, 2018.
2. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* **2008**, *29* (4), 622-55.
3. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M., Comparison of simple potential functions for simulations of liquid water. *Journal of Chemical Physics* **1983**, *79* (2), 926-935.
4. Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E., Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **2019**, *29* (1), 94-103.
5. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14* (1), 33-8, 27-8.

## Appendix III

### **Computational Method**

The crystal structure of the human Fc fragment with PDBid 1FC1 was used as a starting point for the preparation of all our systems. The N-glycans in this structure were functionalized to match our chosen glycoforms, therefore a  $\beta(1-4)$  galactose and  $\alpha(2-6)$  sialic acid were added to the  $\alpha(1-3)$  arm using the CHARMM GUI PDB reader tool (Park, S.J., Lee, J., et al. 2019). This structure was edited using the academic version of Schrödinger's Maestro v.10.7.015 to rotate the torsion angles of the glycosidic linkages to release any steric clash that may have occurred during the build. The values of the glycosidic linkage torsion angles were taken from our earlier work on the unlinked glycans (Harbison, A.M., Brosnan, L.P., et al. 2019). Protein residues Cys 239 to Gly 250 from the human IgG B12 with PDBid 1HZH were linked to the structurally aligned 1FC1 using Maestro to build the disulfide bonds that keep the Fc region stable during the simulation. To ensure by construction that a potential folded conformation of the  $\alpha(1-6)$  arm was sampled, all simulations were started with both outstretched (open) and folded (closed) conformations of the arm. Maestro (Schrodinger, Maestro, 2012) was also used to modify the  $\alpha(1-6)$  glycosidic linkage to get the open and closed conformations for a total of six starting structures.

Each system was prepared through an initial 500k cycles of conjugate gradient minimization with a restraint of  $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  on all heavy atoms. The cutoff for van der Waals interactions was set to 12  $\text{\AA}$  and smoothing functions were applied between 11 and 13.5  $\text{\AA}$ . Particle Mesh Ewald (PME) was used to treat electrostatic interactions with a charge grid of 1  $\text{\AA}$  and a sixth order spline function for mesh interpolation. All non-bonded interactions not directly connected were excluded, namely 1-3 pair interaction, or scaled by 0.8333, namely 1-4 pair interaction. Following energy minimization, the systems were heated from 0 to 300 K over 600 cycles with restraints on all heavy atoms in place. For the temperature Replica Exchange Molecular Dynamics (REMD) (Sugita, Y. and Okamoto, Y. 1999), 90 replicas were generated in the temperature range between 300 and 500 K. The systems were equilibrated with restraints on the protein backbone atoms and on the glycans heavy atoms for 500 ps, followed by 500 ps of unrestrained equilibrations for each replica. The production steps ranged between 13 ns and 11 ns for each replica, with an integrated time step of 2 fs. The SHAKE algorithm was used to restrain bonds to hydrogen atoms. Data analysis and structural alignments were done with VMD v.1.9.3 beta 1 (Humphrey, W., Dalke, A., et al. 1996) and *seaborn* (<https://seaborn.pydata.org>) was used for all statistical data visualization.

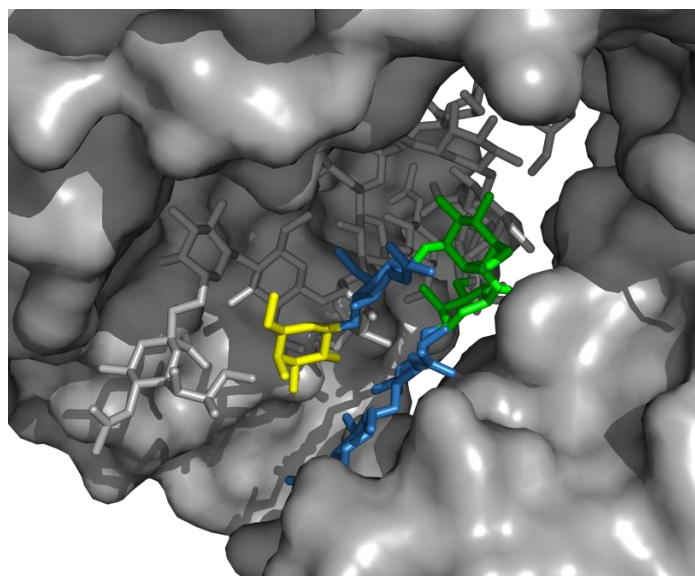


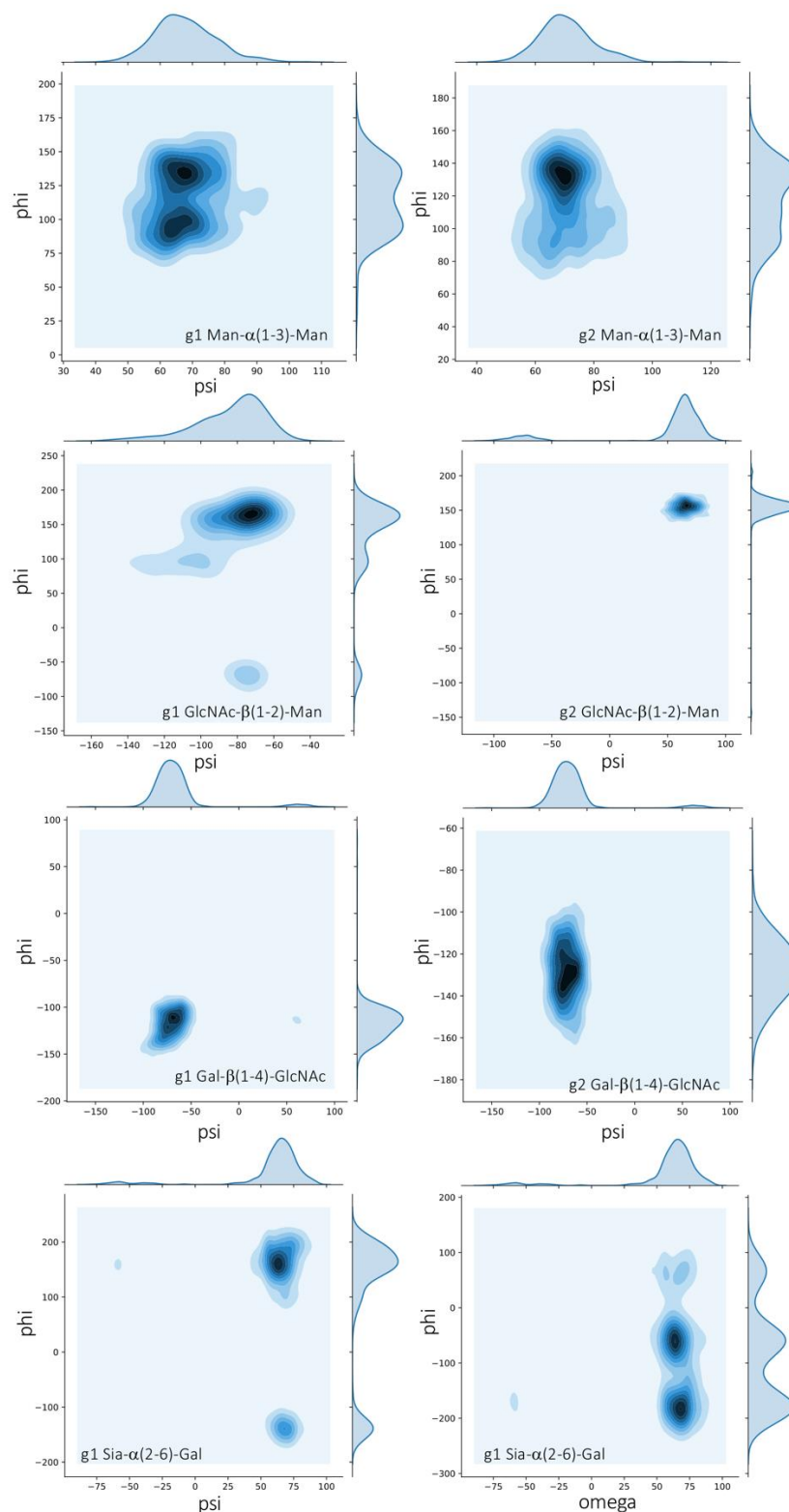
Figure S.2 Structure of the  $\alpha(1-6)$  arm in a snapshot from the end of the *oo* IgG1 Fc closed simulation (11 ns x 90 replicas) shows an extremely compact Fc core where the N-glycans are compressed together. The chitobiose and the  $\alpha(1-6)$  arm are highlighted with SNFG colouring, while the protein is showed through a solvent accessible area in gray.

Table S. I Average protein backbone RMSD values ( $\text{\AA}$ ) calculated through the REMD simulations of the IgG Fc linked to different sets of N-glycans. Sugars *o* and *p* are shown in Figure 1. Standard deviation values are shown in parenthesis. The “open” and “closed” labels correspond to simulations started from an outstretched (open) or from a folded-over (closed) conformation of the  $\alpha(1-6)$  arm.

N297-glycans	Fc		CH2		CH3	
	open	closed	open	closed	open	closed
<i>pp</i>	3.3 (0.9)	3.3 (0.9)	4.4 (1.3)	4.4 (1.2)	1.1 (0.2)	1.0 (0.1)
<i>oo</i>	2.5 (0.6)	3.9 (1.3)	3.3 (0.8)	5.2 (1.8)	1.0 (0.2)	1.1 (0.2)
<i>op (nFuc side)</i>	3.1 (0.8)	3.4 (0.8)	4.1 (1.1)	4.6 (1.0)	1.0 (0.2)	1.1 (0.1)
<i>op (Fuc side)</i>	3.3 (1.2)	4.1 (1.3)	4.4 (1.7)	5.6 (1.8)	1.1 (0.1)	1.1 (0.2)

**Table S. II Average RMSD values (Å) calculated over all N-glycans heavy atoms. For each N-glycan the alignment was done over all the heavy atoms of the core chitobiose and the arms were considered from the central Man. Standard deviation values are shown in parenthesis.**

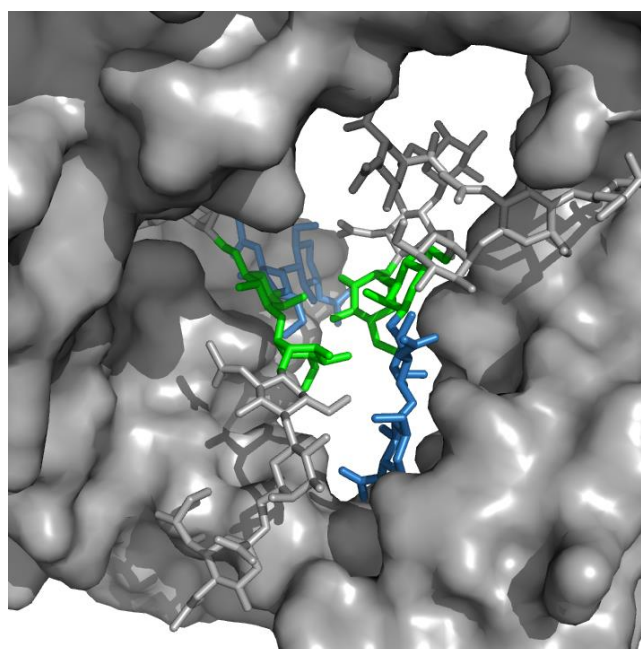
N297-glycans	$\alpha(1-6)$		$\alpha(1-3)$	
	<i>g1</i>	<i>g2</i>	<i>g1</i>	<i>g2</i>
<i>pp</i>	4.1 (1.8)	4.7 (2.2)	9.5 (3.2)	7.1 (2.0)
<i>oo</i>	2.4 (1.1)	4.1 (1.3)	8.0 (1.8)	4.4 (1.1)
<i>op</i>	3.7 (1.9) <i>-nf</i>	5.7 (2.1) <i>-f</i>	10.4 (3.9) <i>-nf</i>	5.8 (1.6) <i>-f</i>



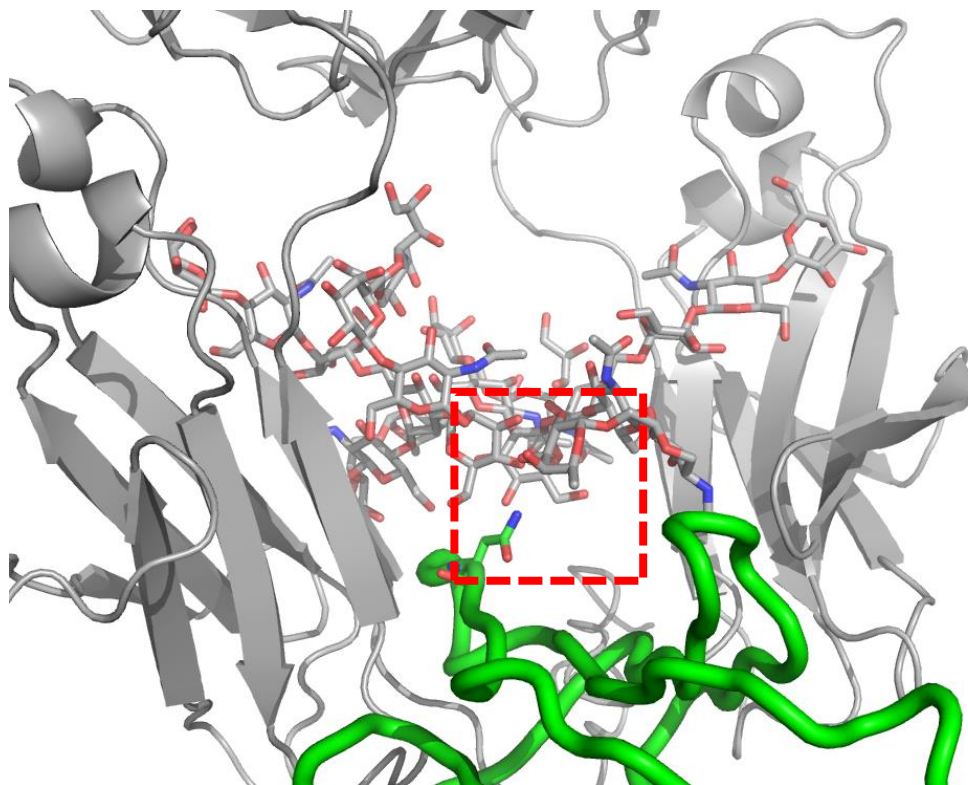
**Figure S. 3** Heat maps (Ramachandran plots) representing the conformational propensities in terms of torsion angles of the  $\alpha$ (1-3) arms throughout the simulation of the pp IgG Fc. Only the data on g1 are shown for the Sia- $\alpha$ (2-6)-Gal linkage for simplicity. The maps are built on 6500 data points per axis and were made with seaborn (<https://seaborn.pydata.org/>).

**Table S. III. Averaged torsion angle values calculated for one of the two  $\alpha(1-3)$  arms (g1) during the simulation of the pp IgG Fc. Standard deviations are indicated in parenthesis and populations (%) are highlighted in red.**

<i>pp</i> $\alpha(1-3)$	phi		psi		omega	
Man- $\alpha(1-3)$ -Man	68 (9) <b>100</b>	-	96 (15) <b>55</b>	138 (11) <b>45</b>	-	-
GlcNAc- $\beta(1-2)$ - Man	-77 (12) <b>82</b>	-116 (14) <b>18</b>	178 (44) <b>79</b>	98 (13) <b>21</b>	-	-
Gal- $\beta(1-4)$ - GlcNAc	-72 (13) <b>95</b>	61 (11) <b>5</b>	-118 (15) <b>100</b>	-	-	-
Sia- $\alpha(2-6)$ -Gal	66 (11) <b>93</b>	-50 (16) <b>7</b>	-178 (27) <b>87</b>	97 (17) <b>11</b>	-177 (15) <b>44</b>	-62 (19) <b>38</b>

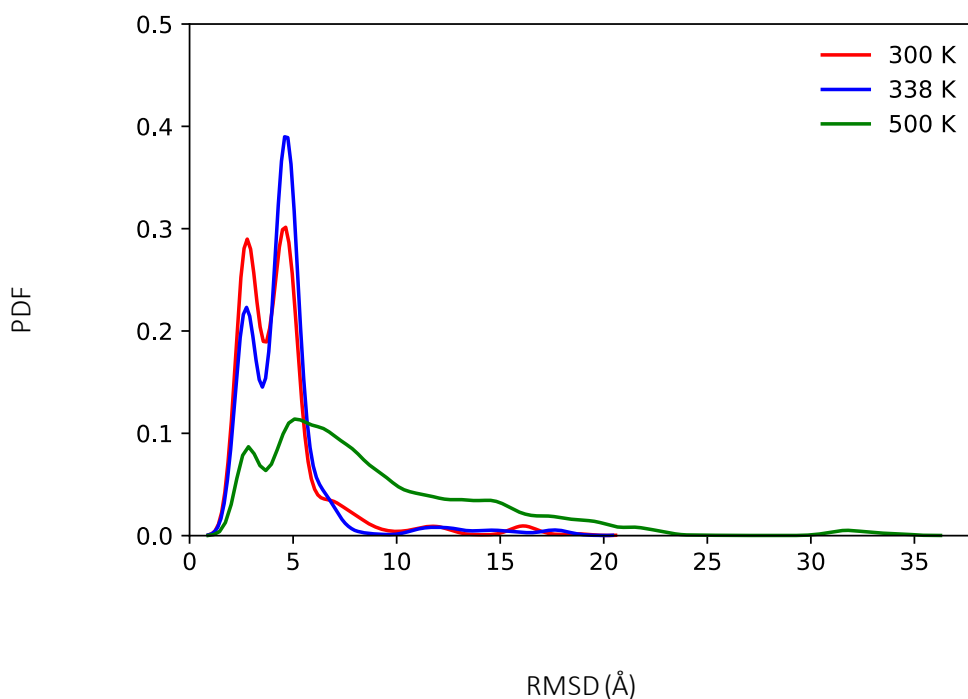


**Figure S. 4 The two trimannose cores (highlighted by SNFG colouring) in the symmetrically opposed N-glycans shown in a snapshot from the simulation of the oo IgG1 Fc. The interactions between the residues are primarily hydrogen bonds and are interchanging continuously throughout the simulation.**



**Figure S. 5** Alignment of a representative structure from the pp IgG1 Fc REMD simulation represented in grey and the structure of the complex between the IgG1 Fc and the Fc $\gamma$ RIII, shown in green tubes, with PDBid 1E4K(Sondermann, P., Huber, R., et al. 2000). The IgG1 Fc from PDBid 1E4K is omitted for clarity. The position of the core fucose is highlighted in a red-dashed frame, the Asn 162 of the Fc $\gamma$ RIII is represented with sticks for clarity.





**Figure S. 6** KDE distributions of the distance (Å) between one of the CH2 Glu 252 carboxylic oxygen and the O4 of the terminal Gal in the N-glycan  $\alpha(1-6)$  arm in function of temperature. This distance was used as a parameter to gauge the position of the  $\alpha(1-6)$  arm relative to CH2 and to identify the bound (values within hydrogen bond distance) from the unbound (values above hydrogen bond distance)

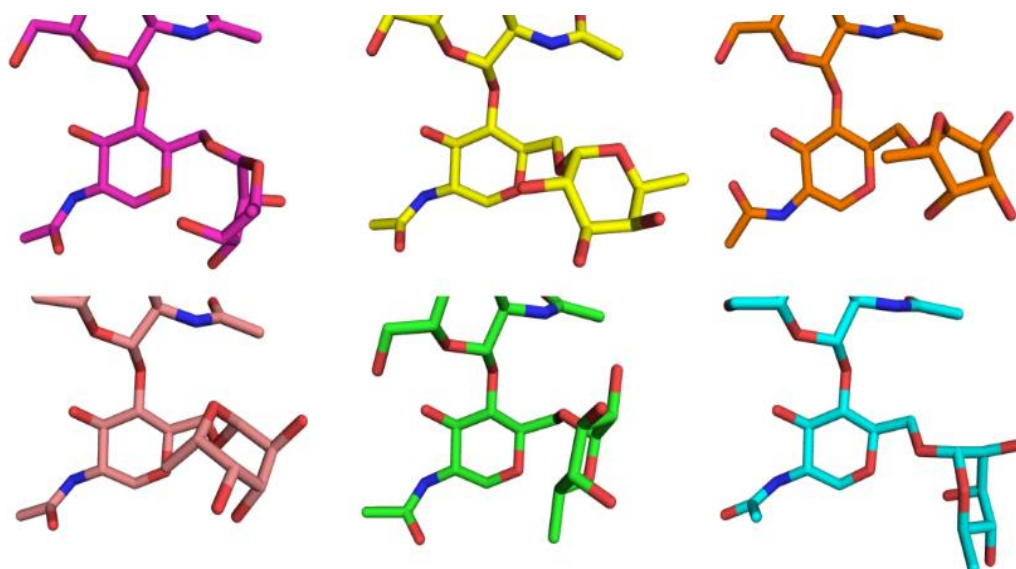
### Rationale for the selection of the starting structure

The glycan fragment database (Jo, S. and Im, W. 2013) was used to search the PDB for IgG crystal structures with N-glycans that were: di-galactosylated, core-fucosylated and sialylated on the  $\alpha(1-3)$  arm. No IgG was found with these specifications and only one with a di-galactosylated N-glycan, namely PDBid 1HZH, which does not fit our requirements as it has an asymmetric glycosylation pattern. Based on these results we decided that the essential minimum requirements to be: a) a human IgG1 structure, b) with core-fucosylation and c) terminal Gal on the  $\alpha(1-6)$  arm and d) with symmetric glycosylation on both sides of the Fc. The terminal Gal requirement was determined by our aim to understand the significance of galactosylation on the conformational propensity of the linked N-glycan and directly compare the results to our earlier study of the isolated N-glycans (Harbison, A.M., Brosnan, L.P., et al. 2019). Out of all the structures in **Table S.IV**, only two structures, namely 1FC1 (2.9 Å resolution) and 3DO3 (2.5 Å resolution), were found to match these minimum criteria and between the two which in many aspects were equivalent, we chose 1FC1. As shown in **Figure S.6** some of the structures in **Table S.IV** were eliminated because of highly speculative ring conformations or wrong anomeric configuration of the fucose.

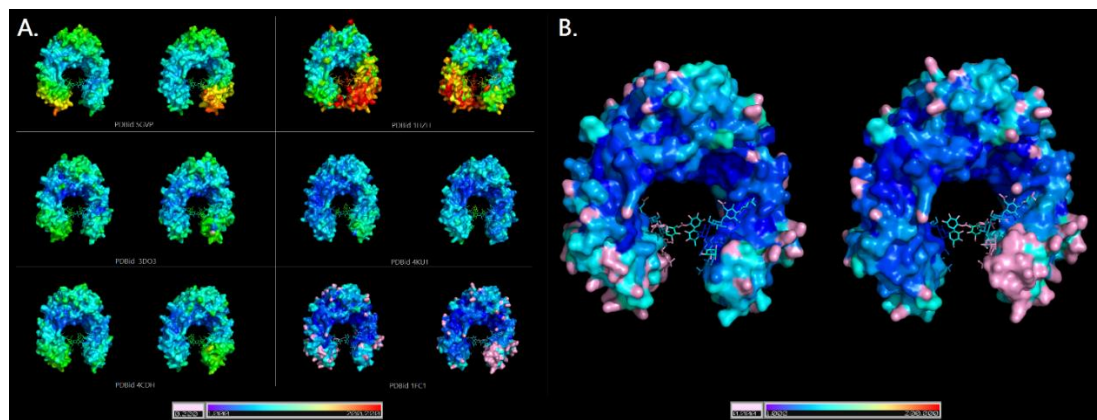
**Table S.IV: Results from the Glycan fragment DB search of the PDB for Fc structures with core-fucosylated, (1-6) galactosylated N-glycans**

<b>PDB id</b>	<b>Contents of Crystal Structure</b>	<b>Resolution (Å)</b>
<b>Exact glycan match</b>		
5VGP	IgG1	2.12
5XJE	IgG1 Fc receptor complex	2.40
5YC5	IgG1 Fc receptor complex	2.71
2RGS	IgG2	2.10
4C55	IgG4	2.35
4W4O	IgG4 receptor complex	1.80
5VAA	mutant IgG	1.55
4BSV	mutant IgG	1.75
4BSW	mutant IgG	2.15
5XJF	mutant IgG	2.50
5TPS	mutant IgG	2.70
5W5L	mutant IgG1	1.90
5D4Q	mutant IgG1	2.39
5HYI	mutant IgG1	2.90
5VAA	mutant IgG4	1.55
5W5N	mutant IgG4	1.85
5W5M	mutant IgG4	1.90
<b>1-6 branch Galactosylated, core-Fucosylated Fragment</b>		
1HZH	IgG1 b12	2.70
4BYH	IgG 1-6 sialylation	2.30
3WKN	IgG complex with artificial protein	2.90
5D6D	IgG1 Fc receptor complex	3.13
<b>1-6 branch Galactosylated Fragment</b>		
1FC1	IgG1	2.90
3DO3	IgG1	2.50
4CDH	IgG1	2.30
4KU1	IgG1	1.90

1FC2	IgG1 (one chain only)	2.80
2VUO	IgG (rabbit)	1.95
1FRT	IgG Fc receptor complex	4.50
1E4K	IgG1 Fc receptor complex	3.20
3AY4	IgG1 Fc receptor complex	2.20
3RY6	IgG1 Fc receptor complex	3.80
4X4M	IgG1 Fc receptor complex	3.49
5VU0	IgG1 Fc receptor complex	2.26
6EAQ	IgG1 Fc receptor complex	2.22
6EA7	IgG1 receptor complex	4.25
2RGS	IgG2B mab (mouse)	2.13
5HVW	IgG4	1.95
2QL1	mutant IgG1	2.53
3C2S	mutant IgG1	2.30
4BSV	mutant IgG1	1.75
4Q74	mutant IgG1	2.19
5M3V	mutant IgG1	1.97
5W5L	mutant IgG1	1.90
6F2Z	mutant IgG1	2.30
3WJJ	mutant IgG1 and complex	2.60
3WJL	mutant IgG1 and complex	2.86
3WN5	mutant IgG1 and complex	2.78
5BW7	mutant IgG1 and complex	3.00
5XJF	mutant IgG1 and complex	2.50
5IW3	mutant IgG1 mab	2.05
5W5N	mutant IgG4	1.85
1L6X	Riuxmab Fc fragment	2.12
4Q6Y	IgG1 1-6 sialylation	3.00



**Figure S.6:** 4KU1 (magenta), 5VGP(yellow), 4CDH(orange), 3DO3(light pink), 1HZH(green) and 1FC1(cyan) glycans from the crystal structure with  $\alpha$  or  $\beta$ -L-fucose with the ring conformation distorted in 4KU1 and 4CDH in particular. The B factors from each crystal structure are shown in **Figure S.7**. 1FC1 has the lowest B factors. However, there are atoms with a B factor of 0.0. The 1FC1 structure is not obtained by molecular replacement.



**Figure S.7.** Panel A. Fc regions from the selected PDBs, coloured based on their B-factors, with dark blue being the lowest B-factors and dark red being the highest B-factors recorded. Those atoms with a B-factor of 0.0 were coloured light pink to differentiate them from those who are represented by the rainbow spectrum, as B-factors of 0.0 can be from error in refinement calculations. Panel B. A closer view of the Fc region and glycans from the 1FC1 crystal structure, coloured using the same colouring scheme for B-factors mentioned above.

Shown in **Table S.V** is a RMSD analysis of based on backbone atoms of the six IgG1 Fc structures from Table S.4 that fit the glycosylation pattern requirements on at least one glycan. This analysis demonstrates how similar all these structures are.

**Table S.V: Calculated RMSD for the selected IgG1 crystal structures by alignment of the Fc protein backbone.**

<b>PDB id</b>	4KU1	5VGP	4CDH	3DO3	1HZH	1FC1
	RMSD (Å)					
4KU1		0.718	0.726	0.771	2.04	0.893
5VGP			0.258	0.294	0.718	0.486
4CDH				0.342	1.429	0.496
3DO3					1.08	0.554
1HZH						1.284
1FC1						

In addition to the analysis of the protein backbone similarities between different crystal structures, we also analysed the Fc N-glycans' structures, shown in **Table S.VI**. Due to the flexibility of the CH2 domain, the position of the two glycans vary between each crystal structure, depending on the conformation of the C'E loop, see **Figure S.8**.

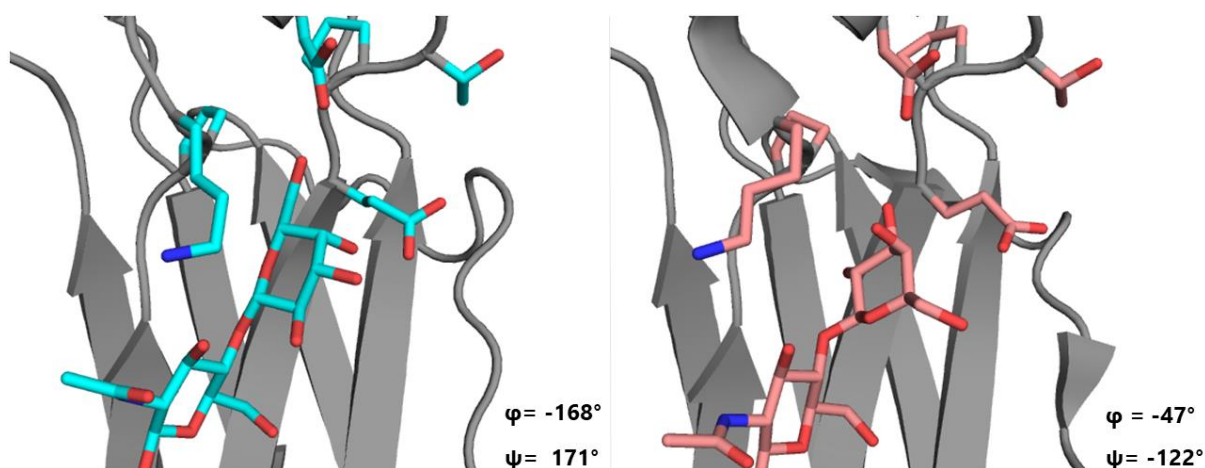
In terms of sequences, the crystal structures 5VGP, 4CDH, 1HZH and 4KU1 are asymmetrically glycosylated, meaning that the two N-glycans have different sequences. While IgGs can be asymmetrically glycosylated, asymmetric glycoforms result in asymmetric protein conformation in the Fc region (Masuda, K., Yamaguchi, Y., et al. 2000). Based on these consideration and observations, we could only consider crystal structures 3DO3 and 1FC1 as suitable as starting structures for our simulations.

**Table S.VI: N-glycans sequence on the two CH2 domains from selected crystal structure. The asterisk (\*) indicates the presence of  $\beta$ -L-fucose instead  $\alpha$ -L. Note, we replaced  $\beta$ -L-fucose with  $\alpha$ -L-fucose in all our starting structures.**

<b>PDB id</b>	<b>Glycan sequence, Chain A</b>	<b>Glycan sequence, Chain B</b>
4KU1	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*	Gal <sub>2</sub> GlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*
5VGP	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc	GlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc
4CDH	Gal <sub>2</sub> GlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub>
3DO3	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*
1HZH	Gal <sub>2</sub> GlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub>	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc
1FC1	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*	GalGlcNAc <sub>2</sub> Man <sub>3</sub> GlcNAc <sub>2</sub> Fuc*

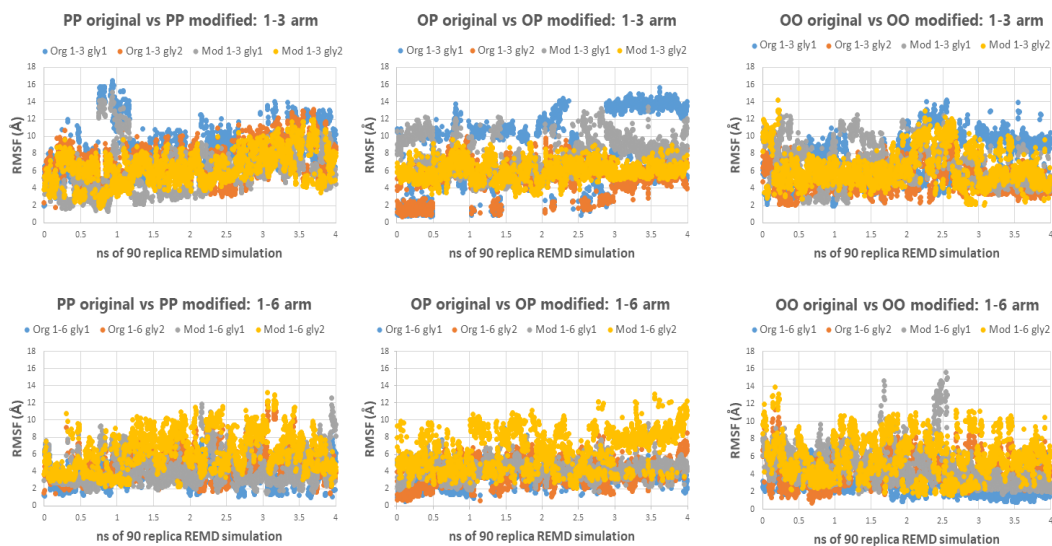
### **Effect of different conformations of the $\alpha$ (1-6) Gal on the arm's dynamics**

A difference between the 1FC1 structure and 3DO3 pre-selected as potential candidates as starting structure is the conformation of the Gal- $\beta$ (1-4)-GlcNAc, see **Figure S.9**. To assess the impact of the starting conformation of the Gal on the conformational propensity of the  $\alpha$ (1-6) arm, we ran a set of simulations of our three differently glycosylated starting structures, namely *pp*, *oo* and *op*, with the conformation of the galactose modified by rotation of the  $\phi$  and  $\psi$  angles to match the conformation seen in 3DO3.



**Figure S.9.** Conformational difference of the bound galactose to the CH2 protein surface in the crystal structure of 1FC1(left, cyan) and 3DO3(right, light pink). Values given for the  $\phi$  and  $\psi$  torsion angles of the glycosidic linkage from the corresponding crystal structure.

Results are presented below in **Figure S.10** in terms of RMSF values calculated throughout the trajectories at 300 K to measure changes in conformational flexibility of both  $\alpha(1-3)$  and  $\alpha(1-6)$  arm from different starting conformation of the Gal- $\beta(1-4)$ -GlcNAc linkage.



**Figure S.10.** Root Mean Square Fluctuation (RMSF) values indicating the different dynamics/flexibility of the  $\alpha(1-3/6)$  arm in function of the different Gal conformation calculated throughout the six trajectories. “Original” indicates the Gal orientation as found in 1FC1, while “modified” represents the orientation found in 3DO3.

The average RMSF values calculated are also shown in **Table S.7** below. The data do not show any significant change in the dynamics of the arms in function of the starting conformation of the Gal.

**Table S.VII: RMSF values calculated for the two arms in simulations started from a conformation of the Gal- $\beta$ (1-4)-GlcNAc linkage corresponding to the 1FC1. Standard deviation values are indicated in parenthesis.**

<b>1FC1</b>	<i>pp</i>	<i>op</i>	<i>oo</i>
$\alpha$ (1-3)	8.3 (2.7)	8.4 (4.2)	7.3 (2.5)
$\alpha$ (1-6)	3.7 (1.4)	3.6 (1.2)	2.7 (1.2)

**Table S.8: RMSF values calculated for the two arms in simulations started from a conformation of the Gal- $\beta$ (1-4)-GlcNAc linkage corresponding to the 3DO3. Standard deviation values are indicated in parenthesis.**

<b>3DO3</b>	<i>pp</i>	<i>op</i>	<i>oo</i>
a(1-3)	6.1 (2.5)	7.8 (2.1)	6.6 (2.4)
a(1-6)	4.4 (1.9)	4.3 (1.1)	4.8 (2.2)

## References

- Harbison AM, Brosnan LP, Fenlon K, Fadda E. 2019. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology*, 29:94-103.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph*, 14:33-38, 27-38.
- Jo S, Im W. 2013. Glycan fragment database: a database of PDB-based glycan 3D structures. *Nucleic Acids Res*, 41:D470-474.
- Masuda K, Yamaguchi Y, Kato K, Takahashi N, Shimada I, Arata Y. 2000. Pairing of oligosaccharides in the Fc region of immunoglobulin G. *Febs Letters*, 473:349-357.
- Park SJ, Lee J, Qi Y, Kern NR, Lee HS, Jo S, Joung I, Joo K, Im W. 2019. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology*, 29:320-331.
- Sondermann P, Huber R, Oosthuizen V, Jacob U. 2000. The 3.2-Å crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature*, 406:267-273.
- Sugita Y, Okamoto Y. 1999. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141-151.



## Appendix IV

### **Table of Contents:**

#### **1. Materials and Methods**

##### **1.1. Computational Methods**

##### **1.2. Experimental Methods**

#### **2. Supplementary Simulations: All-Atom MD Simulations of S Protein Head using Amber FF14SB/Glycam06 Force Fields**

##### **2.1. Computational Methods**

##### **2.2. Results and Discussion**

##### **2.3. Conclusions**

#### **3. Overview of Neutralizing Antibody Epitope Accessibility**

#### **4. Supplementary Figures (S1 to S16)**

#### **5. Supplementary Tables (S1 to S12)**

#### **6. Supplementary Movies (S1 to S3)**

#### **7. Supplementary References**

#### **1. Materials and Methods**

##### **1.1. Computational Methods**

*Model systems.* Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) glycoprotein is a large, glycosylated homotrimer, where each of its three identical monomers (residues 16–1273) can be divided into three main topological domains: the “head,” comprising S1 and S2 subunits until residue 1140; the “stalk,” composed of heptad repeat 2 (HR2) and transmembrane (TM) domains (residues 1141–1234); and the cytoplasmic tail (CT) (residues 1235–1273) (**Figure 1A** main text).<sup>1</sup> Experimental structures of SARS-CoV-2 S have been resolved in two main conformational states, open and closed, that were used in this study to build two complete, fully glycosylated models, referred to in this manuscript as “Open” and “Closed”, respectively. The Closed system is based on a cryo-EM structure of the S protein solved at 2.80Å average resolution (PDB ID: 6VXX),<sup>2</sup> where all receptor binding domains (RBDs) are in the “down” conformation. The Open

system is instead built upon a cryo-EM structure of the S protein solved at 3.46-Å average resolution (PDB ID: 6VSB),<sup>1</sup> where only one RBD (chain A) is in the “up” conformation. A third system, called “Mutant”, was also generated from the Open system upon mutation of N165 and N234 into alanine within all the three monomers, which ablated the respective N-glycan sequons. Although the cryo-EM structures of the S protein already provide critical information about its structure, they are usually incomplete and/or have been modified to increase protein stability.<sup>1,3</sup> For example, the introduction of two consecutive prolines (S-2P variant) in the central helix and/or of an engineered C-terminal foldon trimerization domain<sup>4</sup> has been adopted as a common strategy to stabilize the S protein for cryo-EM.<sup>1,3</sup> In addition, highly flexible protein regions (loops) and glycans beyond the first three sugars often remain unresolved owing to resolution limits. Therefore, several modeling steps were required to produce a full-length model of the wild type protein as described below.

*Missing loops modeling.* The employed cryo-EM structures of the S protein reveal several missing gaps corresponding to flexible loops ranging from 3 to 38 residues. To generate a complete construct, missing gaps were modeled as disordered loops using Modeller9.19.<sup>5</sup> Keeping the cryoEM coordinates fixed, 50 models were independently generated for each monomer, from which the top models were selected and reassembled to recreate the full trimeric head. The alignment between the cryo-EM structure and the FASTA sequence of SARS-CoV-2 spike (QHD43416.1)<sup>6</sup> used by Modeller was generated using Clustal Omega.<sup>7</sup> The top models were further visually inspected to discard those in which loops were entangled in a knot or clashed with the rest of the structure. Finally, the stabilizing proline mutations from the cryo-EM structures were mutated back to wild type, and modeling artifacts were detected and corrected prior to simulations.

*HR2 and TM domain (stalk) modeling.* Both cryo-EM structures employed to build our models were stabilized using an engineered C-terminal foldon trimerization domain.<sup>4</sup> Therefore, the stalk region of the S protein from residues 1147–1234, including the HR2 and TM domains, had to be constructed. Using the Jpred4 server,<sup>8</sup> the secondary structure of the stalk sequence was predicted as three helical segments connected by two unstructured loops (**Figure S13**). Given the amphipathic nature of the helical segments, the three chains were assembled into an alpha-helical coiled-coil trimeric bundle using Modeller9.19.<sup>5</sup> A coiled-coil crystal structure, where the smaller and more hydrophobic residues are positioned inside the bundle and the polar residues are solventexposed, served as a template (PDB ID: 2WPQ).<sup>9</sup> Each alpha-helix was broken into three segments separated by two loops according to Jpred4 secondary structure predictions.

*CT modeling.* The CT of the S protein (residues 1235–1273) was modeled using the i-TASSER software.<sup>10–12</sup> i-TASSER generated five models with confidence scores of -1.25, -2.65, -3.15, -4.33,

and  $-1.33$  (C-score range  $[-5, 2]$ ). Out of these, model 3 was selected because it revealed a helical domain between residues 1238 and 1245, where two cysteines were shown to be palmitoylated in another betacoronavirus, MHV-A59.<sup>13,14</sup> The corresponding cysteines in SARSCoV-2 were C1240 and C1241. The remaining sequence of CT was predicted to be intrinsically disordered. Both cysteines (C1240 and C1241) were palmitoylated using lipid-tail functionality available within *Glycan Reader* in CHARMM-GUI.<sup>15,16</sup>

*Glycosylation.* SARS-CoV-2 S protein features 22 N-glycan sequons (N-X-S/T) per monomer, which have been found to be heterogeneously populated in different glycoanalytic studies.<sup>17-19</sup> Interestingly, two O-glycans have also been characterized at positions T323 and S325.<sup>19</sup> Our modeled constructs have been fully N-/O-glycosylated using the *Glycan Reader & Modeler* tool<sup>20</sup> integrated into *Glycan Reader*<sup>15</sup> in CHARMM-GUI.<sup>16</sup> An asymmetric (i.e., not specular across monomers) site-specific glycoprofile has been derived according to glycoanalytic data reported by Watanabe et al.<sup>17</sup> for N-glycans and by Shajahan et al.<sup>19</sup> for O-glycans. Detailed per-chain descriptions of the site-specific glycoprofile of the S protein systems simulated in this work are shown in **Tables S1-S3**. In the Mutant system, N165A and N234A mutations were introduced to remove the respective N-glycans. This was performed with PSFGEN during system setup. In summary, 70 glycans ( $22 \times 3_{\text{monomers}}$  N-glycans and  $2_{\text{chainA}} + 1_{\text{chainB}} + 1_{\text{chainC}}$  O-glycans) have been added in the Open and Closed systems, whereas 64 have been added in the Mutant system. Our modeled N-glycans account for oligo-mannose from  $\text{Man}_5\text{GlcNAc}_2$  to  $\text{Man}_9\text{GlcNAc}_2$ , complex and hybrid types, displaying one to four antennas. Additional modifications, such as fucosylation and sialylation, have also been site-specifically considered, as reported in Watanabe et al.<sup>17</sup> We remark that the oligosaccharides ( $\text{GlcNAc-}/\text{GlcNAc}_2\text{-}/\text{ManGlcNAc}_2\text{-}$ ) originally solved in the cryo-EM structures have been generally retained or used as a basic scaffold, when possible, to build the full glycans. However, owing to steric clashes arising at particularly buried sites (for example, N122), a manipulation of glycan dihedrals and/or asparagine side chain has been necessary to fit in all the glycans.

*Membrane modeling.* The lipid composition of the membrane patch was selected based on the lipoprofiles<sup>21,22</sup> of the endoplasmic reticulum and trans-Golgi network, organelles in which the coronavirus membranes are known to be constructed.<sup>23,24</sup> A symmetric  $225 \text{ \AA} \times 225 \text{ \AA}$  lipid bilayer patch was generated using CHARMM-GUI's input generator.<sup>16,25</sup> The lipids were packed to an area per lipid of  $70 \text{ \AA}^2$  with the following ratio of phospholipids and cholesterol: POPC (47%), POPE (20%), CHL (15%), POPI (11%), and POPS (7%). The IUPAC names corresponding to these abbreviations are given in **Table S4**. The area per lipid value was selected based on the suggested CHARMM-GUI areas; the equilibrium values for this system were calculated and can be found in **Figure S5**.

*System preparation.* Upon functionalization (i.e., glycosylation and palmitoylation) of the glycoproteins through the *Glycan Reader* module available within CHARMM-GUI, further modifications to the structures were necessary and formatting issues were manually solved. Differently from SARS-CoV, where the S1/S2 site is cleaved prior to fusion (i.e., at the host cell surface), in SARS-CoV-2 the cleavage occurs when the virus is assembled.<sup>2</sup> Therefore, all the three constructs were modeled in their cleaved form, i.e. with the furin site cleaved between residues R685 and S686. However, for convenience, S1 and S2 subunits within each protomer have been assigned to the same chain (referred to as A/B/C), following the scheme used in 6VSB,<sup>1</sup> where RBD of chain A is in the “up” conformation. Glycans were attributed segnames from G1 to G70 (G1–G64 for Mutant), as reported in **Tables S1-S3**. Protonation states were assessed using PROPKA3<sup>26</sup> at pH 7.4 in the presence and absence of glycans, without registering any critical differences. The generated models were parametrized using PSFGEN and CHARMM36 all-atom additive force fields for protein, lipids, and glycans.<sup>27,28</sup> Parameters for palmitoylated cysteine were taken from Jang et al.<sup>29</sup> Sodium and chloride ions were added to neutralize the charge of the system at 150 mM concentration and they were treated using Beglov and Roux force fields.<sup>30</sup> The systems were fully solvated with explicit water molecules described using the TIP3P model.<sup>31</sup> The total number of atoms is 1,693,017 for the Open system (size: 225 Å × 225 Å × 367 Å), 1,658,797 for the Closed system (size: 225 Å × 225 Å × 359 Å), and 1,693,069 for the Mutant system (size: 225 Å × 225 Å × 367 Å).

*Molecular dynamics (MD) simulations.* All-atom MD simulations were conducted on the Frontera computing system at the Texas Advanced Computing Center (TACC) using NAMD 2.14.<sup>32</sup> The systems were initially relaxed through a series of minimization, melting (for the membrane), and equilibration cycles. During the first cycle, the protein, glycans, lipid heads (P atom for POPC, POPI, POPE, and POPS and O3 atom for CHL), solvent, and ions were kept fixed and the systems were subjected to an initial minimization of 10000 steps using the conjugate gradient energy approach. Subsequently, to allow the lipids tails to equilibrate, the temperature was incrementally changed from 10 to 310 K for 0.5 ns at 1 fs/step (NVT ensemble). The following simulation cycle was run at 2 fs/step, 1.01325 bar, and 310 K (NPT ensemble). Next, the systems were simulated with only the protein and glycans harmonically restrained at 5 kcal/mol to allow the full environment to relax in 2500 minimization steps and 0.5-ns simulations. Finally, all the restraints were released, and the systems were equilibrated for additional 0.5 ns. From this point, the production run was started, and frames were saved every 100 ps. Production MD simulations were run in triplicates for ~1 μs for Open and Mutant and ~0.6 μs for Closed (**Table S5**). To further explore the conformational space of the RBD in the “up” conformation, additional adaptive sampling simulations were run for Open and

Mutant, which were also performed in triplicates for  $\sim 0.4 \mu\text{s}$ . Whereas velocities were randomly reinitialized, the initial coordinates were selected after principal component analysis (PCA) of RBD-A. In detail, the minimum (replica 4), mean (replica 5), and maximum (replica 6) along PC1 were identified and the corresponding frames used a starting point for adaptive sampling simulations.

All simulations were performed using periodic boundary conditions and particle-mesh Ewald<sup>33</sup> electrostatics for long-range electrostatic interactions with maximum grid spacing of 2 Å and evaluation every 3 time steps. Non-bonded van der Waals interactions and short-range electrostatic interactions were calculated with a cutoff of 12 Å and a switching distance of 10 Å. The SHAKE algorithm<sup>34</sup> was employed to fix the length of all hydrogen-containing bonds, enabling the use of 2-fs integration time steps. All simulations were performed under the NPT ensemble using a Langevin thermostat<sup>35</sup> (310 K) and a Nosé-Hoover Langevin barostat<sup>36,37</sup> (1.01325 bar) to achieve temperature and pressure control, respectively.

*Accessible surface area (ASA).* ASA was calculated using the *measure sasa* command implemented in VMD,<sup>38</sup> which is based on the Shrake and Rupley algorithm,<sup>39</sup> in combination with in-house Tcl scripts. Three separate ASA analyses were conducted by taking into account the S protein head (residues 16–1140), stalk (residues 1141–1234), and receptor binding motif (RBM) of the RBD (residues 400–508), respectively. The area covered by glycans (i.e., the glycan shield) was obtained after the subtraction of the ASA of the considered domain in the absence of glycans with the ASA in the presence of glycans. This value was calculated along the trajectory with a stride of 150, 20, and 20 frames between each assessment for head, stalk, and RBM-A, respectively. For each system (Open and Closed), the values were averaged across all the respective replicas and standard deviation was computed. Apart from the standard 1.4-Å probe, this analysis was repeated for 14 different (1-Å-interspersed) values of probe radius (from 2 to 15 Å). Note that additional ASA analyses on the whole RBD-A (residues 330–530) and on its noninteracting region (residues 330–399 and 509–530) were also analogously performed (see SI). Similarly, epitope-specific ASA analyses were conducted on the chain A of Open and Closed systems using only 7.2 and 18.6 Å as probe radii. ASA evaluations were conducted with a stride of 20 frames. The residues considered for each epitope are listed in **Table S11**.

*Principal Component Analysis (PCA).* PCA was performed using the *sklearn.decomposition.PCA* function in the *Scikit-learn* library using python3.6.9.<sup>40</sup> First, all simulations were aligned with *mdtraj*<sup>41</sup> onto the same initial coordinates using C $\alpha$  atoms of chainA central scaffold (residues 747–783 and 920–1032). Next, simulation coordinates of RBD-A (residues 330–530) from all systems (Open, Mutant, and Closed) and replicas were concatenated and used to fit the transformation function. Subsequently, the fitted transformation function was applied to reduce the dimensionality of

each system simulation RBD-A C $\alpha$  coordinates. Subsequently, the fitted transformation function was applied to reduce the dimensionality of RBD-A coordinates from each system into the PC space. Note that it is important that the comparative PCA across systems have consistent eigen-basis of the principal components. This was ensured by transforming all systems coordinates into the same PC space. The same procedure was performed for PCA of RBD-B and RBD-C.

*Angles calculation.* The lateral angle and axial angle were calculated using in-house Tcl scripts along with VMD.<sup>38</sup> The axial angle is defined by three points corresponding to (i) the center of mass (COM) of RBD-A  $\beta$ -sheets (residues 394–403, 507–517, and 432–437), (ii - vertex) the COM of the central helices (residues 987–1032), and (iii) the COM of the top section of the central helices (residues 987–993). The lateral angle is described by three points corresponding to the (i) COM of RBD-A at frame 0, (ii - vertex) COM of the top section of the central helices, and (iii) COM of RBD-A  $\beta$ -sheets at frame  $n$ . We note that when calculating the COM of RBD-A we only considered the core residues defining the  $\beta$ -sheets of the domain, i.e. the most stable part of the RBD, discarding instead the highly flexible, solvent exposed loops (like the receptor binding motif) or the hinges that would have altered the position of the COM, thus biasing the angle calculation. In this way it was possible to keep track of the actual core motions of RBD-A. The other COMs were calculated on the central helices (CH). The CH are three alpha-helices (one for each monomer) located around the central axis of the spike trimer, representing the most rigid backbone of the spike's head as shown in **Figure 1** of the main text.

Both angles were evaluated at each frame along the trajectories as a variation (positive or negative) with respect to their initial value. The trajectories were aligned by the S protein central scaffold (residues 747–783 and 920–1032) including the central helices using the coordinates at frame 0 as a reference. Importantly, whereas the axial angle was calculated in a three-dimensional space defined by  $xyz$  coordinates, the lateral angle was assessed by considering the projection of the COMs onto a two-dimensional space defined only by  $xy$  coordinates. In this way, the lateral angle only accounts for lateral tilt/shift of the RBD, discarding any other motion along  $z$ .

*Hydrogen bonds calculation.* Hydrogen bonds were calculated using the *measure hbonds* command implemented in VMD<sup>38</sup> in combination with in-house Tcl scripts. Hydrogen bonds criteria were set as 3.5 Å for distance between heavy atoms and as 45° for angle between AccDon-Hyd. All frames across all replicas were considered for this analysis. Occupancy (%) was determined by counting the number of frames in which a specific hydrogen bond was formed with respect to the total number of frames.

*Root-mean-square-deviation (RMSD)*. RMSD of protein C $\alpha$  atoms was computed using the *measure rmsd* command implemented in VMD<sup>38</sup> in combination with in-house Tcl scripts. Different alignments were done before RMSD calculations using the initial coordinates of C $\alpha$  atoms as a reference. In particular, for RBD-A RMSDs, C $\alpha$  atoms of the S protein central scaffold (residues 747–783 and 920–1032) were used as a reference for alignment, whereas for the head, stalk, and CT RMSDs, the trajectories were aligned onto the C $\alpha$  atoms of the residues of the respective regions.

*Root-mean-square-fluctuations (RMSF)*. RMSF was calculated using in-house python scripts along with *mdtraj*.<sup>41</sup> RMSF was computed for each glycan of every chain across all replicas in Closed, Open, and Mutant. The trajectories were aligned onto the initial coordinates using the C $\alpha$  atoms of the entire protein as a reference.

## 1.2 Experimental Methods

*Protein expression and purification*. The spike S2P variant was expressed using a previously described mammalian expression vector containing proline substitutions at residues 986 and 987 and C-terminal 8xHis and TwinStrep tags.<sup>1</sup> Alanine substitutions were introduced into S2P to yield the N165A and N234A spike variants. To generate the disulfide-locked spike with all RBDs down, we introduced S383C, D985C substitutions into a previously described stabilized spike construct (HexaPro) containing proline substitutions at positions 817, 892, 899, 942, 966 and 987.<sup>42</sup> ACE2 was expressed using a mammalian expression vector encoding residues 1-615 of human ACE2, a C-terminal HRV 3C cleavage site, a mono-Fc tag, and 8xHis. Plasmids were transiently transfected into FreeStyle 293-F cells using polyethylenimine and cultured for 4 days (for spike variants) or 6 days (for ACE2). Spike variants were purified by passing filtered cell supernatant over StrepTactin resin, and ACE2 was purified using Protein A agarose. ACE2 was subsequently cleaved with 3C protease passed over Protein A to remove the Fc-8xHis fragment.

*Biolayer interferometry*. Anti-foldon IgG was immobilized to an anti-human Fc (AHC) Octet biosensor (FortéBio), which was subsequently dipped into the specified spike ectodomain variant for loading. The biosensor was then dipped into 200 nM ACE2 to measure the RBD-spike association signal before being transferred to a well containing buffer only (10 mM HEPES pH 7.5, 150 mM NaCl, 3 mM EDTA, 0.05% Tween 20 and 1 mg/mL bovine serum albumin) to measure the dissociation signal. The total response at the end of the association phase was recorded for each variant and used to quantify the relative proportions of accessible RBDs. The total response at the end of the association phase, where nearly all the RBDs should be saturated, was recorded for each variant and used to quantify the relative proportions of accessible RBDs. Three independent experiments were run for each of the S2P, N165A, and N234A spike variants. Collected data shown

in **Figure 4B** are reported as mean  $\pm$  standard deviation, with the following values: 0.2563 nm  $\pm$  0.0070 nm (S2P); 0.2284 nm  $\pm$  0.0051 nm (N165A); 0.1623 nm  $\pm$  0.0097 nm (N234A). The experiment with the (HexaPro) variant with all the three RBDs locked in the closed conformation was used as control. Raw data are made available as supporting information.

## **2. Supplementary Simulations: All-Atom MD Simulations of S Protein Head using Amber FF14SB/Glycam06 Force Fields**

### **2.1. Computational Methods**

The model of the SARS CoV2 S protein head was built in the open conformation by homology with SWISS MODEL<sup>43</sup> using the cryo-EM structure of the spike trimer in the open state as a template (PDB ID: 6VYB,<sup>44</sup> 3.2 Å resolution) and NCBI YP\_009724390.1 as reference sequence. We note that this structure (6VYB) differs from the one used as template for building the fulllength, Open and Mutant models described in the main text (6VSB). In 6VYB, the RBD “up” belongs to chain B (corresponding to chain A in 6VSB), whereas chains C and A corresponds to chains B and C of 6VSB, respectively. The missing loops in the 6VYB cryo-EM structure were built automatically by SWISS MODEL based on structural libraries of backbone fragments from the PDB with similar sequences. The resulting protein structure exhibits 18 glycosylation sites per protomer, for a total of 54 sites per trimer that can be occupied. Glycosylation on these sites was built by aligning equilibrated structures of complex fucosylated (FA2B) and non-fucosylated (A2B) and of oligomannose (Man5 and Man9) from our in-house database<sup>45</sup> to the resolved GlcNAc residues in the cryo-EM structure. Because not all the GlcNAc residues were resolved in the cryo-EM and because two glycosylation sites per protomer are located in loops also not resolved in the cryo-EM structure, we have built the final 54-glycans model in two phases. In the first phase we built models with 46 glycosylation sites and run a 20 ns equilibration to obtain conformations of the rebuilt loops that allowed for the linking of our glycans. The chosen structures were then completed with glycosylation in all 54 sites, leading to the 54-glycans model.

In this additional set of simulations of the SARS-CoV-2 S protein head, we have considered three slightly different glycosylation profiles, shown in **Table S13**, resulting into three models that differs specifically at position N234, occupied either by a Man9 (Man9-N234) or where N234 is mutated into Ala (N234A) leading to glycan depletion, or where N234 is non-glycosylated. We ran two independent trajectories for the 46-glycans model (i.e., Man9-N234) and one for the related N234A mutant. Moreover, we performed one run each for the 54(53)-glycans models (i.e., Man9-N234, the



N234A mutant, and non-glycosylated N234) for a total of 6 independent MD runs (see **Table S14**). We remark that the N234 N-linked site monitored in this set of simulation belongs to the NTD of chain C (NTD-C), corresponding to NTD-B of the full-length model described in the main text. In these setups of the spike head based on 6VYB and in the respective MD simulations, the protein was described using AMBER ff14SB parameters,<sup>46</sup> whereas the glycans by the GLYCAM06j-1 version of GLYCAM06.<sup>47</sup> An atmosphere of 200 mM NaCl was also included in all simulations, with ions represented by parameters in the AMBER ff14SB set. Water molecules were represented by the TIP3P model.<sup>31</sup> All simulations were run with v18 of the AMBER software package.<sup>48</sup>

The same system preparation and running protocol was used for all MD simulations. The energy of the system built as described above was minimized in two steps of 50,000 cycles of the steepest descent algorithm. During the first minimization all the heavy atoms were kept harmonically restrained using a potential weight of  $5 \text{ kcal mol}^{-1} \text{ \AA}^2$ , while the solvent, counterions and hydrogen atoms were left unrestrained. During the second minimization step only the protein heavy atoms were kept restrained, while the glycans, solvent, counterions and hydrogens were left unrestrained. After energy minimization the system was equilibrated in the NVT ensemble with the same restraint scheme, where heating was performed in two stages over a total time of 1 ns, from 0 to 100 K (stage 1) and then 100 to 300 K (stage 2). During equilibration the SHAKE algorithm was used to constrain all bonds to hydrogen atoms. The Van der Waals and direct electrostatic interactions were truncated at 11 Å and Particle Mesh Ewald (PME) was used to treat long range electrostatics with B-spline interpolation of order 4. Langevin dynamics with collision frequency of  $1.0 \text{ ps}^{-1}$  was used to control temperature, which a pseudo-random variable seed to ensure there are no synchronization artefacts. Once the system was brought to 300 K an equilibration phase in the NPT ensemble of 1 ns was used to set the pressure to 1 atm. The pressure was held constant with isotropic pressure scaling and a pressure relaxation time of 2.0 ps. At this point all restraints on the protein heavy atoms were removed, allowing the system to evolve for 15 ns of conformational equilibration before production. The total simulation times, including equilibration, are shown in **Table S14**.

## 2.2 Results and Discussion

As remarked in the main text, to further assess possible impact of the force fields and/or to the starting cryo-EM structure on the simulations described in the main text, we performed an additional set of simulations of the open SARS-CoV-2 S protein's head (presented here) using AMBER ff14SB/GLYCAM06j-1 force fields<sup>46,47</sup> and an alternative initial cryo-EM structure (PDB ID: 6VYB),<sup>44</sup> which presents the N234 GlcNAc in a slightly different orientation. The results of this set of simulations are described below.

**Man9-N234 model.** We analyzed the relative stability of the RBD domains in chains A, B and C during the MD production by calculating the backbone RMSD values relative to the starting homology model based on the 6VYB template. RMSD values were calculated using different alignments, namely residues 770 to 1255 of chain A, B and C, respectively, to evaluate potential biases. The differences in average RMSD values calculated from different alignments are within the standard deviation values used as error bars, therefore only results from the alignment to chain A (residues 770 to 1255) are shown in **Table S15**. Data were collected after 20 ns of equilibration. Unless otherwise stated, results are shown for the 54(53)-glycans model systems. All simulations of the Man9-N234 model show that the glycan gradually inserts itself in the space left empty by the lifting-up of RBD-B (chain B in this model, corresponding to chain A in the full-length model) (**Movie S2**). This insertion progresses gradually through the formation of hydrogen bonding interactions between Man9 glycan at N234 and Y369 and N370 of RBD-C that evolve to reach the core of the spike's trimer defined by the location of D405, R408 and E409 of RBD-A, located in the diametrically opposite side across the spike apical center (see **Figure S14**). Similar interactions have been registered also in the simulations of the full-length model of the spike described in the main text, where Man9 at N234 is found to establish persistent h-bonds with residues of RBD-C (RBD-A here). The insertion of Man9 into the open pocket and its stable interactions with the trimer core's charged residues D405, R409 and E409 results in a stabilization of the whole structure, which is evident from the backbone RMSD analysis of the three RBD domains, shown in **Figure S15**. As an interesting note, position N357 in SARS-CoV S, corresponding to position N370 in SARSCoV2 S, is part of an NST sequon and it is glycosylated.<sup>49</sup> The sequon is lost in SARS-CoV2 S with a mutation of the Thr to Ala. Because of the important and stable interactions of the Man9 at N234 with N370 along the reaction coordinate that allows it to reach the core of the trimer, it is reasonable to infer that glycosylation at position N370 in SARS-CoV would interfere with this process, potentially affecting or preventing the insertion of the Man9. To understand the significance of the presence of a large glycan such as Man9 at position N234 of NTD-C and its role in filling the empty space left by the opening of the RBD-B, we decided to remove it and to observe the conformational changes through another set of conventional MD simulations.

**N234A mutant and N234-nogly models.** To understand the role of the Man9 at N234 within NTD-C, we designed two models: a model with an N234A mutation in chain C (N234A) and a model where the N234 position also in chain C is non-glycosylated (N234-nogly). For those we analyzed four independent MD trajectories (see **Table S14**), two from the 46(45)-glycans model, one from the 54(53)-glycans model sites in the N234A mutant form, and one from the 54(53)glycans model with N234-nogly. In agreement with the full-length model of SARS-CoV2 S protein described in the main text, we observed a higher degree of dynamics of the open RBD relative to Man9-N234 when the glycan at N234 is missing. As an important note, here the N234A mutant and N234-nogly are missing

Man9 only in one site of the trimer, i.e. chain C (chain B in the full-length model), which is a subtler modification relative to the full system in which the mutant is missing the glycans at N234 and N165 in each protomer. As indicated by the average backbone RMSD values shown in **Table S15**, except for the case of the 54(53)-glycans N234A model, for which the open RBD remains in a stable conformation for the length of the trajectory considered here, in the two simulations of the 46(45)-glycans N234A models and in the simulation of the 54(53)-glycans N234-nogly model, the open RBD domain is largely displaced relative to the original homology model, used as reference structure. In agreement with the full model discussed in the main text, the dynamics of the open RBD in the mutants is quite complex and within this simulation framework cannot (and should not) be defined as part of any specific reaction coordinate, such as a domain closing or unfolding. Nevertheless, in the simulations of the 46(45)-glycans N234A model, the open RBD (chain B) can be described as shifting towards the RBD of chain C, with the flexible loop interacting with the Man5 at N343 within chain C. Meanwhile, the conformational change we observed for the open RBD in the simulations of the 54(53)-glycans N234-nogly model corresponds more to a shift away from the RBD of chain C (**Figure S16**).

### 2.3. Conclusions

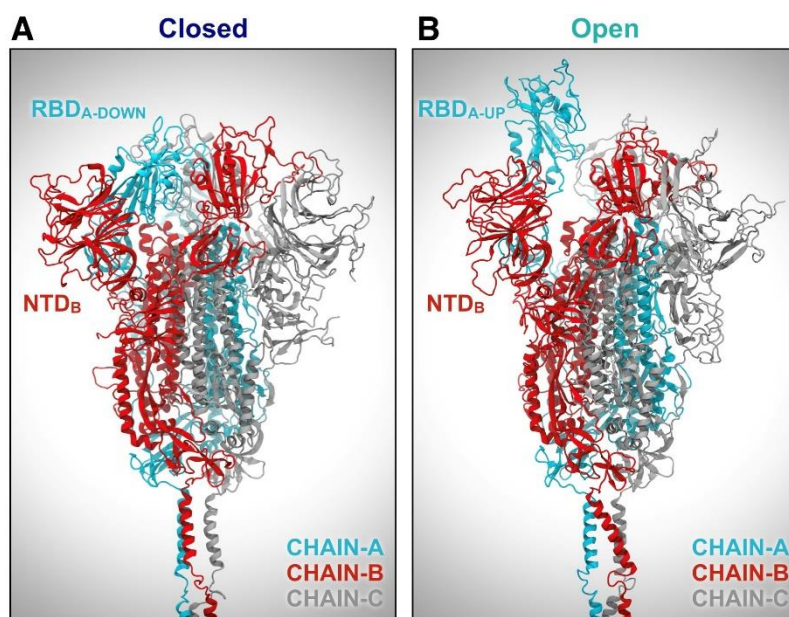
The simulations described in this section of the Supporting Information are presented here as supplementary material in support of the results discussed in the main manuscript based on much larger and complete 3D models of the SARS-CoV2 S glycoprotein. Indeed, despite the differences in the systems sizes, setups, original cryo-EM structure (6VYB vs. 6VSB), force field parameter sets (AMBER vs. CHARMM), MD software packages (Amber vs. NAMD), in the running protocols and details in the models' glycosylation, these two sets of simulations converge in showing that the absence of glycosylation at position N234 and at both positions N234 and N165, causes the open RBD to explore a larger conformational freedom, which may indicate a degree of instability. Therefore, all the simulations combined support the conclusion that the strangely "patchy" glycan shield of the SARS-CoV2 S glycoprotein may very well be engineered to play a structural role in supporting the active structure of the protein, by stabilizing the open or "up" conformation of the RBD. Also, as an important note, a mutation found in SARS-CoV2 S relative to the SARS-CoV S causes the loss of glycosylation at N370. Because of the dynamic process that Man9 follows in accessing the trimer core, which involves stable interactions with N370, a glycan at that location may interfere with insertion of the glycan at N234 into the pocket, which would remain empty, potentially making SARS-CoV S less stable in its open conformation than SARSCoV2 S.

### 3. Overview of Neutralizing Antibody Epitope Accessibility

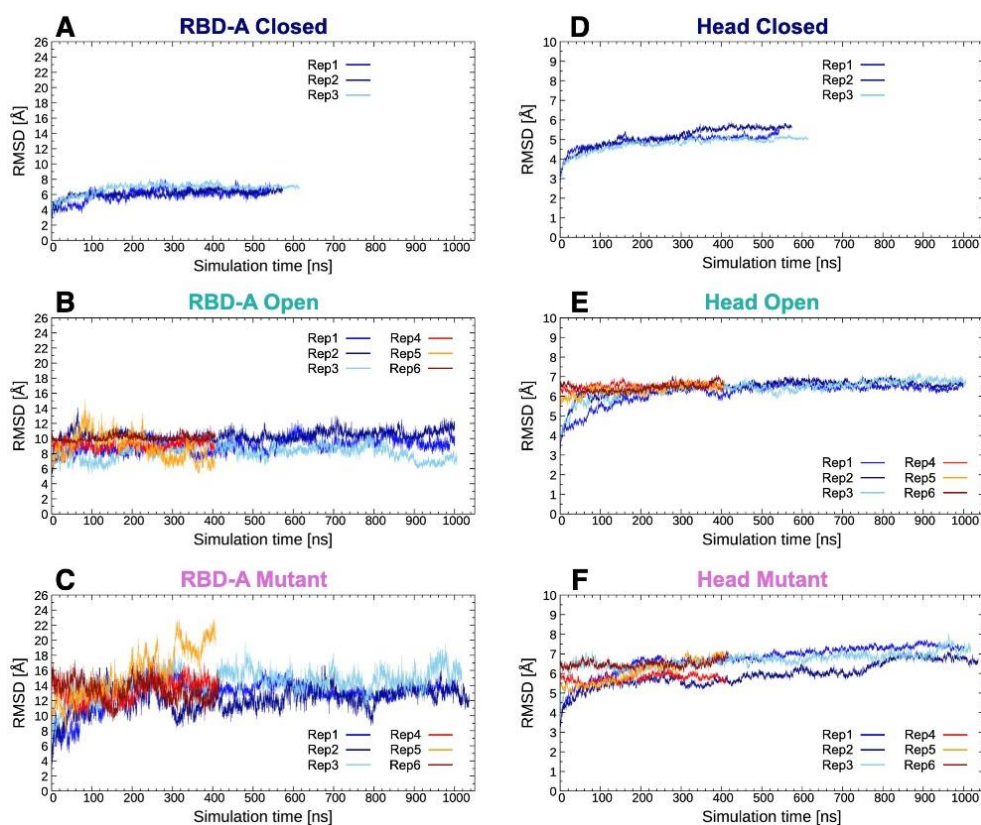
Several SARS-CoV-2 antibodies targeting the S protein have been identified (**Table S11**).<sup>50–61</sup> The majority of these antibodies recognize epitopes on the RBD, while only a few have been shown to address the antigenic regions within the NTD and CD (**Figure S12**). Among the RBD antibodies, B38 interacts with the RBM at the RBD/ACE2 interface,<sup>52</sup> whereas S309 and CR3022 target the side/bottom part of the RBD.<sup>50,51,53</sup> In addition, 4A8 and 1A9 have been found to engage with the NTD and CD, respectively.<sup>55,56</sup> To quantify the effects of glycan shielding on these epitopes, we calculated each epitope's ASA at two probe radii, 7.2 and 18.6 Å, which approximate the size of antibody hypervariable loop and variable fragments domains, respectively (**Figures S12A and S12B**, full data provided in **Table S12**).<sup>62</sup> In Open (RBD “up”), B38 epitope on the RBD/ACE-2 interface shows large ASA that is minimally shielded by glycans (10%/11%, for 7.2 and 18.6 Å probes, respectively) (**Figure S12A**). Antibodies in this region exploit the vulnerability of the S protein when RBD is in the “up” conformation. Conversely, in Closed, the shielding of B38 epitope remarkably increases to 47%/62% (**Figure S12B**). When the RBD is in the “down” conformation, the RBM is buried by the other two neighboring RBDs, which already reduce its overall accessibility by ~40%. These values are in agreement with the RBM ASA trends shown in **Figure 7** in the main text.

The S309 epitope, located on the side of the RBD and near the N-glycan at N343, shows an interesting behavior. When including glycan N343 as a shielding factor, the epitope is covered up to 45%/56% of its total area. However, this glycan has been shown to be incorporated into the recognized epitope, which would considerably increase the antigenic region targeted by S309.<sup>53</sup> Interestingly, no substantial differences in shielding are observed between Open and Closed because this epitope is mostly located on the RBD side, which remains exposed even in the “down” conformation. Considering the bottom part of the RBD, the epitope recognized by the CR3022 antibody is found to be almost completely shielded in Open (69%/94%) and not accessible at all in Closed. This is in agreement with structural data showing that the cryptic epitope engaged by CR3022 is only available when the RBD is both “up” and rotated.<sup>50,51</sup> Remarkably, this epitope partially overlaps with VHH72, an antibody found to neutralize SARS-CoV-2 S pseudotyped viruses.<sup>60</sup> Finally, the 4A8 and 1A9 epitopes located within the NTD and CD, respectively, are not affected by the conformational changes of the RBD.<sup>55,56</sup> Whereas the epitope recognized by 4A8 is about 36%/51% shielded by glycans, the one targeted by 1A9 is almost completely covered at 86%/99%. These results probe further questions on 4A8 and 1A9 binding mode in the presence of glycans.

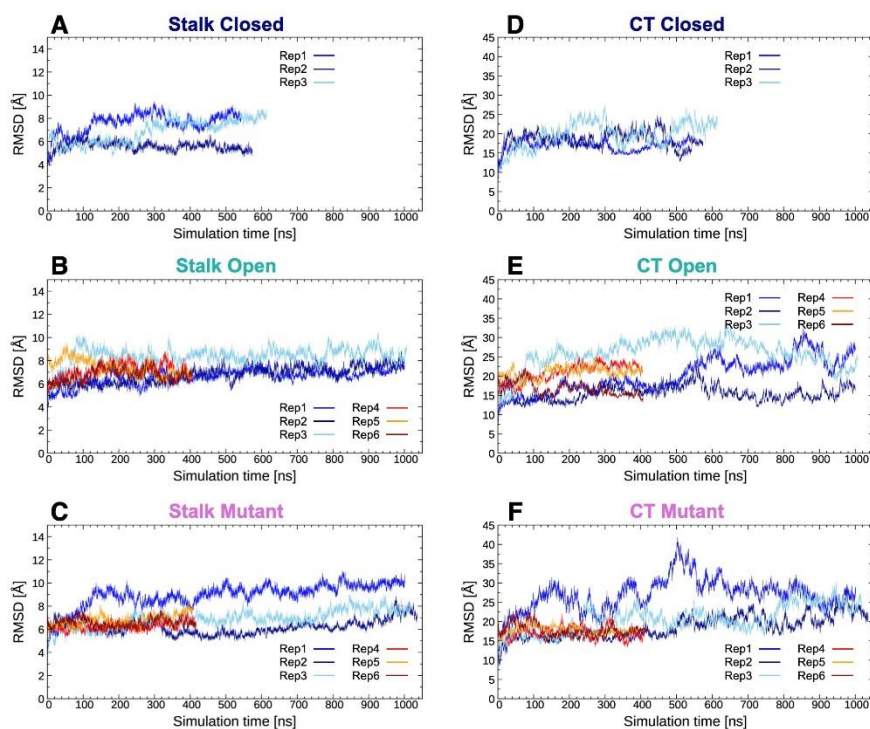
## 4. Supplementary Figures



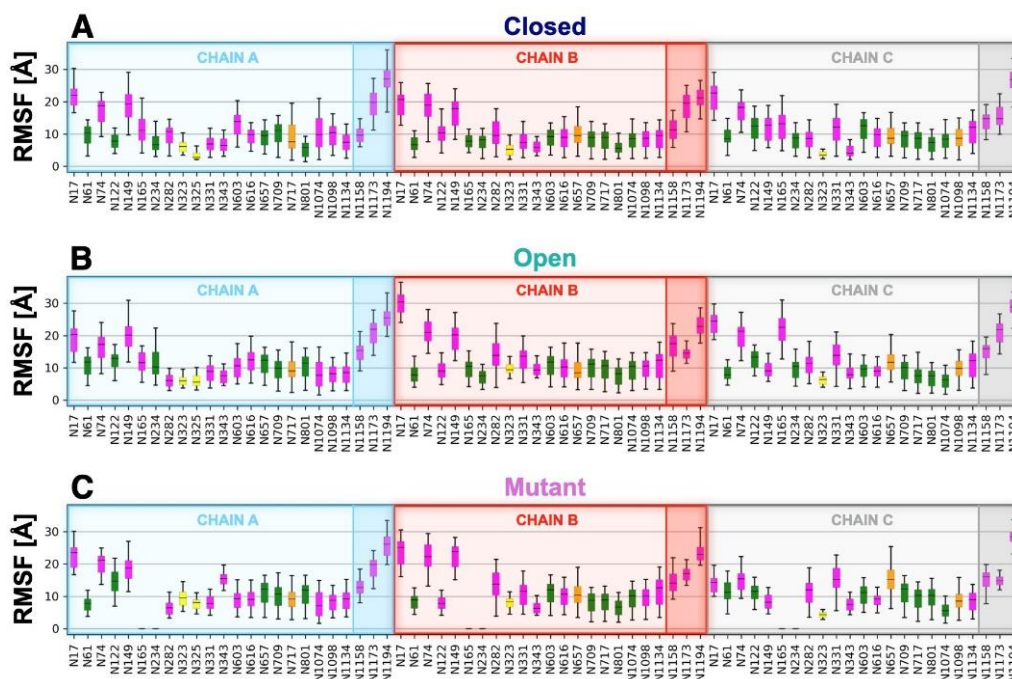
**Figure S1.** Molecular representation of S protein ectodomain in the Closed (A) and Open (B) systems. Protein is shown with cartoons, where chains A-B-C are colored in cyan, red and silver, respectively. Glycans are omitted for clarity.



**Figure S2.** RMSD [Å] vs. time [ns] plots of the receptor binding domain of chain A (A-C) (RBD-A; residues 330-530) and the head (D-F) (residues 16-1140) of the S protein  $\alpha$  atoms in the Closed, Open, and Mutant systems along each replica. For the RBD-A RMSDs,  $\alpha$  atoms of the S protein central scaffold (residue 920 to 1032, 747 to 783) was used for alignment, while for head RMSDs the S protein was aligned onto all the head  $\alpha$  atoms.

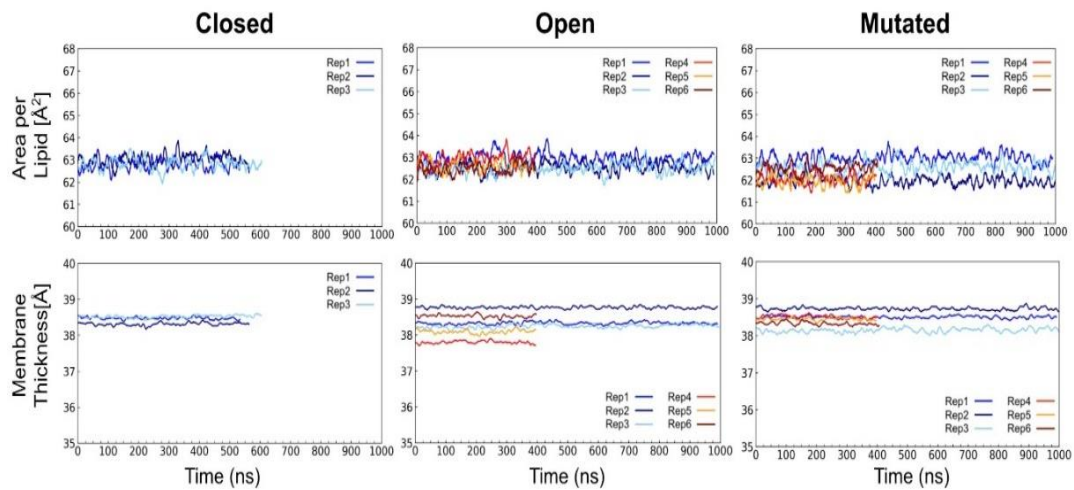


**Figure S3.** RMSD [ $\text{\AA}$ ] vs. time [ns] plots of the stalk (A-C) (residues 1141-1234) and cytoplasmic tail (D-F) (residues 1235-1273)  $\text{C}\alpha$  atoms of the S protein in the Closed, Open, and Mutant systems along each replica. For the stalk RMSDs, the stalk  $\text{C}\alpha$  atoms were used for alignment. Similarly, the cytoplasmic tail was aligned onto its  $\text{C}\alpha$  atoms prior generation of respective RMSDs.

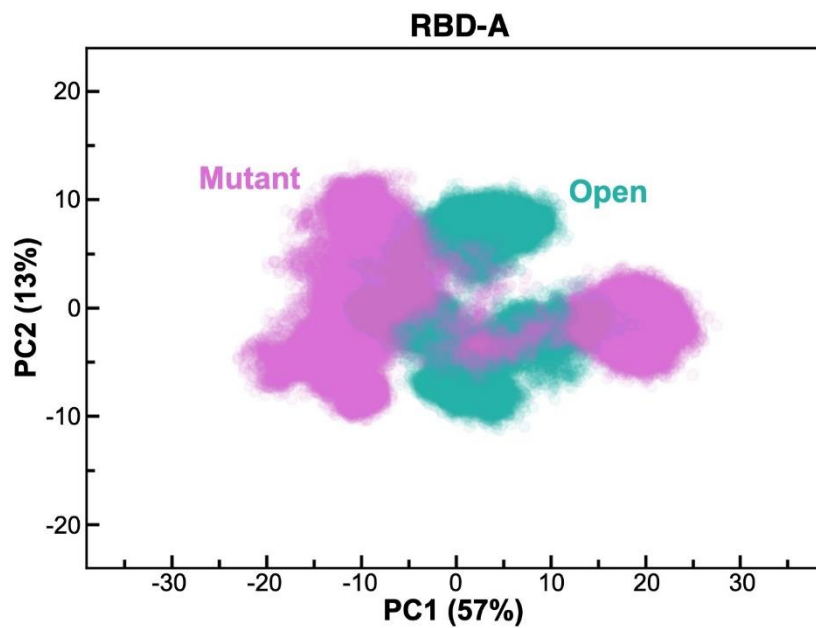


**Figure S4.** RMSF [ $\text{\AA}$ ] of each glycan for all chains across all simulations in the Closed (A), Open (B), and Mutant (C) systems. Glycans are colored based on their structure and composition: complex glycans in magenta, oligomannose glycans in green, hybrid glycans in orange, and O-glycans in yellow. Glycans in each chain are decomposed by domain: head (N17 to N1134), and stalk (N1158 to N1194), highlighted with different level of background opacity.

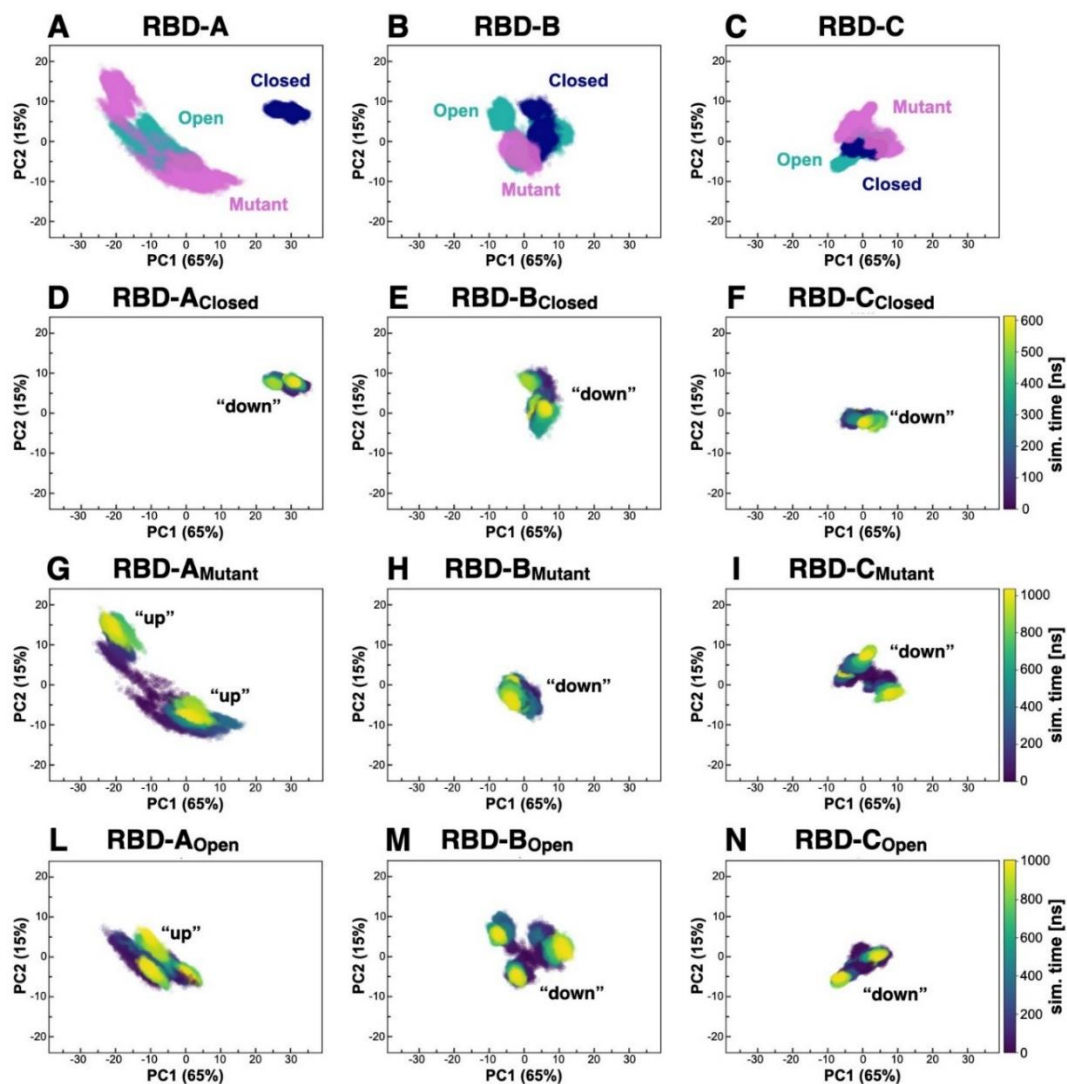




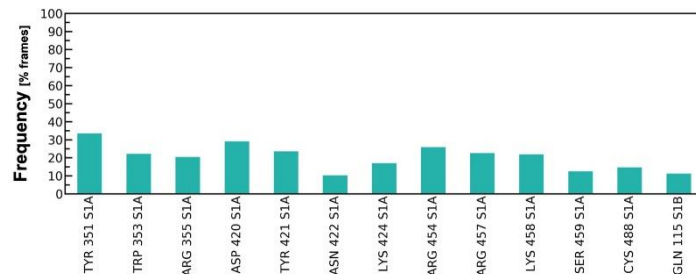
**Figure S5.** Plots of equilibrium area per lipid (top row) and P-P distance indicating membrane thickness (bottom row) of the membranes for the Closed (left), Open (center), and Mutant (right) systems along with each replicate.



**Figure S6.** PCA plot showing PC1 vs PC2 of RBD-A (residues 330-530) in Open and Mutant in teal and magenta, respectively. The amount (%) of variance accounted by each PC is shown between parentheses.

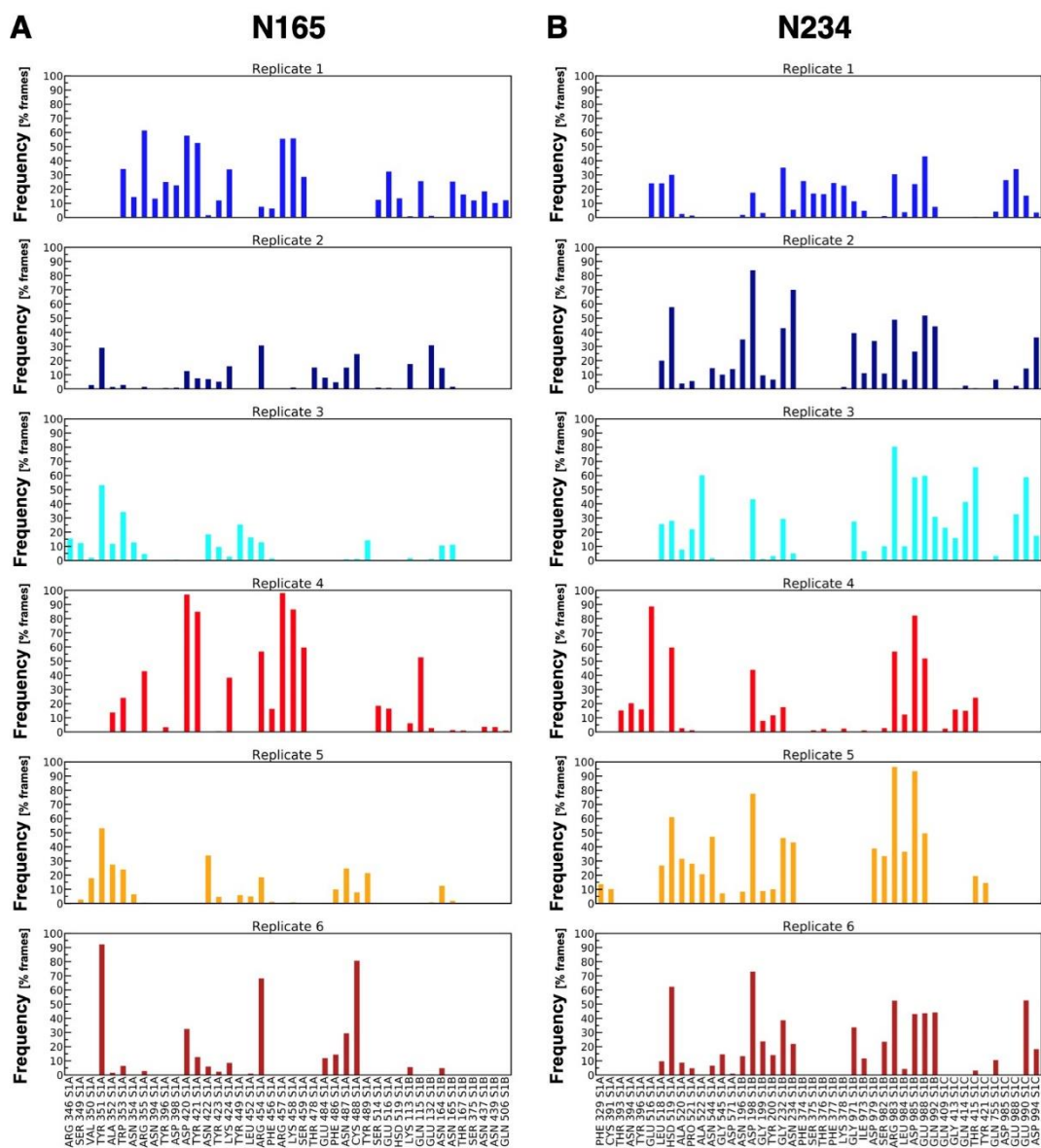


**Figure S7.** (A-C) PCA plots showing PC1 vs PC2 of RBD-A, RBD-B, and RBD-C (residues 330-530) in Closed, Open, and Mutant in blue, teal, and magenta, respectively. (D-N) PCA plots showing each system's time evolution for each replica. The time scale is reported as a color bar for each system. We remark that for Open and Mutant, replicas 4, 5, and 6 were simulated for ~400 ns. The RBD state is annotated within each respective plot as "up" or "down." The amount (%) of variance accounted by each PC is shown between parentheses.

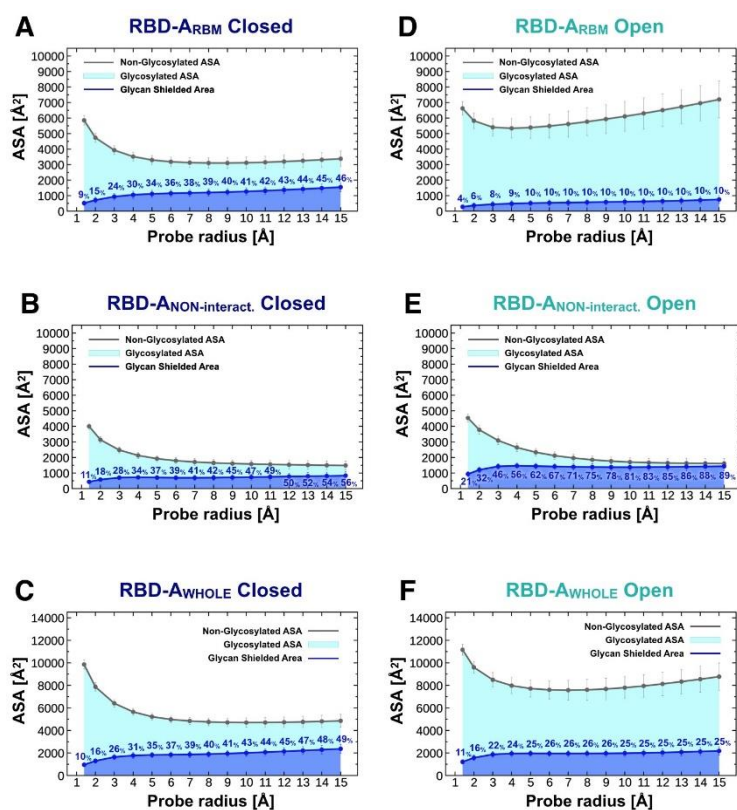


**Figure S8.** The main hydrogen bond interactions of N-glycans at N165 within the Open system are shown as occupancy across all replicas (% frames).

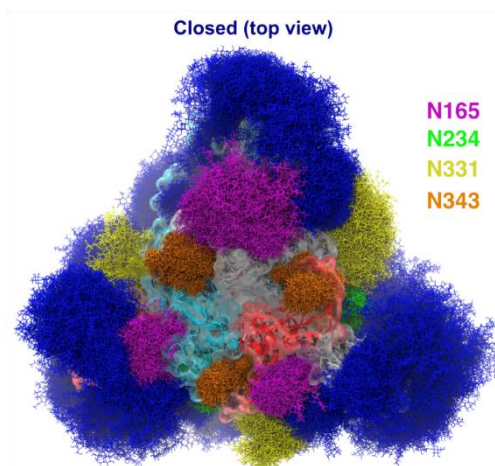




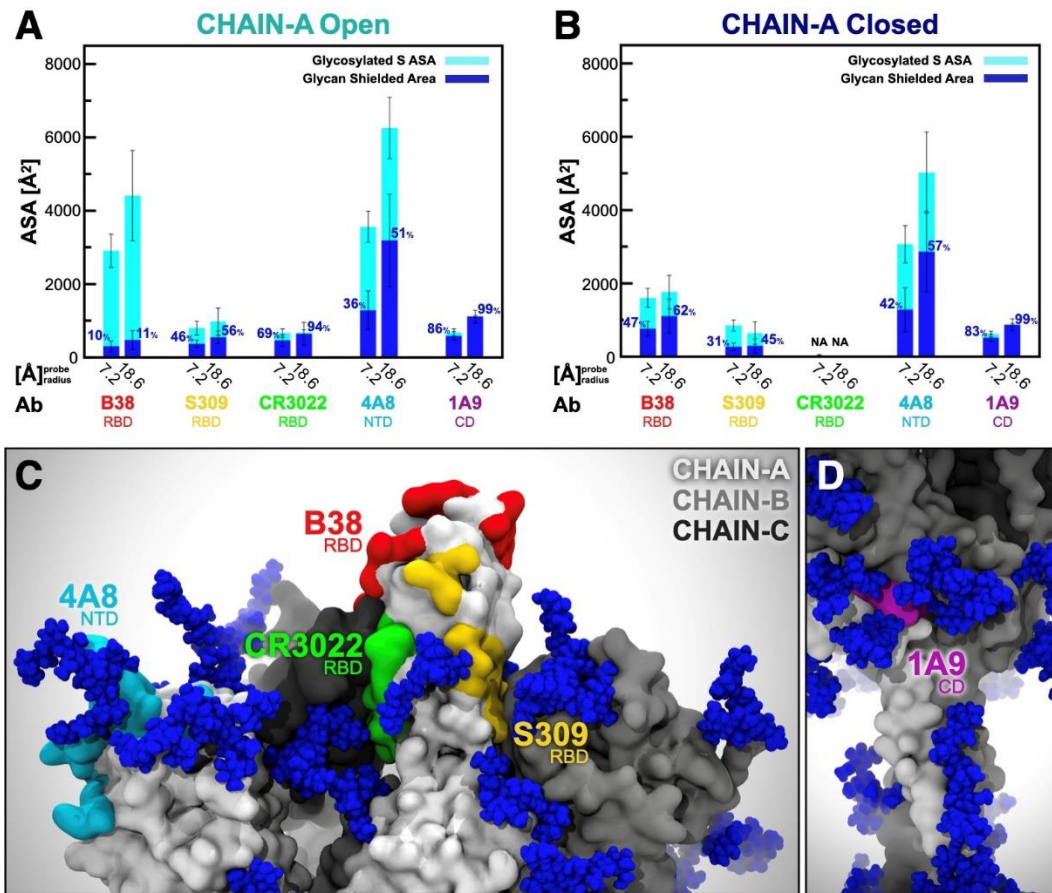
**Figure S9.** Main hydrogen bond interactions of glycan N165 (A) and N234 (B) within each replica of the Open system are shown as occupancy ( $y$ , % frames). Residues interacting with the glycans are reported on the  $x$  axis.



**Figure S10.** The accessible surface area of the RBD-A<sub>RBM</sub> (residues 400 to 509), RBD-A<sub>NON-INTERACTING</sub> (residues 330 to 399 and 509 to 530), RBD-A<sub>WHOLE</sub> (residues 330 to 530) and the area shielded by neighboring glycans in the Closed (A-C, respectively) and Open (D-F, respectively) systems is plotted at multiple probe radii from 1.4 Å (water molecule) to 15 Å. The values have been averaged across replicates and are reported with standard deviation. In blue is the area covered by the glycans (rounded % are reported), while the grey line is the accessible area in the absence of glycans. Highlighted in cyan is the area that remains accessible in the presence of glycans, which is also graphically depicted on the structure in the panels located below the plots.



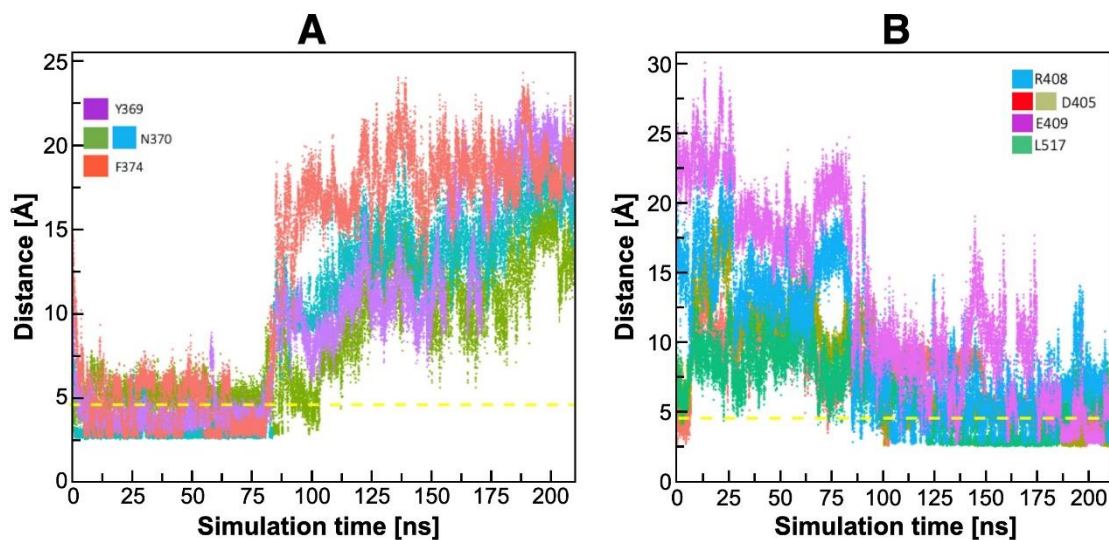
**Figure S11.** Molecular representation of the Closed system from top view. Glycans (blue lines) are represented at several frames equally interspersed along the trajectories (300 frames along 0.55 ns for Closed and 1.0 us for Open), while RBD-A is depicted with cyan cartoons and transparent surface. Chain B and Chain C are shown in red and grey cartoons, respectively, and transparent surface. Glycans at N165, N234, N331 and N343 are colored in purple, green, yellow, and orange, respectively.



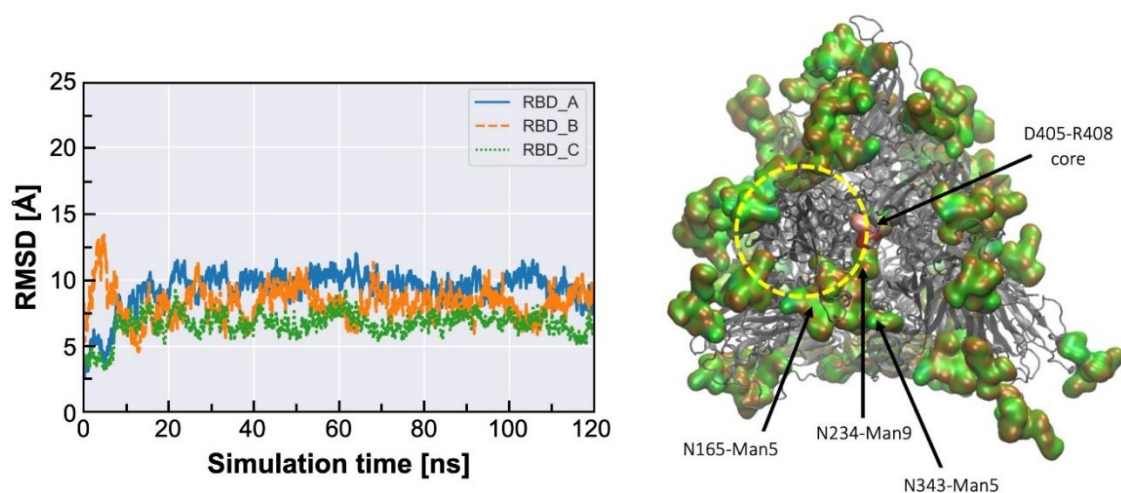
**Figure S12. Accessibility of neutralizing antibody epitopes.** The accessible surface area of antibody epitopes and the area shielded by neighboring glycans are plotted at probe radii 7.2 and 18.6 Å for chain A in Open (A) and Closed (B) systems. The area of the protein covered by the glycans is depicted in blue (rounded % values are reported), whereas highlighted in cyan is the epitope area that remains accessible in the presence of glycans. The values have been averaged across replicas and are reported with standard deviation. (C) Side view of the top region of S protein in the Open system, where the antibody epitopes are highlighted following the color scheme used in panels A and B. (D) Magnified view of 1A9 epitope (highlighted in purple) located within CD.



**Figure S13. Secondary structure of stalk and cytosolic tail regions of the S protein as predicted with Jpred4 server.** The secondary structure of the stalk sequence was predicted as three helical segments (“H”, red) connected by two unstructured loops.

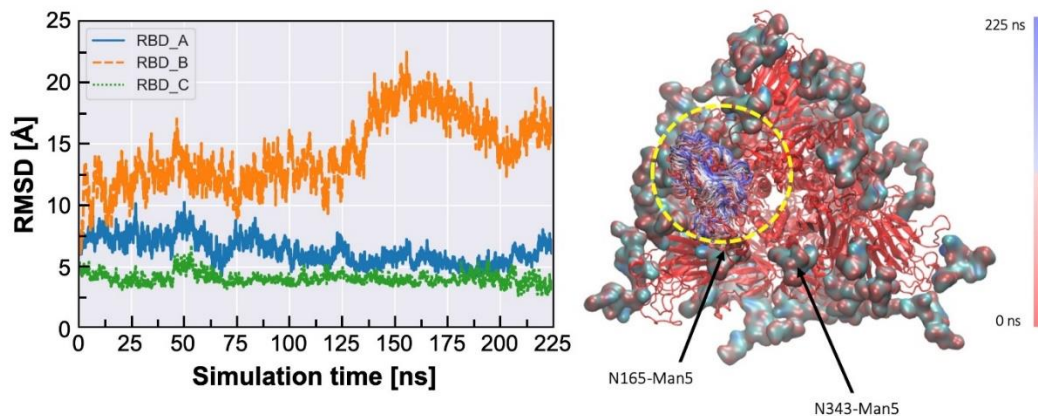


**Figure S14.** (A-B) Hydrogen bonding distances (Å) calculated through the 210 ns trajectories of the 46-glycans Man9N234 spike chosen here for the higher sampling time. (A) Interactions with Y369, N370, F374 belonging to RBD-C. (B) Interactions with R408, D405, E409, L517 belonging to RBD-A. Interactions with F374 and L517 involve backbone atoms and double colors indicates distances between different atoms on the same residues. Similar hydrogen bonding pattern and behavior have been observed in all other simulations of the Man9-N234 systems.



**Figure S15** *Left panel:* Backbone RMSD values calculated for the 54-glycans Man9-N234 system over 120 ns of production. A structural alignment of the protein backbone atoms was done on the stalk residues (resid 770 to 1255) of chain A, see text for more details. *Right panel:* Snapshot of the MD simulation (at 44 ns) showing the insertion of the Man9 at N234 deep into the trimer core highlighted by the residues in red (D405) and white (R408) with labels indicating other important glycans in framing the open conformation of the “up” RBD, highlighted within the yellow circle. All glycans are rendered as Quick Surface (green C atoms) and protein as New Cartoon (all grey). Rendering done with VMD and graphs with *seaborn.pydata.org*.





**Figure S16.** *Left panel:* Backbone RMSD values calculated for the 54-glycans N234-nogly system over 225 ns of production. A structural alignment of the protein backbone atoms was done on the stalk residues (resid 770 to 1255) of chain A, see text for more details. *Right panel:* Snapshot from the MD simulation showing a static representation of the protein (red New Cartoon) and the change in the relative position of the open RBD domain (resid 437 to 508) along the trajectory. The coloring indicates the trajectory progression as indicated in the bar on the right-hand side. The yellow circle highlights the position of the open RBD domain. All glycans are shown with Quick Surface (with C atoms in cyan). Rendering done with VMD and graphs with *seaborn.pydata.org*.

## 5. Supplementary Tables

Table S1. Glycan compositions for chain-A.

#	SITE	TYPE	STRUCTURE	SEQUENCE	
CHAIN A	G1	N17	FA2		bDGlcnAc(1→2)aDMan(1→6)[bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-17
	G2	N61	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-61
	G3	N74	A3		bDGlcnAc(1→6)[bDGlcnAc(1→2)]aDMan(1→6) [bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-74
	G4	N122	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-122
	G5	N149	FA2G2S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcnAc(1→2)aDMan(1→6) [bDGal(1→4)bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-149
	G6	N165	FA2G2S2		xDNeu5Ac(2→6)bDGal(1→4)bDGlcnAc(1→2)aDMan(1→6) [aDNeu5Ac(2→6)bDGal(1→4)bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-165
	G7	N234	M8		aDMan(1→2)aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-234
	G8	N282	FA3		bDGlcnAc(1→6)[bDGlcnAc(1→2)]aDMan(1→6) [bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)[aLFC(1→6)]bDGlcnAc(1→4)PROA-282
	G9	N331	FA2		bDGlcnAc(1→2)aDMan(1→6)[bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-331
	G10	N343	FA2		bDGlcnAc(1→2)aDMan(1→6)[bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-343
	G11	N603	FA2		bDGlcnAc(1→2)aDMan(1→6)[bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-603
	G12	N616	A2		bDGlcnAc(1→2)aDMan(1→6) [bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-616
	G13	N657	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-657
	G14	N709	M6		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-709
	G15	N717	Hybrid G1		bDGal(1→4)bDGlcnAc(1→2)aDMan(1→3)[aDMan(1→6) aDMan(1→3)]aDMan(1→6) [bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-717
	G16	N801	M6		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-801
	G17	N1074	FA2G2S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcnAc(1→2)aDMan(1→3) [bDGal(1→4)bDGlcnAc(1→2)aDMan(1→6)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-1074
	G18	N1098	FA2		bDGlcnAc(1→2)aDMan(1→6)[bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-1098
	G19	N1134	FA1		bDGlcnAc(1→2)aDMan(1→3)[aDMan(1→6)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-1134
	G20	N1158	A2		bDGlcnAc(1→2)aDMan(1→6) [bDGlcnAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4)bDGlcnAc(1→4)PROA-1158
	G21	N1173	FA4		bDGlcnAc(1→6)[bDGlcnAc(1→2)]aDMan(1→6)[bDGlcnAc(1→4) bDGlcnAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-1173
	G22	N1194	FA4G4S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcnAc(1→6)[bDGal(1→4)bDGlcnAc(1→2)]aDMan(1→6) [bDGal(1→4)bDGlcnAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcnAc(1→4) [aLFC(1→6)]bDGlcnAc(1→4)PROA-1194
	G23	T323	O-glycan		aDNeu5Ac(2→3)bDGal(1→3)aDGalNAc(1→4)PROA-323
	G24	S325	O-glycan		aDNeu5Ac(2→3)bDGal(1→3)aDGalNAc(1→4)PROA-325

Table S2. Glycan compositions for chain-B.

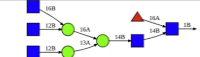

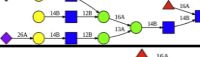
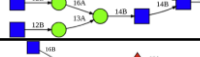
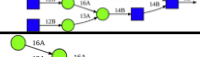

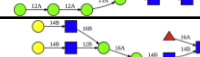
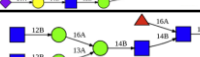
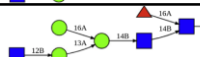
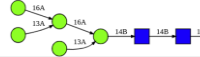
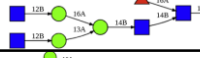
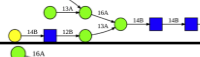
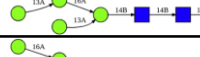
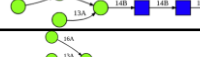
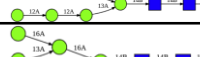

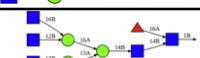
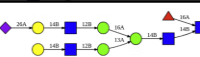

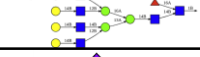



#	SITE	TYPE	STRUCTURE	SEQUENCE	
<b>CHAIN B</b>	G25	N17	FA3		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→→)PROB-17
	G26	N61	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-61
	G27	N74	FA3G3S2		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→6)[bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→6) [aDNeu5Ac(2→6)bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-74
	G28	N122	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-122
	G29	N149	FA3		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→→)PROB-149
	G30	N165	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-165
	G31	N234	M9		aDMan(1→2)aDMan(1→6)[aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→2)aDMan(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-234
	G32	N282	FA3G3S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)]bDGal(1→4)bDGlcNAc(1→6) [bDGal(1→4)]bDGlcNAc(1→2)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-282
	G33	N331	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-331
	G34	N343	FA1		bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-343
	G35	N603	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-603
	G36	N616	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-616
	G37	N657	Hybrid G1		bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-657
	G38	N709	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-709
	G39	N717	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-717
	G40	N801	M7		aDMan(1→2)aDMan(1→2)aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]aDMan(1→6)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-801
	G41	N1074	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-1074
	G42	N1098	A2		bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→→)PROB-1098
	G43	N1134	FA3		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→→)PROB-1134
	G44	N1158	FA2G2S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→6) [bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-1158
G45	N1173	FA4		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)bDGlcNAc(1→4) [bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→→)PROB-1173	
G46	N1194	FA4G4S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→6)[bDGal(1→4)]bDGlcNAc(1→2)]aDMan(1→6) [bDGal(1→4)]bDGlcNAc(1→2)]aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→→)PROB-1194	
G47	T323	O-glycan		aDNeu5Ac(2→3)bDGal(1→3)]aDNeu5Ac(2→6)]aDGalNAc(1→→)PROB-323	

Table S3. Glycan compositions for chain-C.

#	SITE	TYPE	STRUCTURE	SEQUENCE
G48	N17	FA3		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)PROC-17
G49	N61	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-61
G50	N74	A2		bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-74
G51	N122	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-122
G52	N149	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-149
G53	N165	FA2G2S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→6) [bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-165
G54	N234	M9		aDMan(1→2)aDMan(1→6)[aDMan(1→2)]aDMan(1→3)[aDMan(1→6)] [aDMan(1→2)]aDMan(1→2)aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-234
G55	N282	A2		bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-282
G56	N331	FA3G3S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[bDGal(1→4)]bDGlcNAc(1→6) [bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-331
G57	N343	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-343
G58	N603	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-603
G59	N616	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-616
G60	N657	Hybrid G1		bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6)] [aDMan(1→3)]aDMan(1→6)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-657
G61	N709	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-709
G62	N717	M6		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-717
G63	N801	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-801
G64	N1074	M5		aDMan(1→6)[aDMan(1→3)]aDMan(1→6) [aDMan(1→3)]bDMan(1→4)bDGlcNAc(1→4)bDGlcNAc(1→)PROC-1074
G65	N1098	Hybrid G1S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→2)aDMan(1→3)[aDMan(1→6)] [aDMan(1→3)]aDMan(1→6)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-1098
G66	N1134	FA2		bDGlcNAc(1→2)aDMan(1→6)[bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-1134
G67	N1158	A2		bDGlcNAc(1→2)aDMan(1→6) [bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)bDGlcNAc(1→)PROC-1158
G68	N1173	FA4		bDGlcNAc(1→6)[bDGlcNAc(1→2)]aDMan(1→6)[bDGlcNAc(1→4)] [bDGlcNAc(1→2)]aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4)[aLFuc(1→6)]bDGlcNAc(1→)PROC-1173
G69	N1194	FA4G4S1		aDNeu5Ac(2→6)bDGal(1→4)bDGlcNAc(1→6)[bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→6) [bDGal(1→4)]bDGlcNAc(1→2)aDMan(1→3)[bDMan(1→4)]bDGlcNAc(1→4) [aLFuc(1→6)]bDGlcNAc(1→)PROC-1194
G70	T323	O-glycan		bDGal(1→3)aDGalNAc(1→)PROC-323

CHAIN C



**Table S4.** Membrane lipids, their percentages in the membrane patch, and corresponding IUPAC names.

<b>Lipid</b>	<b>Percentage</b>	<b>IUPAC Name</b>
POPC	47%	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	20%	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
CHL	15%	(3 $\beta$ )-cholest-5-en-3-ol
POPI	11%	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol
POPS	7%	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine

**Table S5.** Summary of the full-length S protein all-atom MD simulations.

<b>Replica #</b>	<b>Closed</b>	<b>Open</b>	<b>Mutant</b>
<b>Rep 1</b>	543.60 ns	1000.50 ns	1001.30 ns
<b>Rep 2</b>	573.80 ns	1000.30 ns	1036.20 ns
<b>Rep 3</b>	614.10 ns	1006.30 ns	1018.40 ns
<b>Rep 4</b>	-	404.30 ns	411.70 ns
<b>Rep 5</b>	-	404.30 ns	407.00 ns
<b>Rep 6</b>	-	406.50 ns	416.60 ns
<b>Total</b>	<b>1731.50 ns</b>	<b>4222.20 ns</b>	<b>4291.20 ns</b>

**Table S6.** Accessible Surface Area (ASA) values for protein S' head in Open (A) and Closed (B). Glycan shielded area is the area covered by glycans. Glycosylated P ASA is the area effectively accessible in the presence of glycans. Non-Glycosylated P ASA is the accessible area in the absence of glycans (i.e. of the nude protein). AVG is average, ST.DEV is standard deviation. A full description is provided in Material and Methods section.

A OPEN							B CLOSED						
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]
1.4	18779.8	1571.4	159922.7	3174.9	141142.9	4145.5	1.4	18521.6	1571.9	157977.0	2438.0	139455.3	3746.0
2.0	24698.4	1665.1	130357.0	3444.4	105658.6	4398.0	2.0	24625.2	1744.2	128117.3	2254.9	103492.1	3617.8
3.0	31100.5	1434.8	107239.5	3488.8	76138.9	4272.8	3.0	31256.1	1547.3	104823.6	1963.9	73567.5	2961.4
4.0	35301.9	1075.7	96504.0	3490.1	61202.1	3955.2	4.0	35675.1	1210.0	93827.7	1860.4	58152.6	2343.5
5.0	38191.1	822.2	90792.7	3412.8	52601.6	3498.2	5.0	38897.3	1038.3	88215.8	1722.4	49318.5	1767.4
6.0	40356.3	986.2	87786.2	3345.6	47429.9	3006.9	6.0	41537.9	1192.1	85642.1	1590.7	44104.2	1467.4
7.0	42281.5	1328.9	86439.4	3325.4	44157.8	2699.9	7.0	43882.8	1438.8	84670.6	1487.4	40787.8	1393.2
8.0	44131.3	1613.1	86039.7	3301.1	41908.4	2567.3	8.0	45990.7	1660.3	84545.2	1431.6	38554.6	1501.2
9.0	45968.6	1808.5	86208.7	3195.1	40240.1	2499.6	9.0	47982.3	1879.2	84845.5	1393.0	36863.2	1628.8
10.0	47841.9	1921.5	86837.2	3074.8	38995.3	2502.5	10.0	49919.4	2037.1	85500.4	1345.2	35581.0	1744.0
11.0	49744.9	2002.0	87790.2	2951.8	38045.4	2530.3	11.0	51813.2	2212.0	86438.5	1373.7	34625.4	1861.0
12.0	51698.8	2098.7	89020.7	2884.3	37324.0	2602.2	12.0	53749.4	2315.6	87602.2	1393.9	33852.7	2005.5
13.0	53652.4	2229.5	90407.6	2870.1	36755.2	2710.7	13.0	55669.8	2427.7	88927.4	1392.7	33257.6	2132.1
14.0	55595.5	2382.7	91942.5	2864.8	36347.0	2831.9	14.0	57634.5	2520.2	90436.6	1424.9	32802.0	2235.4
15.0	57532.3	2558.8	93593.0	2881.6	36060.8	2962.7	15.0	59581.6	2574.8	92011.2	1422.9	32429.5	2333.3

**Table S7.** Accessible Surface Area values for protein S' stalk in Open (A) and Closed (B). Same notations as Table S2.

A OPEN							B CLOSED						
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]
1.4	2514.0	294.1	14591.3	590.3	12077.2	645.3	1.4	2302.4	246.1	14791.3	696.8	12488.9	666.3
2.0	3292.5	348.5	13132.1	582.5	9839.6	675.3	2.0	3002.9	270.0	13272.3	647.1	10269.4	618.6
3.0	4244.9	359.5	12129.2	506.5	7884.3	612.3	3.0	3977.4	297.4	12284.7	533.3	8307.3	528.7
4.0	5139.4	359.8	11865.1	476.1	6725.8	566.3	4.0	4946.9	336.3	12073.8	484.5	7126.9	495.8
5.0	6030.3	366.8	11872.6	476.2	5842.4	540.1	5.0	5913.4	386.2	12137.6	471.5	6224.2	481.5
6.0	6903.7	382.7	11994.4	489.5	5090.7	524.3	6.0	6863.0	442.9	12318.8	478.3	5457.7	468.7
7.0	7741.0	412.3	12170.6	509.2	4429.6	511.0	7.0	7777.2	503.6	12553.7	490.7	4776.5	459.4
8.0	8529.4	450.4	12374.3	531.2	3844.9	499.5	8.0	8630.6	558.3	12800.5	500.1	4169.8	454.0
9.0	9258.8	496.9	12591.7	558.3	3332.9	489.5	9.0	9419.1	612.3	13057.7	517.9	3638.6	453.3
10.0	9927.0	549.6	12810.0	584.9	2882.9	481.9	10.0	10138.2	661.8	13313.4	536.9	3175.2	450.1
11.0	10527.8	601.0	13026.0	611.3	2498.2	475.9	11.0	10782.2	705.1	13563.3	552.6	2776.1	454.3
12.0	11066.5	653.6	13233.3	641.6	2166.8	468.9	12.0	11362.0	744.6	13799.7	568.7	2437.7	461.3
13.0	11548.5	700.9	13432.7	674.2	1884.3	464.0	13.0	11872.3	786.3	14017.8	584.6	2145.5	472.2
14.0	11979.2	745.6	13621.2	702.9	1642.0	454.4	14.0	12332.3	828.3	14228.7	604.4	1896.4	484.4
15.0	12360.8	789.2	13794.6	735.9	1433.8	444.1	15.0	12740.7	857.2	14426.7	624.0	1686.0	496.8

**Table S8.** Accessible Surface Area values for protein S' RBD<sub>RBM</sub> in Open (A) and Closed (B). Same notations as Table S2.

A OPEN							B CLOSED						
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]
1.4	281.2	133.9	6625.4	439.8	6344.2	461.7	1.4	517.9	142.2	5855.3	301.2	5337.4	332.1
2.0	368.9	162.1	5825.7	497.9	5456.8	535.8	2.0	716.3	176.1	4732.7	291.5	4016.4	315.5
3.0	445.9	175.3	5408.3	560.6	4962.4	577.1	3.0	934.3	215.0	3931.5	284.7	2997.2	294.3
4.0	486.0	181.9	5341.8	621.8	4855.9	615.7	4.0	1053.5	225.7	3520.0	277.8	2466.5	284.2
5.0	517.7	186.3	5389.5	692.0	4871.8	667.9	5.0	1117.5	222.6	3297.2	280.6	2179.7	285.2
6.0	538.2	187.4	5486.8	767.1	4948.7	723.3	6.0	1152.0	221.2	3184.6	293.3	2032.6	295.8
7.0	554.5	190.8	5615.6	837.2	5061.1	776.8	7.0	1176.1	220.4	3129.0	312.7	1953.0	316.9
8.0	572.9	194.1	5766.7	891.4	5193.9	827.9	8.0	1200.1	218.6	3104.0	332.4	1903.9	342.5
9.0	591.6	196.2	5932.8	931.2	5341.1	876.7	9.0	1231.0	220.8	3103.0	352.1	1872.0	371.2
10.0	608.3	197.5	6110.5	966.6	5502.3	928.3	10.0	1269.6	223.6	3119.3	371.5	1849.7	400.4
11.0	626.4	200.6	6301.9	1004.4	5675.5	982.9	11.0	1315.8	229.8	3151.8	391.6	1836.1	429.1
12.0	649.2	207.3	6509.5	1046.2	5860.3	1041.2	12.0	1369.7	240.9	3198.3	412.9	1828.6	455.9
13.0	678.1	217.7	6727.7	1090.6	6049.6	1099.9	13.0	1426.2	254.3	3253.1	440.8	1826.9	482.9
14.0	713.2	232.1	6958.3	1137.5	6245.1	1159.7	14.0	1486.2	268.6	3312.8	467.6	1826.6	509.1
15.0	753.3	247.7	7199.9	1188.8	6446.6	1222.3	15.0	1546.5	286.4	3377.0	499.1	1830.4	537.4

**Table S9.** Accessible Surface Area values for protein S' RBD<sub>NON-INTERACTING</sub> in Open (A) and Closed (B). Same notations as Table S2.

A OPEN								B CLOSED							
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	RBD NON-int	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	
1.4	930.9	156.2	4536.2	257.8	3605.3	299.0		1.4	433.9	72.6	4000.4	195.1	3566.5	184.4	
2.0	1206.9	175.4	3776.9	274.7	2570.0	302.9		2.0	577.8	90.2	3137.9	195.7	2560.0	183.4	
3.0	1428.6	167.7	3093.0	272.7	1664.4	305.1		3.0	704.7	104.1	2476.8	190.2	1772.1	169.3	
4.0	1466.4	171.4	2641.0	264.8	1174.5	292.1		4.0	723.9	125.8	2130.8	191.3	1406.9	161.4	
5.0	1445.1	176.7	2332.6	245.3	887.5	260.7		5.0	708.7	135.7	1921.7	193.6	1213.0	164.2	
6.0	1418.8	190.1	2118.9	241.7	700.2	230.0		6.0	696.1	145.3	1792.3	201.8	1096.2	172.7	
7.0	1399.2	206.9	1964.4	247.3	565.2	204.9		7.0	692.8	158.8	1708.1	213.1	1015.3	184.4	
8.0	1383.7	220.4	1848.1	256.5	464.4	185.8		8.0	700.8	169.9	1651.0	220.5	950.2	194.9	
9.0	1375.5	231.8	1763.1	261.5	387.6	171.8		9.0	717.5	182.0	1609.4	230.2	891.8	204.0	
10.0	1373.8	243.6	1703.9	268.2	330.1	160.7		10.0	737.9	195.9	1579.4	241.3	841.4	211.6	
11.0	1380.3	250.9	1666.6	272.6	286.3	152.1		11.0	755.7	208.5	1553.0	251.2	797.3	218.5	
12.0	1392.4	259.3	1643.7	280.0	251.2	145.9		12.0	773.8	220.2	1532.4	261.8	758.6	225.3	
13.0	1407.8	267.4	1630.4	290.0	222.6	140.9		13.0	791.7	233.1	1513.6	270.9	721.9	232.8	
14.0	1421.0	275.9	1618.6	301.5	197.5	136.6		14.0	809.6	246.8	1497.7	279.3	688.1	240.5	
15.0	1432.5	286.1	1608.7	314.9	176.2	132.5		15.0	828.3	260.0	1485.3	290.0	657.0	247.2	

**Table S10.** Accessible Surface Area values for protein S' RBD<sub>WHOLE</sub> in Open (A) and Closed (B). Same notations as Table S2.

A OPEN							B CLOSED							
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	WHOLE RBD	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]
1.4	1211.6	200.4	11159.1	477.4	9947.5	551.2		1.4	952.4	157.6	9855.1	400.7	8902.7	425.7
2.0	1574.9	239.8	9599.5	526.3	8024.6	636.4		2.0	1295.6	199.2	7869.4	369.2	6573.8	393.8
3.0	1874.3	246.7	8499.3	644.6	6625.0	731.0		3.0	1641.7	236.5	6407.7	327.0	4766.0	349.0
4.0	1953.5	257.6	7980.0	708.5	6026.5	768.9		4.0	1778.3	248.0	5651.2	299.5	3872.8	311.8
5.0	1965.1	249.8	7719.0	749.4	5753.9	805.0		5.0	1826.7	246.0	5220.2	287.9	3393.5	294.5
6.0	1959.1	244.3	7601.5	809.0	5642.5	843.4		6.0	1848.2	250.3	4978.4	296.0	3130.1	296.5
7.0	1955.1	255.7	7573.2	877.6	5618.1	880.6		7.0	1868.6	259.4	4838.8	317.5	2970.2	313.1
8.0	1957.4	271.2	7604.7	929.4	5647.3	916.7		8.0	1901.1	262.6	4756.5	342.3	2855.4	334.6
9.0	1964.8	282.1	7679.9	952.9	5715.1	952.0		9.0	1949.7	274.5	4714.0	373.7	2764.2	360.6
10.0	1979.1	291.1	7797.0	972.1	5817.9	996.1		10.0	2008.5	290.5	4699.2	405.7	2690.7	389.1
11.0	2004.9	301.5	7950.0	1003.3	5945.1	1045.8		11.0	2072.5	304.7	4704.3	437.7	2631.8	420.3
12.0	2040.0	314.8	8131.4	1047.0	6091.4	1100.4		12.0	2142.2	323.8	4727.4	467.7	2585.2	447.9
13.0	2084.0	329.7	8333.2	1096.7	6249.2	1156.6		13.0	2213.6	347.2	4759.0	500.4	2545.5	479.4
14.0	2132.6	346.2	8549.3	1149.6	6416.7	1213.2		14.0	2290.7	373.4	4803.3	536.8	2512.6	514.5
15.0	2181.7	366.1	8775.7	1207.9	6594.0	1272.6		15.0	2368.4	404.4	4854.0	576.6	2485.5	546.4

**Table S11.** Summary of principal antibody epitopes reported in the literature. Antibodies denoted by \* refer to partial epitopes identified through mutational studies.

Antibody	Spike Domain	Epitope Residue Numbers	References
COVA1-22	NTD	141-156, 246-260	51
4A8		141-156, 246-260	50

B38	RBD	403-409, 415-421, 455-459, 473-479, 486-505	54
47D11		RBD core (338-437, 507-527)	57
S309		337-344, 356-361, 440-444, glycan at N343	53
CR3022		369-392, 427-430, 515-517	49,55
VHH-72		partial overlap with CR3022	59
COV2-2196*		F486, N487	52
COV2-2165*		F486, N487	52
COV2-2130*		K444, G447	52
MAb362*		Y449, F456, Y489	60
18F3*		D405, V407	58
7B11*		L441, S443, L452	58
n3021*		T500, N501, G502	62
n3113*		N354	62
n3130*		D428, F429, E516	62
1A9	CD	1111-1130	56

**Table S12.** Accessible Surface Area values for protein S' epitopes in Open (A) and Closed (B). Same notations as Table S2.

A		OPEN						B		CLOSED					
Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]	Probe Radius [Å]	Glycan shielded area AVG [Å <sup>2</sup> ]	Glycan shielded area ST.DEV. [±Å <sup>2</sup> ]	Non-Glyc. P ASA AVG [Å <sup>2</sup> ]	Non-Glyc. P ASA ST.DEV. [±Å <sup>2</sup> ]	Glyc. P. ASA AVG [Å <sup>2</sup> ]	Glyc. P. ASA ST.DEV. [±Å <sup>2</sup> ]		
7.2	293.8	149.9	2899.4	420.9	2605.6	452.1	B38	7.2	757.8	205.3	1599.4	297.9	841.6	256.9	
18.6	465.8	257.2	4406.9	1272.7	3941.1	1232.0	B38	18.6	1098.1	462.7	1757.4	692.6	659.3	454.0	
7.2	368.5	94.6	795.4	193.7	426.9	193.1	S309	7.2	264.0	101.3	844.3	202.9	580.4	152.6	
18.6	542.8	169.7	970.3	412.0	427.5	370.2	S309	18.6	287.2	190.7	643.6	405.2	356.4	305.9	
7.2	454.4	158.7	654.5	208.1	200.0	113.0	CR3022	7.2	23.9	24.0	25.4	24.9	1.5	4.1	
18.6	636.3	309.2	674.8	335.6	38.5	71.7	CR3022	18.6	0.0	0.0	0.0	0.0	0.0	0.0	
7.2	1279.9	526.7	3557.6	610.8	2277.7	427.5	4A8	7.2	1277.1	600.2	3060.0	748.5	1783.0	510.3	
18.6	3186.0	1264.5	6252.3	1297.9	3066.2	835.3	4A8	18.6	2859.2	1097.1	5020.3	1542.7	2161.1	1108.9	
7.2	577.3	120.2	675.1	80.2	97.8	93.1	1A9	7.2	508.2	100.5	609.9	81.9	101.7	80.1	
18.6	1106.7	172.0	1113.9	175.0	7.2	24.2	1A9	18.6	866.3	159.7	872.8	159.9	6.6	26.0	

## 6. Supplementary Movies

### Movie S1.

[https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl\\_file/oc0c01056\\_si\\_002.mp4](https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl_file/oc0c01056_si_002.mp4)

**Glycosylated full-length model of SARS-CoV-2 spike protein.** The movie shows, in the first part, the structure of the glycosylated full-length model of SARS-CoV-2 S protein in the open state (i.e. with 1 RBD in the “up” conformation, namely RBD-A) referred to as “Open” in the main text. The different domains and the color code used for lipids and glycans are indicated in the movie. In the second part, the movie shows the MD dynamics (with CHARMM36 force fields) of the same Open system. Only one MD replica was selected for illustrative purposes.

### Movie S2.

[https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl\\_file/oc0c01056\\_si\\_003.mp4](https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl_file/oc0c01056_si_003.mp4)

**N-glycan at N234 progressively inserts itself into the cavity left empty upon the lifting-up of the open RBD.** The movie shows the MD dynamics of the glycosylated head-only model of SARS-CoV-2 S protein described in Section 2 of the Supporting Information. These simulations were conducted with AMBER and GLYCAM force fields. N-glycan at N234 (highlighted with a magenta surface) progressively inserts in the space left empty by the lifting-up of the open RBD. We remark that in this system, based on a different cryo-EM structure (PDB ID: 6VYB), the open RBD is within chain B, here highlighted with an orange transparent surface. Nglycans at N165 and N343 are depicted with steel blue and cornflower blue surfaces, respectively. All the remaining glycans are shown with an admiral blue surface, whereas the protein is represented with gray cartoons.

### Movie S3.

[https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl\\_file/oc0c01056\\_si\\_004.mp4](https://pubs.acs.org/doi/suppl/10.1021/acscentsci.0c01056/suppl_file/oc0c01056_si_004.mp4)

**N-glycans at N165 and N234 “lock-and-load” the open RBD for infection.** The movie illustrates the structural role of the N-glycans at N165 and N234 in modulating the RBD conformational plasticity. By means of a closed-up view, the movie shows the MD dynamics of the open RBD (i.e., RBD-A) within the glycosylated full-length model of the SARS-CoV-2 S protein (wild-type, referred to as “Open” in the main text). Only one MD replica was selected for illustrative purpose. N-glycans at N165 and N234, and the RBD, are indicated with respective labels in the movie. All the remaining glycans are depicted with a per-residue colored licorice representation (GlcNAc in blue, Fucose in red, Galactose in yellow, Sialic Acid in purple, Mannose in green). Chain A, B and C of the spike trimer are depicted with cyan, red and gray cartoons, respectively.

## 7. Supplementary References

- (1) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* **2020**, *1263* (March), 1260–1263. <https://doi.org/10.1126/science.abb2507>.
- (2) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, 1–12. <https://doi.org/10.1016/j.cell.2020.02.058>.
- (3) Pallesen, J.; Wang, N.; Corbett, K. S.; Wrapp, D.; Kirchdoerfer, R. N.; Turner, H. L.; Cottrell, C. A.; Becker, M. M.; Wang, L.; Shi, W.; Kong, W. P.; Andres, E. L.; Kettenbach, A. N.; Denison, M. R.; Chappell, J. D.; Graham, B. S.; Ward, A. B.; McLellan, J. S. Immunogenicity and Structures of a Rationally Designed Prefusion MERS-CoV Spike Antigen. *Proc. Natl. Acad. Sci. U. S. A.* **2017**. <https://doi.org/10.1073/pnas.1707304114>.
- (4) Miroshnikov, K. A.; Marusich, E. I.; Cerritelli, M. E.; Cheng, N.; Hyde, C. C.; Steven, A. C.; Mesyanzhinov, V. V. *Engineering Trimeric Fibrous Proteins Based on Bacteriophage T4 Adhesins*; 1998; Vol. 11.
- (5) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- (6) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579* (7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
- (7) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539. <https://doi.org/10.1038/msb.2011.75>.
- (8) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. JPred4: A Protein Secondary Structure Prediction Server. *Nucleic Acids Res.* **2015**, *43*, 389–394. <https://doi.org/10.1093/nar/gkv332>.
- (9) Hartmann, M. D.; Ridderbusch, O.; Zeth, K.; Albrecht, R.; Testa, O.; Woolfson, D. N.; Sauer, G.; Dunin-Horkawicz, S.; Lupas, A. N.; Alvarez, B. H. A Coiled-Coil Motif That Sequesters Ions to the Hydrophobic Core. *Proc. Natl. Acad. Sci. U. S. A.* **2009**. <https://doi.org/10.1073/pnas.0907256106>.
- (10) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9* (1), 1–8. <https://doi.org/10.1186/1471-2105-9-40>.
- (11) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738. <https://doi.org/10.1038/nprot.2010.5>.
- (12) Yang, J.; Roy, A.; Zhang, Y. Protein-Ligand Binding Site Recognition Using Complementary Binding-Specific Substructure Comparison and Sequence Profile Alignment. *Struct. Bioinforma.* **2013**, *29* (20), 2588–2595. <https://doi.org/10.1093/bioinformatics/btt447>.
- (13) Shulla, A.; Gallagher, T. Role of Spike Protein Endodomains in Regulating Coronavirus Entry. *J. Biol. Chem.* **2009**, *284* (47), 32725–32734. <https://doi.org/10.1074/jbc.M109.043547>.
- (14) Thorp, E. B.; Boscarino, J. A.; Logan, H. L.; Goletz, J. T.; Gallagher, T. M.

- Palmitoylations on Murine Coronavirus Spike Proteins Are Essential for Virion Assembly and Infectivity. *J. Virol.* **2006**, *80* (3), 1280–1289.  
<https://doi.org/10.1128/jvi.80.3.12801289.2006>.
- (15) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *J. Comput. Chem.* **2011**. <https://doi.org/10.1002/jcc.21886>.
- (16) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 2967–2970.  
<https://doi.org/10.1002/jcc>.
- (17) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* **2020**.  
<https://doi.org/10.1126/science.abb9983>.
- (18) Zhang, Y.; Zhao, W.; Mao, Y.; Wang, S.; Zhong, Y.; Su, T.; Gong, M.; Lu, X.; Cheng, J.; Yang, H. Site-Specific N-Glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins Using High-Resolution Mass Spectrometry. *bioRxiv* **2020**.  
<https://doi.org/10.1101/2020.03.28.013276>.
- (19) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. Deducing the N-and OGlycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* **2020**, 1–8.  
<https://doi.org/10.1093/glycob/cwaa042>.
- (20) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**. <https://doi.org/10.1093/glycob/cwz003>.
- (21) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are and How They Behave. *Nature Reviews Molecular Cell Biology*. February 2008, pp 112–124.  
<https://doi.org/10.1038/nrm2330>.
- (22) Casares, D.; Escribá, P. V.; Rosselló, C. A. Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues. *Int. J. Mol. Sci.* **2019**, *20* (9). <https://doi.org/10.3390/ijms20092167>.
- (23) Krijnse-Locker, J.; Ericsson, M.; Rottier, P. J. M.; Grittiths, G. *Characterization of the Budding Compartment of Mouse Hepatitis Virus: Evidence That Transport from the RER to the Golgi Complex Requires Only One Vesicular Transport Step.*