Understanding Glycobiology through Glycans Structure and Dynamic Signatures: From Glycan Biosynthesis to Bacterial Adhesion

Maynooth University
National University of Ireland Maynooth

Akash Satheesan B.Sc. (Hons)

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

Doctor of Philosophy

By

Akash Satheesan B. Sc.

Hamilton Institute

Maynooth University

February 2025

Research supervisor: Prof David Malone and Dr Elisa Fadda

Head of Department: Prof David Malone

Table of Contents

Acknowledgements ***i***

PhD Timeline ***ii***

Abbreviations ***iii***

Declaration ***vii***

Abstract ***viii***

List of Publications ***ix***

Chapter 1. Introduction ***1***

1.1 Background **1**

1.2 Computational Method **10**

1.2.1 Fundamentals of Molecular Dynamics 10

1.2.2 Simulation Set-Up 14

1.2.3 Practical Tips for MD Simulations 19

1.3 Aims and Objectives **21**

References: **24**

Chapter 2. My Contribution to the GlycoShape Glycan 3D Structures Database ***32***

2.1 Introduction **32**

2.2 Methods **33**

2.2.1 Molecular Dynamics 33

2.2.1 Data Processing and Clustering 35

2.3 Results **35**

2.3.1 Triantennary *N*-Glycans 35

2.3.2 Blood Group Antigens 40

2.3.3 Lewis Antigens 41

2.3.4 α2,8-linked polysialic acids 42

2.4 Discussion **43**

2.5 Conclusion **45**

References: **46**

Chapter 3. Effect of bisection on the structure, dynamics, maturation and recognition of N-glycans ***50***

3.1 Introduction **50**

3.2 Methods **54**

3.2.1 Computational Methods 54

3.3 Results **55**

3.3.1 Biantennary *N*-glycans 55

3.3.2 Bisected and triantennary *N*-glycans 56

3.3.3 Effect of bisection on the *N*-glycan architecture 57

3.3.4 Effect of the third antenna on the *N*-glycan architecture 58

3.3.5 Assessing the suitability of A2B and A3 as substrates for b4GalT1 59

3.3.6 Assessing recognition of A2B and A3 by FUT8 62

3.4 Discussion **63**

3.5 Conclusion **66**

References: **67**

Chapter 4. Structure, dynamics, and glycan-binding specificities of the N. meningitidis Type IV Pili ***73***

4.1 Introduction **73**

4.2 Methods **76**

4.2.1 Computational Methods. 76

4.2.2 Surface Plasmon Resonance (SPR). 78

4.3 Results **79**

4.3.1 Structure and Dynamics of the *Nm* T4P 26-mer: The PilE system 79

5.3.2 Structure and Dynamics of the *Nm* T4P 26-mer: The PilE/PilV system 82

4.3.3 Finding the glycan binding site: Molecular Docking and Dynamics on a PilE Dimer 84

4.3.5 ‘Pilin Array’ Screening by MD simulations 87

4.3.6 Binding of Polysialic acids 89

4.3.7 Binding of Sialylated and non-sialylated LacNAc 91

4.3.8 Binding to Heparan Sulfate 92

4.3.9 Binding of Lewis B antigen 94

4.3.10 Experimental Binding Assays 94

4.4 Discussion: **96**

4.5 Conclusion **98**

References: **99**

Chapter 5. Structure, dynamics, and glycan-binding specificities of the N. gonorhoeae Type IV Pilus (T4P) ***104***

5.1 Introduction **104**

5.2 Methods **106**

5.2.1 Computational Method 106

5.2.2 Surface Plasmon Resonance 108

5.3 Results **109**

5.3.1 Structure and Dynamics of the *Ng* pilE System 109

5.3.2 'Pilin Array' Screening by MD Simulation 113

5.3.3 Binding of the α2-8 linked polysialic acid 115

5.3.4 sialyl-α(2-3)-LacNAc 115

5.3.5 Blood Groups A and B 116

5.3.6 Experimental Binding Assays 117

5.4 Discussion **119**

5.5 Conclusion **122**

References: **124**

Conclusion ***128***

Acknowledgements

First, I would like to thank my supervisor Dr. Elisa Fadda for the incredible amount of support and guidance you've given me through these last 4/5 years. I couldn't ask for a better supervisor and I truly appreciate your patience, kindness and encouragement throughout my PhD. While we may never fully agree on my accent, it has been an absolute pleasure getting to know you as a person.

I would also like to thank the members of my lab – Bea, Ojas and Silvia for all their support and endless chats. A special thanks to the past members of the lab, in particular Callum and Carl, who were great mentors and always reliable for a good laugh.

Many thanks to those in the Hamilton Institute, including those in my CRT cohort, especially Chang, Darshana, Fergal, Nathan, Jonny, Bill and Shauna for keeping the office environment so lively and supportive. A special thanks to Anthony, Ahmed, Conor, Dara and Paddy, for always entertaining my random ramblings.

I would like to give thanks to the staff of the Hamilton Institute in particular Joanna, Kate, Rosemary and David for all their support throughout my degree.

I want to express my deepest gratitude to my parents, Anitha and Satheesan, and my sister, Devika, for their unwavering support and unconditional love. I couldn't have done this without you.

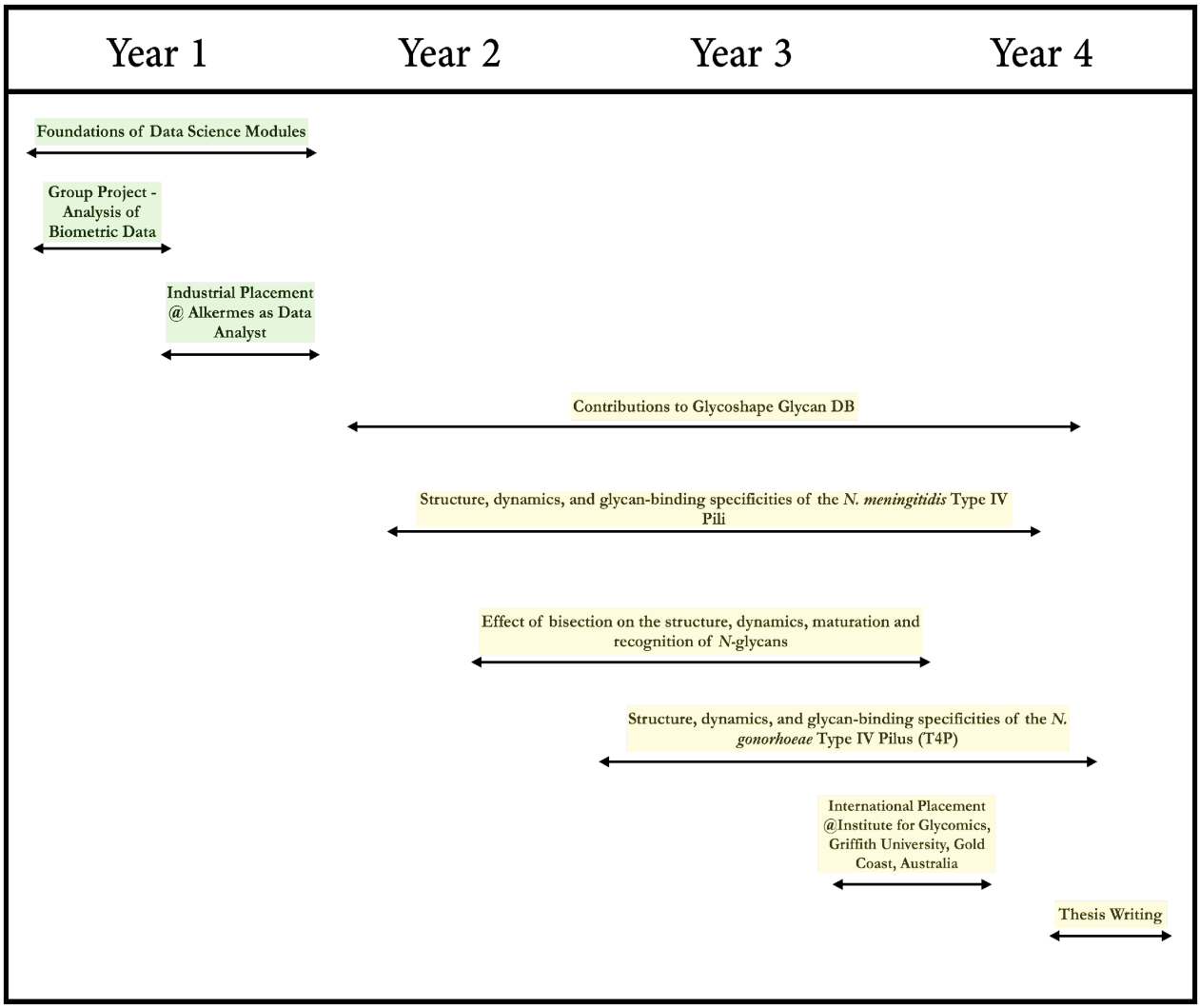
I want to thank Eleni, Evan, Jonny, Orla and Rory for being great friends that I could rely on. Special thanks to Blake and YC for basically being my pseudo-siblings through these years.

Finally, I would like to thank my wonderful partner, Neisha, for all her support and patience through my PhD. You have been my rock through all the tough times and I can't express enough how lucky I am to have you by my side.

PhD Timeline

My PhD programme was funded by the Science Foundation Ireland Centres for Research Training (SFI-CRT) in Foundations of Data Science. Alongside my specific PhD research, I underwent training in the core principles of data science, including machine learning and statistical analysis. This training included specialised data science modules, a group project where I analysed biometric data to evaluate its value as predictors for mood, and industrial placement at the pharmaceutical company Alkermes, where I worked as a data analyst between year 1 and 2 of my PhD.

A timeline of my PhD training is shown in the graphic below with details on the programme-specific training highlighted in green and PhD research activities highlighted in yellow.



Abbreviations

AMBER	Assisted Model Building with Energy Refinement
Ala	Alanine
Asn	Asparagine
B4GALT1	Beta-1,4-Galactosyltransferase 1
BGA	Blood Group Antigen A
BGB	Blood Group Antigen B
CHARMM	Chemistry at HARvard Macromolecular Mechanics
ChoP	Phosphorylcholine
CPS	Capsular Polysaccharide
CR3	Complement Receptor 3
DATDH	2,4-Diacetamido-2,4,6-trideoxyhexose
Dol-P	Dolichol Phosphate
ER	Endoplasmic Reticulum
Fuc	Fucose
FUT8	Fucosyltransferase 8
GAFF	General Amber Force Field
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GBP	Glycan Binding Protein

Glc	Glucose
GlcNAc	N-Acetylglucosamine
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	GROningen MOlecular Simulation
GAG	Glycosaminoglycan
HA	Hemagglutinin
HPC	High-Performance Computing
HS	Heparan Sulfate
IgG	Immunoglobulin G
KD	Dissociation Constant
LeA	Lewis A Antigen
LeB	Lewis B Antigen
LeX	Lewis X Antigen
LeY	Lewis Y Antigen
LINCS	Linear Constraint Solver
MGAT3	Beta-1,4-Mannosyl-Glycoprotein Beta-1,4-N-Acetylglucosaminyltransferase 3
MD	Molecular Dynamics
NCAM	Neural Cell Adhesion Molecules
NHBA	Neisserial Heparin Binding Antigen
NMR	Nuclear Magnetic Resonance

NOESY	Nuclear Overhauser Effect Spectroscopy
Nm	*Neisseria meningitidis*
Ng	*Neisseria gonorrhoeae*
Neu5Ac	N-Acetylneuraminic Acid
OPLS-AA	Optimized Potentials for Liquid Simulations-All Atom
OST	Oligosaccharyltransferase
PAFr	Platelet-Activating Factor Receptor
PDB	Protein Data Bank
PC	Phosphatidylcholine
PE	Phosphoethanolamine
PTM	Post-Translational Modification
pLDDT	Predicted Local Distance Difference Test
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
REMD	Replica Exchange Molecular Dynamics
Ser	Serine
SLTCAP	Statistical Learning Theory for Computational Applications in Physics
sLeA	Sialyl-Lewis A
sLeX	Sialyl-Lewis X
SPR	Surface Plasmon Resonance

STI	Sexually Transmitted Infection
T4P	Type IV Pili
UDP	Uridine Diphosphate
VMD	Visual Molecular Dynamics
WHO	World Health Organization

Declaration

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

Akash Satheesan

Abstract

Complex carbohydrates or glycans are one of the four main biomolecules essential for life. They play crucial roles in molecular recognition events that regulate immune responses, cellular communications and pathogen-host interactions. Despite their importance, characterising glycan structure and dynamics remains a significant challenge for structural biologists due to their inherent flexibility, combinatorial complexity and structural heterogeneity. Throughout my PhD, I used molecular dynamics (MD) simulations to characterise glycan structures and dynamics, revealing how their conformational behaviour dictates recognition and binding to protein receptors and enzymes, regulating processes from glycan biosynthesis to bacterial adhesion.

Based on collaborative projects with experimental glycobiologists and microbiologists, the my research focused on identifying the distinctive structural and dynamic features, or signatures, of free glycan structures regulating their molecular recognition. More specifically, I analysed multiatennary *N*-glycans, ABH and Lewis blood group antigens, and α(2-8)-linked polysialic acids. The data I generated from the MD analysis of these glycan structures contributed to the GlycoShape Glycan 3D Structure Database (https://glycoshape.org), a web-based open access (OA) resource designed, developed and curated by our research group to advance structural glycobiology.

Bisected *N*-glycan structures have been linked to specific disease states and progression. Using comparative analysis between free biantennary and triantennary *N*-glycan structures, I explored the structural consequences of *N*-glycan bisection, demonstrating how this modification alters glycan architecture and disrupts interactions with key enzymes in the *N*-glycosylation maturation pathway, namely B4GalT1 and FUT8, thereby preventing further functionalisation of the antennae. My results confirm and reconcile apparently discordant experimental results, ultimately suggesting an alternative biosynthetic pathway for the maturation of bisected *N*-glycan forms. Experimental validation of such pathway is in progress through a collaboration with Prof Daniel Kolarich and Dr Andrea Maggioni at the Institute of Biomedicine and Glycomics, Griffith University, QLD, Australia.

To understand how glycan structure and dynamics modulates recognition, I investigated glycan-protein interactions in the context of bacterial adhesion, focusing on Type IV pili (T4P) of *Neisseria meningitidis* and *Neisseria gonorrhoeae,* causative agents of meningococcal disease and gonorrhoea, respectively. My MD simulations revealed that T4P subunits form multi-subunit carbohydrate-binding pockets, enabling high-avidity interactions with α(2-8)-linked polysialic acids, mediated by conserved polar residues and post-translational modifications (PTMs) such as phosphorylcholine (ChoP) and O-linked bacterial glycosylation. Binding assays by surface plasmon resonance (SPR), I ran with guidance and support by Dr Chris Day and Dr Freda Jen during my internship in Prof Michael Jennings laboratory at the Institute of Biomedicine and Glycomics, Griffith University, QLD, Australia, confirmed the *Ng* and *Nm*T4P binding specificity for α(2-8)-polysialic acids motifs. Further to this, the SPR data indicate that T4P mutants lacking bacterial glycosylation showed an increased binding affinity, suggesting that T4P glycosylation may hinder binding possibly by restricting access to the glycan-binding site.

Through the scope of MD simulations, my thesis provides further insight into how glycan sequence and branching regulate their structure and dynamics which in turn can affect biosynthetic pathways, and molecular recognition, such as in glycan-mediated bacterial adhesion. My findings highlight the intricate relationship between glycan architecture and function and represent a 3D template that can used to inform the design of glycan-based diagnostics and glycomimetic therapeutics.

List of Publications

I. Ives, Callum M., Ojas Singh, Silvia D'Andrea, Carl A. Fogarty, Aoife M. Harbison, **Akash Satheesan**, Beatrice Tropea, and Elisa Fadda. 2024. "Restoring Protein Glycosylation with GlycoShape." *Nature Methods*, October, 1–11.

II. Harbison, Aoife M., Carl A. Fogarty, Toan K. Phung, **Akash Satheesan**, Benjamin L. Schulz, and Elisa Fadda. 2022. "Fine-Tuning the Spike: Role of the Nature and

Topology of the Glycan Shield in the Structure and Dynamics of the SARS-CoV-2 S." *Chemical Science (Royal Society of Chemistry: 2010)* 13 (2): 386–95.

Chapter 1. Introduction

In this chapter, I provide an overview of the background and context of my research, followed by a summary of the computational and analytical methods employed throughout this work. The chapter concludes with a statement of the research aims and objectives, and an outline of how these are addressed in the subsequent chapters.

1.1 Background

Complex carbohydrates, also known as glycans, are the most abundant biomolecules in Nature. These biomolecules are found in all living cells and are highly diverse in terms of size, branching patterns and 3D architectures. This level of diversity and abundance allows glycans to be involved in a large number of biological processes across various living systems.

Glycans are polymers of monosaccharides linked together by glycosidic bonds. There are many different types of monosaccharides, but they have the general formula of $C_x(H_2O)_n$ and are classified as polyhydroxyl aldehydes or ketones depending on the nature of their carbonyl group. Monosaccharides can be categorised into D or L configurations based on the orientation of the hydroxyl group on the chiral carbon furthest away from the carbonyl group based on the Fischer projection, see **Figure 1.1**(Seeberger 2015).

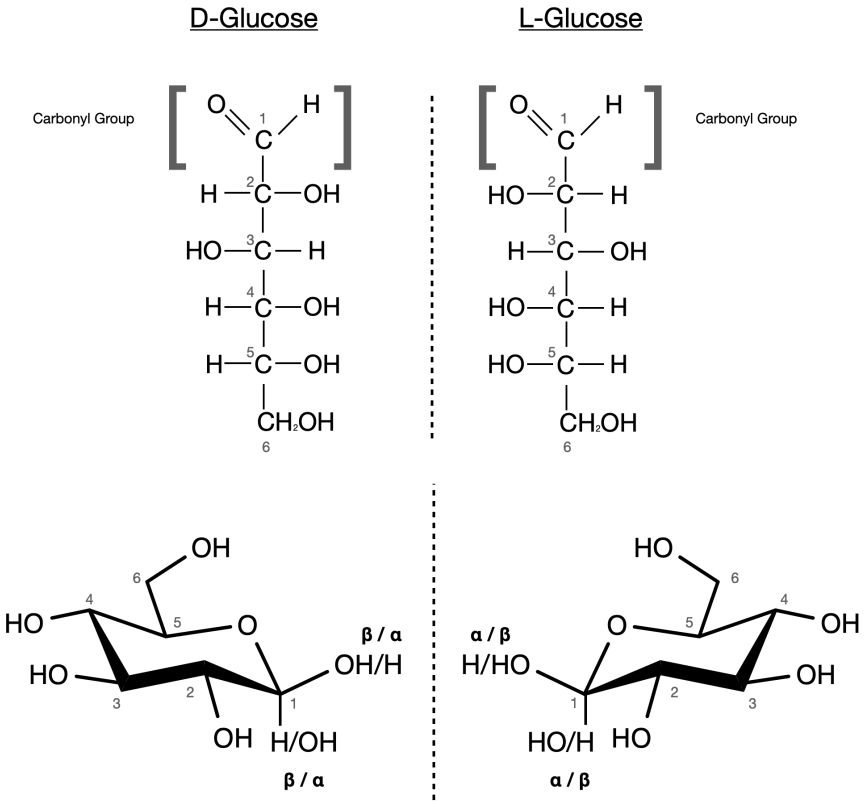


Figure 1.1: Fischer projections of the D- and L-glucose in an open chain form (top), with the corresponding chair conformations of the two enantiomers (bottom). The carbonyl group is highlighted in the Fischer projections and the position of the hydroxyl group in α/β positions around the anomeric carbon are labelled on both chair conformations.

While the Fischer projection represents the monosaccharide in an open (linear) conformation, in solution monosaccharides exist predominantly as five- or six-membered cyclic forms. In its cyclic form, the monosaccharide can exist in two different anomeric configurations, known as α or β depending on the position of the hydroxyl group around the C1 anomeric carbon, see **Figure 1.1**.

The ring structures is flexible to different degrees depending on the monosaccharide, and can access multiple conformations as a result of a process known as ring puckering(Chan *et al*. 2021; Perez & Makshakova 2022), which can be plotted on a Cremer−Pople sphere (Agirre 2017; Perez & Makshakova 2022), see **Figure 1.2**. In the case of five-membered rings, the ring structure alternates between envelope and twist conformations, whereas six-membered rings interconvert between stable chair conformations.

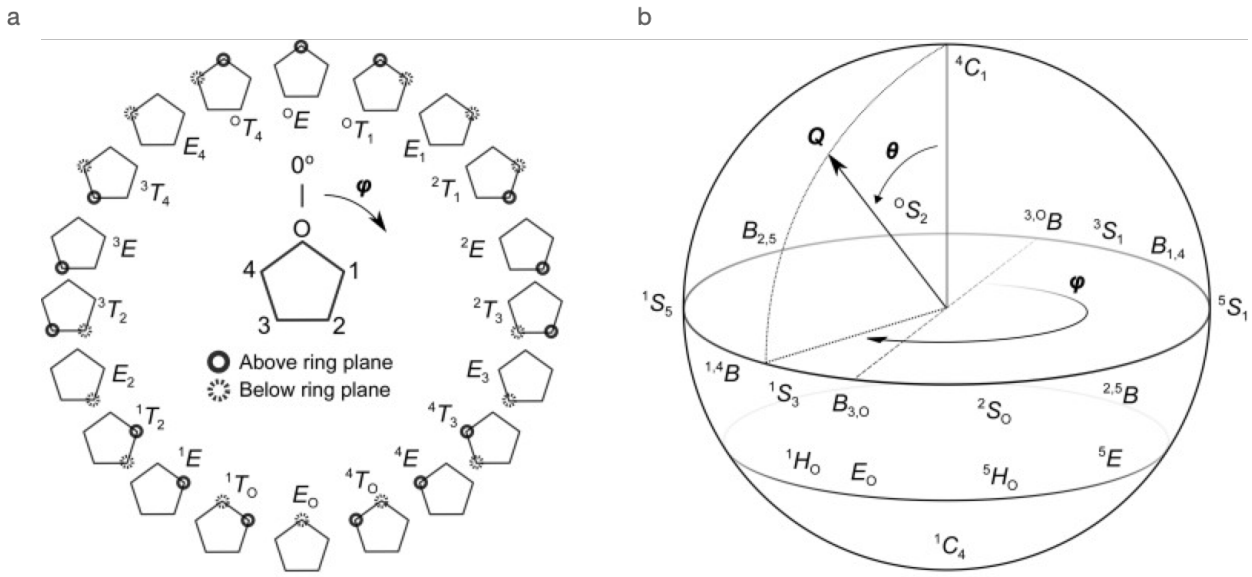


Figure 1.2 a.) 2D mapping of the main confirmations adopted by five-membered furanose rings, highlighting the twist (T) and envelope (E) conformations. **b)** Cremer-Pople sphere showing the possible configurations adopted by a six-membered pyranose ring, highlighting the chair (C), half-chair (H), envelope (E) and skew-boat (S) conformations. Adapted from reference(Agirre 2017).

The glycosidic linkages between monosaccharide units are formed through condensation reactions between the anomeric carbon of one monosaccharide and the hydroxyl group of another. These linkages can vary, resulting in different regioisomers and two possible stereoisomers (α or β) at the anomeric carbon of the monosaccharide(Seeberger 2015). Additionally, monosaccharides can form multiple linkages leading in branched structures. As the number of monosaccharides increases, the possible configuration of stereoisomers, regioisomers and branched structures grows exponentially resulting in a combinatorial explosion of unique glycan structures(Laine 1994). Glycosidic linkages can be relatively flexible due to their chemical nature. Indeed, multiple rotational degrees of freedom around the torsion angles, namely phi (φ), psi (ψ), and in the case of 1/2-6 linkages omega (ω), see **Figure 1.3**. These torsion angles can be fully accessible at room temperature with fluctuations in the range of up to 20° around equilibrium values leading to a high conformational diversity.

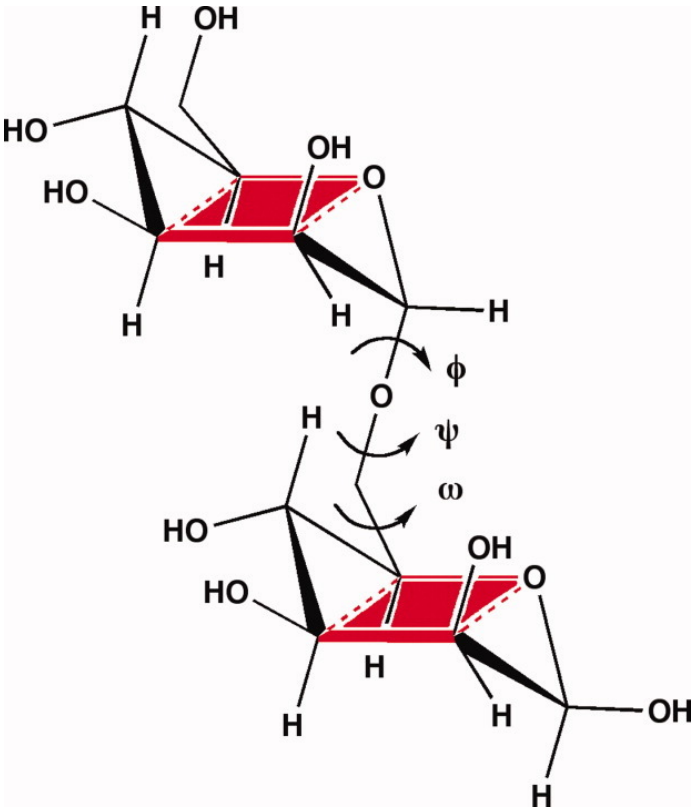


Figure 1.3 Graphical representation of the phi (ϕ), psi (ψ), and omega (ω) torsion angles formed by glyocsidic bonds. Adapted from reference(Salisburg et al., 2009). The phi (ϕ) angle is defined by the atoms O5–C1–O1–Cx, psi (ψ) by C1–O1–Cx–Cx+1, and omega (ω) by O1–Cx–Cx+1–Cx+2.

Unlike proteins, glycan structures are not encoded in the genome but are secondary gene products synthesised by specific enzymes, namely glycosyltransferases (GTs) and glycoside hydrolases (GHs) (Rini *et al*. 2009; Varki & Sharon 2009). However, the diversity of glycan structures in a given biological system is not random but tightly regulated by the expression of the GTs and GHs responsible for specific glycan structures(Nairn *et al*. 2008).

Glycan structures are more often found in the context of glycoconjugates, i.e. covalently attached to other biomolecules, such as proteins and lipids. The functionalisation of proteins with glycans is a process known as glycosylation, i.e. one of the most common post-translational modifications (PTMs). The two most common forms of protein glycosylation in eukaryotic cells are *N*- and *O*- glycosylation. *N*-glycans are covalently linked to an Asn residue within a sequon N-X-(S/T), where X is any residue except Pro, through the formation of a glycosidic bond to its amide N atom. All eukaryotic *N*-glycans share a common pentasaccharide core, Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1– 4GlcNAcβ1–Asn that can be further functionalised and classified into 3 types, namely oligomannose, complex and hybrid

N-glycans, based on their monosaccharide makeup extending from the core(Stanley *et al*. 2022), see **Figure 1.4.**

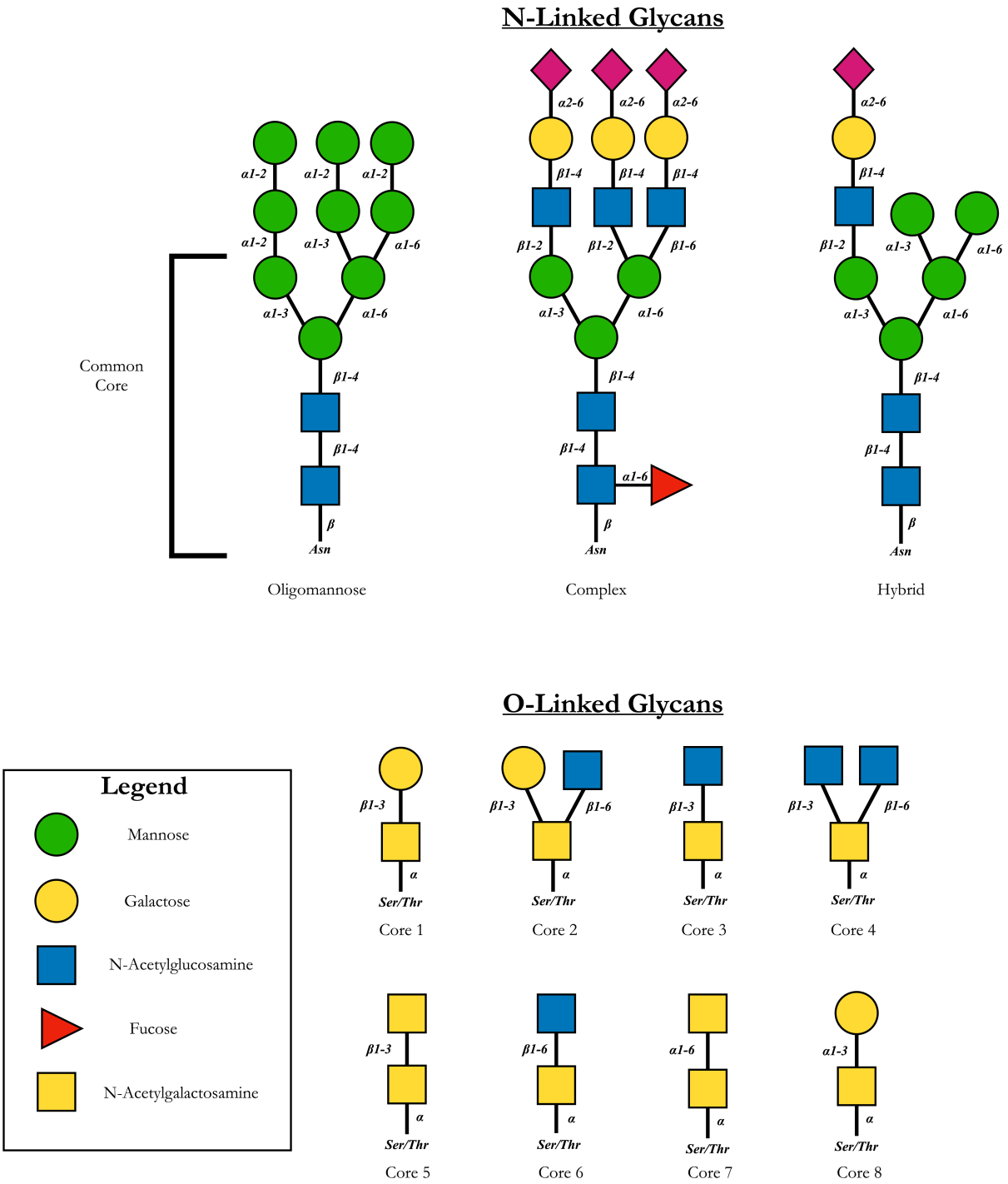


Figure 1.4: System Nomenclature for Glycans (SNFG)(Thieker *et al*. 2016) representation of the structures of the different classes of *N*- and *O*- linked glycans with a legend provided.

O-glycans are covalently linked to the O atom of the hydroxyl group of the sidechain of Ser/Thr residues. Unlike *N*-glycans, *O*-glycans do not have a common core that is shared among all glycan structures, see **Figure 1.3**. O-linked glycosylation occurs through the transfer of a single

monosaccharide, namely Man, Xyl, Fuc, Gal, GalNAc, Glc or GlcNAc, to the Ser/Thr residue(Wilkinson & Saldova 2020). Among these different types, one of the most common is the *O*-GalNAc linkage(Nielsen *et al*. 2022; Wilkinson & Saldova 2020). The *O*- linked GalNAc root, also known as Tn antigen, can be extended to match a set of eight core structures, see **Figure 1.4,** which can be further functionalised to form more complex structures, such as the blood group and Lewis antigens(Brockhausen *et al*. 2022; Wilkinson & Saldova 2020).

Alongside their structure and glycosylation sites, *N*- and *O*- linked glycans are also distinct in their biosynthesis. *N*-glycan biosynthesis starts in the endoplasmic reticulum (ER) membrane with the transfer of a 14 sugar long glycan from the lipid carrier Dol-P to the Asn residue of the N-X-(S/T) sequon by the oligosaccharyltransferase (OST)(He *et al*. 2024; Stanley *et al*. 2022). *N*-glycans are then processed and functionalised through the activity of glycosidases and glycotransferases as they move down the lumen of the ER and through the Golgi apparatus, see **Figure 1.5**.

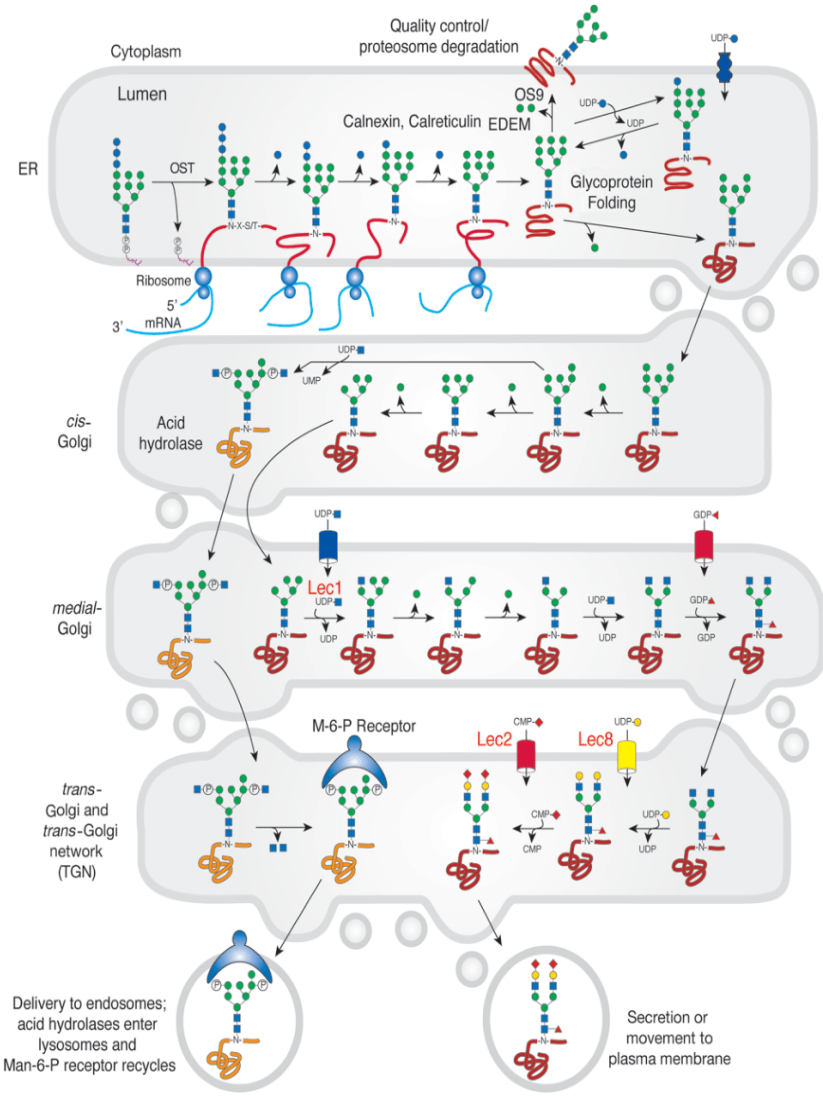


Figure 1.5: Biosynthesis pathways of N-glycans in the cell. Adapted from reference(Stanley *et al*. 2022)

O-linked glycosylation occurs in the Golgi. In the case of *O*- linked GalNAc glycans, the process begins with the transfer of the UDP-Gal to the Ser/Thr residue by a polypeptide GalNAc-transferase (GALNT)(Brockhausen *et al*. 2022). The resulting Tn antigen is then extended to the various cores and further functionalised by the action of a variety of GTs(Wilkinson & Saldova 2020; Brockhausen *et al*. 2022). Unlike *N*-glycans, whose maturation involves both GTs and GHs, the maturation of *O*-glycans relies exclusively on GTs.

The glycosylation pathway is both sequential and highly competitive, with glycosyltransferases and glycosidases often competing for the same glycan substrate(Zacchi & Schulz 2016; Reily *et al*. 2019; Kim *et al*. 2009). These enzymes often operate with sub-optimal efficiency, resulting in incomplete glycan structures(Zacchi & Schulz 2016). This enzymatic inefficiency

gives rise to two distinct forms of structural diversity in glycoproteins: macroheterogeneity and microheterogeneity(Zacchi & Schulz 2016; Varki & Sharon 2009; Čaval *et al*. 2021). Macroheterogeneity arises from the incomplete transfer of glycans to specific glycosylation sites, leading to variability in the presence or absence of a glycan at a particular site, while microheterogeneity arises from the difference in glycan structure present at a particular site as a result of incomplete or differential enzymatic activity during glycan processing. Both macro- and microheterogeneity are inherent features of glycosylation, influenced by various factors such as enzyme expression levels, glycan substrate availability, accessibility of glycosylation sites, as well as the physiological state of the cell(Zacchi & Schulz 2016). This heterogeneity reflects the dynamic and intricate nature of the glycosylation machinery, highlighting the importance of specific glycan structures in ensuring efficient functionality in a given biological system.

Glycans play crucial roles in countless biological processes across various systems. Their biological roles are diverse and complex but they can be loosely broken down into three broad categories: structural contributions, energy metabolism and information carriers(Gagneux *et al*. 2022). While all three functions are complex and important in their own right, for the purposes of this thesis, I will be looking at their role as information carriers, more specifically exerted through molecular recognition events.

Different glycan 3D structures and the dynamic equilibrium regulating their accessibility function as molecular signatures, mediating interactions with glycan-binding proteins (GBPs). These can be subdivided into two categories: lectins and sulfated GAG-binding proteins(Taylor *et al*. 2022). Lectins are involved in a wide range of recognition processes, while sulfated GAG-binding proteins interact with glycosaminoglycans (GAGs) to support structural organisation. Lectins serve in both self and pathogenic recognition(Schnaar 2016). In humans, lectins interact with self-glycans to maintain normal biological processes, facilitating cell-cell communication, glycoprotein trafficking and immune regulation(Taylor *et al*. 2022; Gagneux *et al*. 2022). In pathogens, many lectins function as adhesins, surface proteins that enable the attachment of the pathogen to the host cell and promote colonisation of host cells, ultimately playing a crucial role in infection(Taylor *et al*. 2022).

The recognition and binding of glycans by lectin hinges on a combination of non-covalent interactions, including electrostatic interactions and van der Waals contacts as well as other

contributing factors such as the hydrophobic effect(J. Angulo *et al*. 2022; Pérez & Tvaroška 2014). In terms of binding affinity the glycan-protein interactions are typically weaker than protein-protein interactions. However, affinity is usually increased through multivalency, where multiple interactions between the different lectin domains and the glycan antigen enhance binding strength(Kiessling *et al*. 2000; Quintana *et al*. 2023; Zeng *et al*. 2012).

Specificity plays a key role in lectin binding(Raposo *et al*. 2021; Sood *et al*. 2018). Lectins recognise glycans based on key structural features such as terminal sugar sequences, glycosidic linkages, and branching patterns(Taylor *et al*. 2022). A clear example of this specificity is seen in influenza viruses, where the hemagglutinin (HA) glycoprotein of human influenza viruses preferentially binds to glycans containing α2,6 linked terminal Neu5Ac, abundant in the human upper respiratory tract, whereas in avian influenza viruses the HA glycoprotein preferentially targets glycans containing α2,3 linked terminal Neu5Ac, which is a common epitope in birds respiratory epithelium, but only present in the lower respiratory tract and conjunctival cells in humans(Kumlin *et al*. 2008). Modifications like sulfation, fucosylation, and sialylation also create unique binding epitopes, as seen with selectins binding sulfated Lewis antigens(Chugh *et al*. 2015).

The characterisation of glycosylation patterns and of their corresponding 3D structures is a difficult task by means of traditional structural biology techniques, even under cryogenic conditions(Nagae & Yamaguchi 2012; A. Angulo *et al*. 2022; Gimeno *et al*. 2020; Atanasova *et al*. 2020) and glycans are often partially or completely removed from proteins for crystallisation(Agirre *et al*. 2015). Adding to the challenge is the structural diversity of glycans, which results from the wide variety of regioisomers, stereoisomers, branching patterns for a given glycan sequence, and the inherent heterogeneity. However, the recent advancement of glycoinformatics tools and databases(Alocci *et al*. 2019; York *et al*. 2020; Tiemeyer *et al*. 2017) alongside the rapid improvement of high-performance computing molecular simulations(Fadda 2022; Fadda *et al*. 2024) have risen to meet this challenge allowing us to gather detailed analysis of glycan structure and dynamics giving us valuable insight to their function in various biological process.

1.2 Computational Method

This section outlines the molecular dynamics (MD) simulation techniques I used in my PhD work, beginning with the fundamental principles of classical mechanics that underpin MD simulations, including force fields approximation, integration methods, and thermodynamic ensembles. I will then describe the step-by-step process of setting up an MD simulation, from system preparations such as selecting structures, assigning force fields, and solvating the system to running simulations on high-performance computing (HPC) infrastructure. Finally, I will share insights gained throughout my PhD, highlighting best practices and useful strategies for conducting reliable and efficient simulations.

1.2.1 Fundamentals of Molecular Dynamics

Molecular Dynamics (MD) simulations(Fadda *et al*. 2024; Fadda 2022; Braun *et al*. 2019) have become progressively a more common and useful computational tool for studying the structure, dynamics, and energetics of biomolecular interactions at the atomistic-level of detail. MD simulations allow us to simulate the movement of atoms and molecules over time and over phase-space, offering insights that complement the static structures captured by experimental techniques such as X-ray crystallography and cryo-EM, or predicted by machine learning (ML) models. MD are particularly useful to supplement experiments by characterising the structural and mechanistic aspects of complex biological processes involving glycans, which are highly flexible and heterogeneous, thus largely invisible through structural biology techniques.

Within the framework of classical mechanics, which forms the foundation of MD simulations, atoms are assumed to be hard impenetrable spheres with specific masses, sizes and point charges. The motion of each atom of a molecule is governed by Newtonian physics, which describes how the force acting on an atom influences its acceleration and subsequent trajectory of motion. This relationship can be expressed using Newton's second law in **Eq. 1**,

$$F = ma \tag{1}$$

where F is the force, m is the atom's mass, and a its acceleration. The force acting can be expressed as the negative of the derivative of the potential energy in function of the displacement, as shown in **Eq. 2,**

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

where *V(r)* represents the potential energy of the molecule as a function of the position *r*. Within the classical mechanics approximation, the potential energy can be defined by an empirical force field, which is generally expressed by the function in **Eq. 3,**

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

where the potential energy of the molecule *V(r)* is a function of the position of the N atoms in the system and can be obtained as the sum of the potential energy functions deriving from bonded (covalent) and non-bonded (non-covalent) interactions. The first three terms in Eq 3 correspond to the potential energy contributions from the covalent interactions and include bonds, bond angles and torsion (dihedral) angles. The bonded interactions potential is approximated through a Hooke's potential to represent the energy of bond vibrations and angle bending. Within this framework torsions are represented by a sinusoidal function because of their periodicity. The last term of Eq 3 corresponds to the non-bonded interactions, namely dispersion interactions, approximated via Lennard-Jones potential, and electrostatic interactions, represented by a Coulomb potential.

Force fields are separated into additive or polarisable forcefields. In the case of additive or non-polarisable forcefields, which are the most widely used in MD simulations(Lin & MacKerell 2019), the partial charges, *q*, are fixed or static around the respective atom while in polarisable forcefields the partial charges are variable, approximating explicitly the electron polarisation effects exerted by the molecular environment surrounding the system.

The force field parameters determine the energy potential of the target system, and as such it is important to select a parameter set that has been specifically designed and developed to represent as accurately as possible the type of biomolecule(s) you are simulating. In terms of glycans and glycoconjugates, the most widely used parameter sets are AMBER/GLYCAM06(Kirschner *et al*. 2008) and CHARMM/CHARMM36(Guvench *et al*. 2011). GLYCAM06 is specifically designed for modelling carbohydrates. It is complementary in its structure and implementation to the AMBER force field, yet it is independent of AMBER and could be used in principle in combination with any force field (Kirschner *et al*. 2008). CHARMM36 provides specialised parameters for glycans, including glycosidic linkages, and glycoprotein linkages. It is compatible with the broader CHARMM all-atom biomolecular force fields, enabling the modelling of glycoproteins, glycolipids, and other glycan-containing systems. Other carbohydrate forcefield sets include GROMOS(Nester *et al*. 2019), which uses a united-atom approach for modelling biomolecules, OPLS-AA, an all-atom force field parameter set, and DRUDE(Aytenfisu *et al*. 2018), a polarisable force field that extends CHARMM to include electronic polarisation effects. The parameterisation protocols for these forcefield sets are shown in **Figure 1.6.**

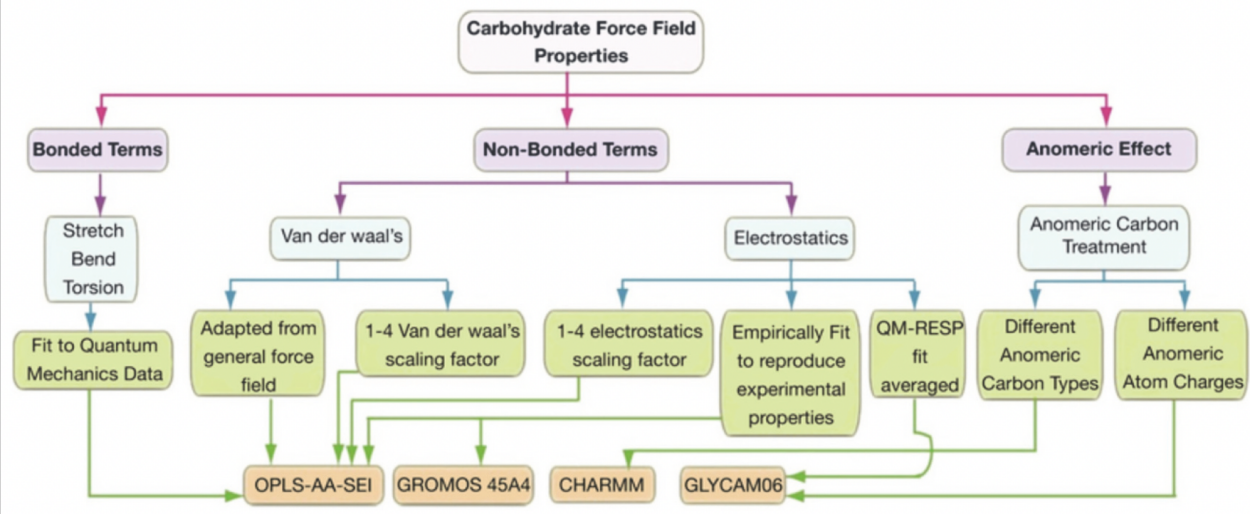


Figure 1.6: Parameterisation protocols for various carbohydrate forcefields. Adapted from reference(Perez *et al*. 2021).

Once the force field has been selected, an MD trajectory can be generated by integrating the Newton's equations of motion in **Eq. 2** with numerical integrators. As an example of commonly used integrators, the Verlet integrator(Verlet 1967) and its variant, the Leapfrog integrator, are shown in **Eq. 4** and **Eq. 5**, respectively.

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

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

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\,\delta t^2 \tag{4}$$

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

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + a(t)\,\delta t \tag{5}$$

The Verlet integrator shown in **Eq. 4**, updates atomic positions at time t+δt using the current acceleration at time *t* and previous positions at time *t* - δt. The basic Verlet integrator does not directly calculate or store velocities, which may result in inaccuracies over time. The leapfrog integrator shown in **Eq. 5**, addresses this issue by computing velocities at half-time steps $t + \frac{1}{2}\,\delta t$ and using them to compute positions at full-time steps $t + \delta t$.

The integration time step, δt, is important as it determines how long the simulation will take to run depending on your computational resources. The time-step has to be one order of magnitude lower than the fastest bond vibration in your system which in most systems is the C-H bond vibration which approximates to 10 fs, leading us to a time-step of 1 fs. By placing constraints on bond lengths involving hydrogen atoms, such as those applied with the SHAKE(Ryckaert *et al.* 1977) and LINCS (Hess *et al.* 1997) algorithms, we can extend the time step to 2 fs and access much longer simulation times without losing accuracy.

MD simulations can be run in different thermodynamic ensembles, where the observables of interest can be calculated through time or phase-space averages. In a microcanonical ensemble (NVE), the number of particles (N), volume (V), and total energy (E) remain constant. This ensemble is useful for studying the intrinsic dynamics of isolated systems and their energy conservation as there are no external temperature and pressure constraints. In a canonical ensemble (NVT), the number of particles (N), volume (V), and temperature(T) are kept

constant. NVT systems requires regulation of temperatures at equilibrium with specific algorithms aimed at rescaling velocities, such as the one proposed by Berendsen(Berendsen *et al*. 1984), Nosé-Hoover(Nosé 1984), and Langevin(Brooks *et al*. 1980). The NVT ensemble is mainly used for systems that require temperature and simulation volume control. The isothermal-isobaric (NPT) ensemble extends the NVT ensemble by incorporating a barostat, such as the one proposed by Berendsen(Berendsen *et al*. 1984), Parrinello-Rahman(Parrinello & Rahman 1981), and Monte Carlo(Åqvist *et al*. 2004), to regulate pressure (P), allowing for fluctuations in volume. This ensemble is widely used for solvated systems, as it mimics experimental conditions by maintaining both temperature and pressure constant as it is likely the case in a laboratory setting.

1.2.2 Simulation Set-Up

In this section, I will outline the process of setting up an MD simulation of a glycoprotein under physiological conditions on high-performance computing (HPC) infrastructure, from preparing your system to running the MD simulation as described in the workflow schematic shown in **Figure 1.7.**

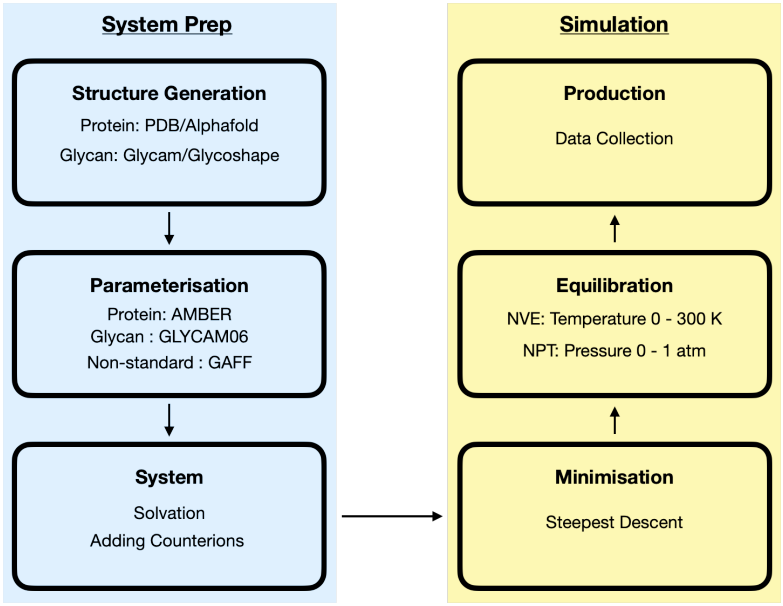


Figure 1.7: Workflow for setting up and running an MD simulation of a glycoprotein.

1.2.2.1 System Preparation

The first step in preparing the system would be generating PDB structures for the protein and glycan(s). The crystal structure of the protein can be sourced from databases such as

PDB(Berman *et al*. 2000) and structures can now be predicted using a protein sequence with AlphaFold(Jumper *et al*. 2021; Varadi *et al*. 2022). As a fundamental first step, it is important to assess the quality of the chosen starting structure in terms of accuracy and suitability to address the question we are trying to answer by MD. For example, in the case of crystal structures it is important ensure the resolution is suitable for the study(Burley *et al*. 2022) and that all the key parts of the protein are included in the 3D structure. For AlphaFold predicted structures, the pLDDT confidence score should be evaluated and, where possible, the structure should be compared against known experimental structures to evaluate its reliability and the potential of alternative conformations.

Where the target glycan is not available in the GlycoShape database, starting structures can be generated from the GLYCAM Carbohydrate Builder tool (www.glycam.org). As an important caveat, the generated structures do not necessarily reflect the dynamic ensemble of the glycan at equilibrium in a solvated environment. MD simulations and analysis of the structures may be required to identify the representative conformers. Glycan structures from all significantly (>10%) populated conformations at equilibrium can be sourced from the GlycoShape(https://glycoshape.org/)(Ives *et al*. 2023), a 3D structure database designed, developed and curated in our research lab that provides glycan structures from extensive MD simulations of over 640 glycans. The development and application of GlycoShape are discussed in further detail in **Chapter 2**.

Once the protein and glycan structures have been chosen or generated, the glycan must be linked to the protein at the appropriate site. This can be done using graphic user interfaces (GUI) such as PyMOL (www.pymol.org) which has a bond functionality that can be used to manually create the linkage. If the protein structure was crystallised with a fragment of the glycan, the fragment can serve as an alignment reference to correctly position the full glycan structure using PyMOL's alignment function.

A much more efficient and precise approach is to use GlycoShape Re-Glyco, a tool available on (https://glycoshape.org/reglyco), which streamlines the attachment of pre-equilibrated glycans from the GlycoShape database and optimises their orientation by minimising steric clashes with the protein that may occur at the glycosylation site.

Once the glycoprotein 3D structure has been rebuilt and saved in PDB format with the atom names corresponding to the AMBER/GLYCAM nomenclature, we can generate the corresponding topology file by linking the structure to the force field parameters. This can be done using tools developed specifically for the chosen MD simulation software. Here I will describe the process using the AMBER 18(Lee *et al*. 2018) software package. Topology files are generated using *tleap*, a program within AMBER 18 that builds molecular systems, assigns force field parameters, and outputs topology and coordinate files. Before using *tleap*, it may be necessary to pre-process the PDB file using *pdb4amber*, another tool from the AMBER 18 package, to ensure format compatibility. Once the PDB file is cleared, it can be loaded into *tleap*, where the parameters are assigned to each atom according to their atom types. In the work I have done during my PhD I used the AMBER force field ff14SB(Maier *et al.* 2015) to represent protein atoms and GLYCAM06(Kirschner *et al.* 2008) parameters to represent carbohydrates. For systems containing non-standard residues, i.e. residues not included in either protein or carbohydrate force fields, we used the General Amber Force Field (GAFF), an AMBER-type force field with parameters suitable to represent small organic molecules, ligands, and modified/non-natural amino acids(Wang *et al*. 2004).

In **Chapter 3**, I describe the process of generating parameters in the context of pilin E (PilE), the major pilin subunit of *Neisseria meningitidis*. PilE contains phosphorylcholine (ChoP), a non-standard residue not included in the AMBER or GLYCAM06 parameter set. The introduction of GAFF parameters in the topology is slightly cumbersome. It involves first generating a 3D structure for the isolated ChoP. First I generated a phosphatidylcholine (PC), a lipid present in the AMBER lipid17 force fields, using *tleap* and saved it as a PDB file. I used PyMOL to remove the hydrophobic fatty acid chains and glycerol backbone, leaving only the phosphate and choline groups. I saved the resulting structure as a PDB file and processed it using *antechamber*, a tool within the AMBER 18 package, that assigns GAFF atom types and atomic partial charges. To identify any missing parameters, I used the *parmchk2* tool to generate a *.frcmod* file, which listed undefined force field terms. I derived the missing parameters from chemically similar structures covered within GAFF. Once I assigned the necessary parameters, I generated a prep file to define the internal molecular structure, connectivity, and charges of ChoP in AMBER format. Finally, I loaded the *.frcmod* and *.prep* files into *tleap*, where I integrated the ChoP within the glycoprotein topology.

After generating the topology, the system must be solvated using a water model to mimic biological conditions. The most commonly used water model in combination with AMBER and GLYCAM-06 parameter sets is TIP3P(Jorgensen *et al.* 1983) due to its complementarity and computational efficiency. It is important to understand that such water model is quite simplistic and contributes to overestimate electrostatic interactions, while underestimating solvation free energies(Onufriev & Izadi 2018). For better accuracy, more sophisticated water models, such as TIP4P or TIP5P with their many different flavours available in terms of dispersion(Piana *et al.* 2015) and or electrostatic interactions scaling(Cruces Chamorro *et al.* 2024), may be preferred, however these come with an increased computational demand(Onufriev & Izadi 2018; Vega & Abascal 2011) and are not required in the present context.

To prevent boundary effects and ensure bulk-like behaviour, the solvated system is enclosed within a periodic box, which replicates the simulation cell in all directions to infinity. The size of the box must be carefully chosen to maintain a sufficient buffer distance between the solute and its own image, preventing self-interactions (Braun *et al*. 2019). Yet a box that is too large will increase the computational cost, so the right choice is highly system dependent. In AMBER, periodic boxes can be selected and built in *tleap* using the 'solvateBox' function.

Once the system is solvated, ions should be added to neutralise the total electrostatic charge charge of the system at the target pH, and to mimic physiological ion concentration using the 'addIons' function in *tleap*. For all glycans included in the GlycoShape database the simulation conditions include a NaCl concentration between 150–200 mM, as generally recommended for MD of physiological conditions (Ross *et al*., 2018), unless weaker or stronger ionic conditions are specifically required. Several online tools, such as SLTCAP(Schmit *et al*. 2018), can assist in calculating the appropriate ion concentration for a given system. Once ions are added, the system is complete and the topology and coordinate files can be generated for the MD simulation to start.

1.2.2.2 Simulation Protocol

Running MD simulations requires a careful system set-up, which can be automated by creating the appropriate input files and submission scripts to run progressively every stage of the simulation. The input files define the simulation control parameters necessary for the

simulation, including time step, simulation length, ensemble type (NVT, NPT, or NVE), position or distance restraints or constraints, temperature and pressure coupling and output frequency. To assist with the initial setup, template files and tutorials are available on the AMBER website(https://ambermd.org/), along with a comprehensive manual that provides detailed guidance on fine-tuning the parameters.

The submission files we use are *bash* scripts used to manage job execution on an HPC cluster, requesting the necessary computing resources, including CPU/GPU allocation and wall time, and listing the commands to execute the different simulation steps. Additionally, the scripts handle input file execution, output generation, and error logging for job monitoring and troubleshooting. Computing resources vary across different HPC clusters and should be always chosen based on accurate benchmarking.

Our typical MD simulation protocol begins with an energy minimisation or relaxation step, which aims to find the nearest energy minimum by eliminating steric clashes and high-energy contacts within the system. This phase ensures a stable system that will not "blow up"(Braun *et al*. 2019). Energy minimisation is performed using standard optimisation algorithms, such as the steepest descent (SD) method. In most of my simulations I set the SD to run for 500,000 cycles with all heavy atoms restrained in their starting positions, leaving only the water molecules, counterions, and hydrogen atoms free. As a note the hydrogen atoms are restrained through SHAKE to allow us a longer integration step, as discussed earlier.

After the energy minimisation phase, the system undergoes a short equilibration phase to bring it to room temperature (298 K or 300 K) and atmospheric pressure (1 atm or 1 bar). In AMBER the equilibration is performed in two stages: 1) an NVT equilibration where the system is brought up to constant room temperature, followed by 2) an NPT equilibration where the pressure is set to atmospheric pressure, matching standard laboratory or physiological conditions. In the simulation protocol I followed, the temperature is incrementally increased to 300 K in two steps over 1 ns using a Langevin thermostat. Meanwhile, to regulate the pressure up to 1 atm I used a Berendsen barostat during 1 ns of simulation.

Once the system has reached equilibrium values of temperature and pressure, the most difficult part of the simulation begins, which corresponds to a conformational equilibration, i.e. bringing the system to a conformational state that is realistic or near the target state we would like to

draw observables from. To achieve this goal the positional restraints on the heavy atoms need to be removed either abruptly or gradually, depending on the biological system, starting conformation and question asked. Indeed some restraints may need to be maintained through a long equilibration phase to ensure stability. When the system has reached conformational stability, or if you are happy with how it looks, you can start the production phase. This latter phase can be relatively short or extremely long, depending on the stability of the system and the complexity of the information and observables you want to collect.

1.2.3 Practical Tips for MD Simulations

In this section, I will list some practical advice and good habits related to MD simulations that I have learned and made my own over the course of my PhD.

1. **Review the literature carefully**

This may seem obvious, but thoroughly understanding your system is crucial before setting up a simulation. Reviewing the literature helps in making informed decisions during system preparation. A solid grasp of the biomolecule's biological environment, glycosylation patterns, and functional role ensures that the simulation setup is both realistic and scientifically relevant.

2. **Check your files (and do it again!)**

Encountering errors and troubleshooting are natural parts of setting up and running MD simulations. Many issues can be resolved by ensuring that all input files are correctly formatted. This includes verifying that residue nomenclature in the PDB file matches the force field definitions, checking that TER cards are correctly placed, and ensuring that no atoms or linkages are missing. These checks should not be limited to the initial structure. Once the topology file is generated, make sure everything is in order before you start your simulation.

3. **Benchmark your resources to get "the best bang for your buck"**

Benchmarking allows you to optimise the use of your computational resources by assessing the performance of your simulation setup before committing to a full production run. Running short test simulations on different numbers of CPUs/GPUs helps determine the optimal hardware configuration, balancing speed and efficiency. This ensures that computational resources are used effectively, minimising queue times on HPC systems and avoiding

unnecessary delays in data collection. On top of that, benchmarking provides an estimate of simulation speed, allowing you to adjust production times accordingly.

4. **You may need to release the positional restraints gradually**

When working with large systems containing multiple components or subunits, it is crucial to remove constraints stepwise rather than all at once. Sudden removal of restraints can lead to instability, high-energy fluctuations and system collapse. A gradual approach allows the system to adjust naturally, ensuring the stability of the system and minimising the risk of system failure.

5. **Backup your data regularly, you won't regret it!**

Ensure you are backing up your data often. MD simulations generate large amounts of data, and most HPC clusters have limited storage quotas or may periodically purge old files. It is good practice to store essential input files, scripts, and key output files on external drives, cloud storage, or institutional repositories.

6. **Check the output of your running simulations**

Do not assume that a simulation is running correctly just because it is producing output files. Regularly inspect the trajectory or the structure files it produces to ensure that everything is running smoothly.

7. **Know what information you are looking for**

"How long should I run a simulation for?" This is probably one of the most frequently asked questions when setting up an MD simulation. While it might be tempting to dismiss the response with, *"How long is a piece of string?"*, the answer is not straightforward as there is no one-size-fits-all approach. The duration of a simulation depends on the system being studied, the biological process being observed and the timescale needed to observe meaningful changes. The simulation time must be long enough to capture relevant events and to see those happening again and again withing uncorrelated time frames. Short simulations may be sufficient to observe conformational transitions occurring at room temperature, i.e. where the thermal energy accessible is sufficient to promote interconversion between states. The simulation of large-scale conformational changes, of ligand binding events, or of protein-protein interactions may require microsecond- or even millisecond-long simulations which are simply inaccessible through a deterministic approach and require enhanced sampling or non-

equilibrium approaches. The size and complexity of the system also matters as larger, more complex systems may need longer simulations to ensure adequate sampling. Alternatively, within the deterministic sampling scheme a useful strategy is to run multiple independent MD replicas with different initial velocities, and different starting structures of the same system allowing for better statistical averaging and the reproducibility of simulation results.

Defining the specific goals of the simulation before starting the simulation is crucial. Asking questions such as, "what do I hope to observe?", "is the simulation the primary focus of my study, or is it intended to support and validate other findings?", "has previous work produced similar results?". Answering these questions will help you assessing the required simulation length and determine whether enhanced sampling techniques may be necessary. If experimental data is available or can be collected, comparing simulated observables to experimental results can provide additional validation and strengthen the reliability of the findings.

Ultimately, simulations as every experiment represents a model. All models are reductive and flawed, but some are useful. A well-thought and run MD simulation can be extremely valuable to provide insight unavailable in any other manner and to build new hypothesis or refine old ones for further investigation. To be useful MD simulations need to be interpreted within their limitations.

1.3 Aims and Objectives

In this work, I performed molecular dynamics (MD) simulations, discussed in detail in section **1.2,** to study the structure and dynamic of glycan structures in the context of different biological processes. In **Chapter 2,** I will be discussing the structure and dynamic of the glycan structures I contributed to the GlycoShape Glycan 3D Structure database(https://glycoshape.org)(Ives *et al*. 2023) developed and curated in our research group. In particular, I highlighted the difference in conformational diversity and behavioural patterns between distinct groups of glycans, namely triantennary *N*-glycans, ABH blood group antigens, Lewis antigens and α(2-8)-linked polysialic acid structures.

I present the work I have done on the structural characterization of triantennary *N*-glycans in **Chapter 3,** making a direct comparisons to biantennary *N*-glycan structures(Harbison *et al*.

2019) and bisected *N*-glycan structures. Using extensive sampling through deterministic MD simulations, I was able to understand how bisection and addition of the third arm affects the architecture relative to the biantennary *N*-glycan system. Structural alignments supplemented by MD simulations revealed how bisection hinders binding to FUT8 and b4GalT1. This disruption prevents maturation through core fucosylation and galactosylation of the glycan arms, consistent with findings from previous studies (Järvå *et al*., 2020; Tseng *et al*., 2017; Nakano *et al*., 2019). In contrast, the triantennary structure is a viable substrate for both of these enzymes(Bydlinski *et al*., 2018; Zhao *et al*., 2023; Tseng *et al*. 2017; Tomida *et al*. 2024). Nevertheless, bisected and core-fucosylated structures bearing highly functionalised arms have been observed in immunoglobulins (IgGs), tumor cells, and pregnant bovine cells(Bondt *et al*., 2014; Link-Lenczowski *et al*., 2018; Klisch *et al*., 2008). This led us to propose an alternative pathway for the formation of functionalised bisected structures, where *N*-glycan bisection can occur after maturation, and thus it depends on the locus of expression of beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (GNT-3 or MGAT3) within the Golgi apparatus. We are unable to test this hypothesis by means of molecular simulations as currently there are no available 3D structures of MGAT3 in the Protein Databank (PDB; www.rcsb.org) and the models predicted by AlphaFold 3 (AF3; https://alphafoldserver.com/) are unsuitable for MD in the absence of further mechanistic and structural information. To circumvent this obstacle, we are collaborating with Dr Andrea Maggioni and Prof Daniel Kolarich at the Institute for Biomedicine and Glycomics at Griffith University, QLD, Australia, who (to date) have successfully expressed a soluble construct of human MGAT3 and are currently investigating its activity through glycomics profiling by mass spectrometry (MS).

In **Chapter 4**, I explored the bacterial adhesion of glycans through the Type IV pili(T4P) of *Neisseria meningitidis* (*Nm*), a Gram-negative bacterium known to cause meningococcal disease. The T4P of *Nm* is a filamentous protein that extends out of the bacterial outer membrane and plays a crucial role in adhesion and mobility (Craig *et al*. 2019). Given its carbohydrate-rich surface(Mubaiwa *et al*. 2017) and interactions with host glycan structures(Gasparini *et al*. 2015), I investigated whether *Nm* could use its T4P as an adhesin to interact with host glycan structures for attachment and whether environmental glycan structures could aid in colonisation. To this end, I reconstructed a 26-pilin subunit long T4P and used it as a "virtual array" system to test its glycan specificity for various glycan structures among those available in the epithelial cells surface environment. All glycan structures I tested are now available in the Glycoshape 3D database. The results from the extensive MD

simulation, showed how the T4P model was able to create potential multi-subunit carbohydrate binding pockets between the globular domains of adjacent pilin subunits and how this unique arrangement facilitates a remarkably high avidity for glycan interactions. Analysis of the MD data also indicated a preference for sialylated glycans and in particular for α2,8-linked polysialic acid with a minimum motif of a trimer. This result was particularly interesting because the capsular polysaccharide of *Nm* serogroup B, which I chose as my structural and sequence T4P model, consists of repeating α2,8-linked Neu5Ac, suggesting that the capsular polysaccharide may interact with T4P to aid colonisation. Glycan binding was primarily mediated by polar residues conserved within the proposed binding pocket, with additional contributions from post-translational modifications such as phosphorylcholine (ChoP) and the O-linked bacterial trisaccharide Galβ1-4Galα1-3DATDH.

In **Chapter 5**, I extended this study to another member of the *Neisseria* family, *Neisseria gonorrhoeae* (*Ng*), to determine whether this behaviour is conserved across species and also in view of the experience and expertise in *Ng* infection of our collaborators in Prof. Michael Jennings' group at the Institute for Glycomics at Griffith University, Gold Coast, QLD, Australia. The results showed that the T4P systems of both *Ng* and *Nm* displayed similar binding patterns and specificities, with many of the residues responsible for binding in the *Nm* pilin system conserved in the *Ng* pilin system. I was able to confirm by SPR that both *Nm* and *Ng* T4P systems have a distinct preference for α2,8-linked polysialic acids. As for *Nm* the *Ng* pilin mutants lacking bacterial glycosylation displayed an increase in binding affinity across all tested glycans compared to wild-type pilin.

Ultimately, my work contributes important atomistic-level insight into the role of glycan 3D structure in fundamental molecular recognition events, with examples from enzyme driven biosynthetic processes to bacterial infection. The results of the studies I present in this thesis all support a paradigm where glycan sequence and branching function as molecular descriptors within a 'glycocode', with the reading key required to decipher the biological messages lying in the 3D structure and dynamics that these sequences and branching uniquely determine.

References:

1. Agirre, J. *et al*., 2015. Carbohydrate anomalies in the PDB. *Nature chemical biology*, 11(5), p.303.
2. Agirre, J., 2017. Strategies for carbohydrate model building, refinement and validation. *Acta crystallographica. Section D, Structural biology*, 73(Pt 2), pp.171–186.
3. Alocci, D. *et al*., 2019. GlyConnect: Glycoproteomics goes visual, interactive, and analytical. *Journal of proteome research*, 18(2), pp.664–677.
4. Angulo, A. *et al*., 2022. Rediscovery of the endangered Amphilophus lyonsi (Gosse 1966) (Cichliformes: Cichlidae) in Costa Rica; extended diagnosis and description and new ecological data. *Zootaxa*, 5182(6), pp.528–540.
5. Angulo, J. *et al*., 2022. Structural biology of glycan recognition. *Essentials of Glycobiology, 4th edition.*
6. Åqvist, J. *et al*., 2004. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical physics letters*, 384(4-6), pp.288–294.
7. Aytenfisu, A.H., Yang, M. & MacKerell, A.D., Jr, 2018. CHARMM Drude polarizable force field for glycosidic linkages involving pyranoses and furanoses. *Journal of chemical theory and computation*, 14(6), pp.3132–3143.
8. Atanasova, M., Bagdonas, H. & Agirre, J., 2020. Structural glycobiology in the age of electron cryo-microscopy. *Current opinion in structural biology*, 62, pp.70–78.
9. Berendsen, H.J.C. *et al*., 1984. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8), pp.3684–3690.
10. Berman, H.M. *et al*., 2000. The Protein Data Bank. *Nucleic acids research*, 28(1), pp.235–242.
11. Braun, E. *et al*., 2019. Best practices for foundations in molecular simulations [article v1.0]. *Living journal of computational molecular science*, 1(1). Available at:

http://dx.doi.org/10.33011/livecoms.1.1.5957.

12. Brockhausen, I. *et al*., 2022. O-GalNAc glycans. In *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 117–128.
13. Brooks, C.L., III, Berkowitz, M. & Adelman, S.A., 1980. Generalized Langevin theory for many-body problems in chemical dynamics: Gas-surface collisions, vibrational energy relaxation in solids, and recombination reactions in liquids. *The Journal of chemical physics*, 73(9), pp.4353–4364.
14. Burley, S.K. *et al*., 2022. Protein Data Bank: A comprehensive review of 3D structure holdings and worldwide utilization by researchers, educators, and students. *Biomolecules*, 12(10), p.1425.
15. Čaval, T., Heck, A.J.R. & Reiding, K.R., 2021. Meta-heterogeneity: Evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Molecular & cellular proteomics: MCP*, 20, p.100010.
16. Chan, L., Hutchison, G.R. & Morris, G.M., 2021. Understanding ring puckering in small molecules and cyclic peptides. *Journal of chemical information and modeling*, 61(2), pp.743–755.
17. Chugh, S. *et al*., 2015. Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochimica et biophysica acta*, 1856(2), pp.211–225.
18. Craig, L., Forest, K.T. & Maier, B., 2019. Type IV pili: dynamics, biophysics and functional consequences. *Nature reviews. Microbiology*, 17(7), pp.429–440.
19. Cruces Chamorro, V., Jungwirth, P. & Martinez-Seara, H., 2024. Building water models compatible with charge scaling molecular dynamics. *The journal of physical chemistry letters*, 15(10), pp.2922–2928.
20. Fadda, E., 2022. Molecular simulations of complex carbohydrates and glycoconjugates. *Current opinion in chemical biology*, 69(102175), p.102175.

21. Fadda, E., Makshakova, O. & Perez, S., 2024. Understanding glycobiology through multiscale molecular dynamics simulations. In *Translational Glycobiology in Human Health and Disease*. Elsevier, pp. 379–396.
22. Gagneux, P., Hennet, T. & Varki, A., 2022. Biological functions of glycans. In *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 79–92.
23. Gasparini, R. *et al*., 2015. How the Knowledge of Interactions between Meningococcus and the Human Immune System Has Been Used to Prepare Effective Neisseria meningitidis Vaccines. *Journal of immunology research*, 2015, p.189153.
24. Gimeno, A. *et al*., 2020. Glycan structures and their interactions with proteins. A NMR view. *Current opinion in structural biology*, 62, pp.22–30.
25. Guvench, O. *et al*., 2011. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *Journal of chemical theory and computation*, 7(10), pp.3162–3180.
26. Harbison, A.M. *et al*., 2019. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology*, 29(1), pp.94–103.
27. He, M., Zhou, X. & Wang, X., 2024. Glycosylation: mechanisms, biological functions and clinical implications. *Signal transduction and targeted therapy*, 9(1), p.194.
28. Hess, B. *et al*., 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12), pp.1463–1472.
29. Ives, C.M. *et al*., 2023. Restoring Protein Glycosylation with GlycoShape. *bioRxiv*
30. Jorgensen, W.L. *et al*., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), pp.926–935.
31. Jumper, J. *et al*., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp.583–589.

32. Kiessling, L.L., Gestwicki, J.E. & Strong, L.E., 2000. Synthetic multivalent ligands in the exploration of cell-surface interactions. *Current opinion in chemical biology*, 4(6), pp.696–703.

33. Kim, P.-J., Lee, D.-Y. & Jeong, H., 2009. Centralized modularity of N-linked glycosylation pathways in mammalian cells. *PloS one*, 4(10), p.e7317.

34. Kirschner, K.N. *et al*., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates: GLYCAM06. *Journal of computational chemistry*, 29(4), pp.622–655.

35. Kumlin, U. *et al*., 2008. Sialic acid tissue distribution and influenza virus tropism. *Influenza and other respiratory viruses*, 2(5), pp.147–154.

36. Laine, R.A., 1994. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, 4(6), pp.759–767.

37. Lee, T.-S. *et al*., 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. *Journal of chemical information and modeling*, 58(10), pp.2043–2050.

38. Lin, F.-Y. & MacKerell, A.D., Jr, 2019. Force fields for small molecules. *Methods in molecular biology (Clifton, N.J.)*, 2022, pp.21–54.

39. Maier, J.A. *et al*., 2015. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation*, 11(8), pp.3696–3713.

40. Mubaiwa, T.D. *et al*., 2017. The glycointeractome of serogroup B Neisseria meningitidis strain MC58. *Scientific reports*, 7(1), p.5693.

41. Nagae, M. & Yamaguchi, Y., 2012. Function and 3D structure of the N-glycans on glycoproteins. *International journal of molecular sciences*, 13(7), pp.8398–8429.

42. Nairn, A.V. *et al*., 2008. Regulation of glycan structures in animal tissues. *The journal of biological chemistry*, 283(25), pp.17298–17313.

43. Nester, K., Gaweda, K. & Plazinski, W., 2019. A GROMOS force field for furanose-based carbohydrates. *Journal of chemical theory and computation*, 15(2), pp.1168–1186.

44. Nielsen, M.I. *et al*., 2022. Global mapping of GalNAc-T isoform-specificities and O-glycosylation site-occupancy in a tissue-forming human cell line. *Nature communications*, 13(1), p.6257.

45. Nosé, S., 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics*, 81(1), pp.511–519.

46. Onufriev, A.V. & Izadi, S., 2018. Water models for biomolecular simulations: Water models for biomolecular simulations. *Wiley interdisciplinary reviews. Computational molecular science*, 8(2), p.e1347.

47. Parrinello, M. & Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of applied physics*, 52(12), pp.7182–7190.

48. Perez, S., Fadda, E. & Makshakova, O., 2021. Computational Modeling in Glycoscience. In J. J. Barchi, ed. *Comprehensive Glycoscience*. Elsevier, pp. 374–404.

49. Perez, S. & Makshakova, O., 2022. Multifaceted computational modeling in glycoscience. *Chemical reviews*, 122(20), pp.15914–15970.

50. Pérez, S. & Tvaroška, I., 2014. Carbohydrate-protein interactions: molecular modeling insights D. Horton, ed. *Advances in carbohydrate chemistry and biochemistry*, 71, pp.9–136.

51. Piana, S. *et al*., 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The journal of physical chemistry. B*, 119(16), pp.5113–5123.

52. Quintana, J.I. *et al*., 2023. Exploring multivalent carbohydrate-protein interactions by

NMR. *Chemical Society reviews*, 52(5), pp.1591–1613.

53. Raposo, C.D., Canelas, A.B. & Barros, M.T., 2021. Human lectins, their carbohydrate affinities and where to find them. *Biomolecules*, 11(2), p.188.
54. Reily, C. *et al*., 2019. Glycosylation in health and disease. *Nature reviews. Nephrology*, 15(6), pp.346–366.
55. Rini, J., Esko, J. & Varki, A., 2009. Glycosyltransferases and glycan-processing enzymes. In *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
56. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C., 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3), pp.327–341.
57. Salisburg, A.M. *et al.* (2009) 'Ramachandran-type plots for glycosidic linkages: Examples from molecular dynamic simulations using the Glycam06 force field', *Journal of computational chemistry*, 30(6), pp. 910–921.
58. Schmit, J.D. *et al*., 2018. SLTCAP: A simple method for calculating the number of ions needed for MD simulation. *Journal of chemical theory and computation*, 14(4), pp.1823–1827.
59. Schnaar, R.L., 2016. Glycobiology simplified: diverse roles of glycan recognition in inflammation. *Journal of leukocyte biology*, 99(6), pp.825–838.
60. Seeberger, P.H., 2015. Monosaccharide diversity. In *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 19–30.
61. Sood, A. *et al*., 2018. Defining the specificity of carbohydrate-protein interactions by quantifying functional group contributions. *Journal of chemical information and modeling*, 58(9), pp.1889–1901.
62. Stanley, P. *et al*., 2022. N-glycans. In *Essentials of Glycobiology*. Cold Spring Harbor

(NY): Cold Spring Harbor Laboratory Press, pp. 103–116.

63. Taylor, M.E. *et al*., 2022. Discovery and classification of glycan-binding proteins. In *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 375–386.
64. Thieker, D.F. *et al*., 2016. 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology*, 26(8), pp.786–787.
65. Tiemeyer, M. *et al*., 2017. GlyTouCan: an accessible glycan structure repository. *Glycobiology*, 27(10), pp.915–919.
66. Tomida, S., Nagae, M. & Kizuka, Y., 2024. Distinctive domains and activity regulation of core fucosylation enzyme FUT8. *Biochimica et biophysica acta. General subjects*, 1868(4), p.130561.
67. Tseng, T.-H. *et al*., 2017. Substrate preference and interplay of fucosyltransferase 8 and N-acetylglucosaminyltransferases. *Journal of the American Chemical Society*, 139(28), pp.9431–9434.
68. Varadi, M. *et al*., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1), pp.D439–D444.
69. Varki, A. & Sharon, N., 2009. Historical background and overview. *Essentials of Glycobiology. 2nd edition*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK579927/.
70. Vega, C. & Abascal, J.L.F., 2011. Simulating water with rigid non-polarizable models: a general perspective. *Physical chemistry chemical physics: PCCP*, 13(44), pp.19663–19688.
71. Verlet, L., 1967. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1), pp.98–103.

72. Wang, J. *et al*., 2004. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9), pp.1157–1174.

73. Wilkinson, H. & Saldova, R., 2020. Current methods for the characterization of O-glycans. *Journal of proteome research*, 19(10), pp.3890–3905.

74. York, W.S. *et al*., 2020. GlyGen: Computational and informatics resources for glycoscience. *Glycobiology*, 30(2), pp.72–73.

75. Zacchi, L.F. & Schulz, B.L., 2016. N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconjugate journal*, 33(3), pp.359–376.

76. Zeng, X. *et al*., 2012. Carbohydrate-protein interactions and their biosensing applications. *Analytical and bioanalytical chemistry*, 402(10), pp.3161–3176.

Chapter 2. My Contribution to the GlycoShape Glycan 3D Structures Database

2.1 Introduction

Glycosylation is one of the most common protein post-translational modifications (PTM). The term glycosylation encompasses a great variety of molecular structures and complexities that can occur co- and/or post-translationally. The two most common types of glycosylation in eukaryotes are *N*- and *O*-glycosylation. *N*-glycosylation refers to the covalent functionalization of the amide N atom of specific Asn residues within the protein sequence with a glycan, whereas in *O*-glycosylation the covalent linkage is to the O atom of the hydroxyl group of Ser or Thr residues(Varki *et al*. 2022).

Protein glycosylation fulfils many biological functions across different systems and organisms, including but not limited to roles in protein folding, stability, cell adhesion and signalling(He *et al*. 2024; Varki & Gagneux 2017). Changes in glycosylation can also be a very useful diagnostic of various diseases, where abnormal glycosylation patterns may be indicative of disease state in cells(Reily *et al*. 2019). For example, the upregulation of sialyltransferases leading to an increase in sialylation of glycoconjugates has been linked to cancer(He *et al*. 2024; Pinho & Reis 2015). Sialic acids have been reported to aid immune evasion by acting as a molecular mask to avoid immune recognition and surface-exposed sialylated glycans such as sialyl-Lewis antigens and polysialic acids have been recognised as biomarkers for various cancers(He *et al*. 2024; Pinho & Reis 2015; Munkley 2022; Schauer 1985)**.** An increased expression levels of fucosyltransferases, such as FUT8, which catalyses the transfer of a fucose to the core of an *N*-glycan (i.e. -α(1-6)Fuc), and of specific GlcNAc transferases, such as MGAT3, which transfers GlcNAc to the central mannose (i.e. -β(1-4)GlcNAc) of the *N*-glycan pentasaccharide scaffold, have also been linked to cancer progression. We will discuss these enzymes in further detail in **Chapter 3.**

Unlike proteins, glycan structures are not directly encoded in the genome, and glycosylation is a highly diverse and dynamic process. The nature of glycan biosynthesis determines the characteristic micro- and macro-heterogeneity of glycans, which refers to the variety of glycan structures present concurrently at the same glycosylation sites, and the different degrees of

occupancy of those sites, respectively(Trbojević-Akmačić *et al*. 2022; Zacchi & Schulz 2016). Glycosylation is also susceptible to environmental conditions(Varki & Sharon 2009).

Glycan 3D structures are highly flexible because of the chemical nature of the glycosidic linkages. This feature makes the experimental characterisation of their structure highly challenging even in cryogenic settings(Nagae & Yamaguchi 2012; Angulo *et al*. 2022; Gimeno *et al*. 2020). The difficulty in resolving the glycans 3D structures poses a significant challenge to understanding their functions in different environments(Esmail & Manolson 2021)**.** Within this context, the advancement of glycoinformatics tools and databases(Alocci *et al*. 2019; York *et al*. 2020; Tiemeyer *et al*. 2017) along with high-performance computing (HPC) molecular simulations(Fadda 2022) have played a significant role in addressing this challenge, bridging the gap where traditional experimental techniques fall short(Malhotra & Ramsland 2020).

To aid this effort our lab has developed an open access (OA) glycan 3D structure database named GlycoShape (https://glycoshape.org) that contains structural information collected by our research group through molecular dynamics (MD) simulations. The database holds data of over 640 glycan structures to date, corresponding to over 1 ms of cumulative sampling through uncorrelated replicas of deterministic MD simulations. Throughout my PhD, I have consistently contributed to the database by characterising the structure and dynamics of glycan structures directly connected to different projects I have been working on, as well as other structures requested by users or otherwise needed. I contributed to the database a total of 121 glycan 3D structures. In this Chapter, I will present the results of this work focusing on the main classes of glycans I built and analysed, namely triantennary *N*-glycans, the ABH and Lewis blood group antigens, and polysialic acids. I will discuss the projects where these glycans are implicated in the following Chapters.

2.2 Methods

2.2.1 Molecular Dynamics

The 3D starting structures of all glycans were built with the carbohydrate builder tool on GLYCAM-WEB (https://glycam.org/). Conformational sampling is based on a minimum of three replicas for each glycan structure. For glycan structures that theoretically corresponded to two energetically available conformers, I used two replicas for each conformer. For

structures that theoretically corresponded to three or more energetically available conformers, we used one replica for each conformer.

Topology files were built with the *tleap* program included in the AMBER 18(Lee *et al*. 2018) software package. The carbohydrate atoms were represented using the GLYCAM06j-1version of the GLYCAM06 force field(Kirschner *et al*. 2008), while the counterions were represented using the AMBER ff14SB(Maier *et al*. 2015). Water molecules were represented using the TIP3P water model(Jorgensen *et al*. 1983). All MD simulations were run in 200 mM NaCl concentration using AMBER18(Lee *et al*. 2018) or GROMACS 2022.4(GMX)(Abraham *et al*. 2015) depending on the HPC infrastructure available.

All AMBER18 simulations started with an energy minimisation phase of 500,000 steps of steepest descent. During minimisation only hydrogen atoms, water molecules and counterions were left unrestrained, while the position of all other atoms was kept restrained with a force constant of 5 kcal/mol·Å^{-2}. Following minimisation, the system was brought up to standard temperature through a heating phase in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During phase one the temperature was raised from 0 to 100 K where the volume was kept constant and from 100 to 300 K during phase two where the pressure was kept constant. The system was then equilibrated for 1 ns in the NPT ensemble to a 1 am pressure with the Berendsen barostat. During these equilibration stages, we used the same position restraints we used during the energy minimisation. A conformational equilibration was performed for 10 ns with all positional restraints removed. The production phase followed 500 ns for each replica of the glycan structure with all positional restraints removed.

For the GMX simulations the Amber topology (.prm7) and structure (.rst7) files were converted to the corresponding GMX topology (.top) and structure (.gro) files using ACPYPE(Bernardi *et al*. 2019). The GMX simulations protocol is analogous to the one we used to set up the AMBER18 simulations, starting with an energy minimisation stage through 500,000 steps of steepest descent. Following minimisation, the system was then equilibrated in the NVT ensemble using a Langevin thermostat to reach the equilibrium temperature of 300 K proceeded by equilibration in the NPT ensemble using a Parrinello–Rahman barostat to reach

the equilibrium pressure of 1 bar. Production runs followed where we run a minimum of 500 ns of unrestrained MD trajectories for each structure replica.

2.2.1 Data Processing and Clustering

Upon completion of the production phases, the data is processed and clustered through the Glycan Analysis Pipeline(GAP) as described in the reference(Ives *et al*. 2024). In this context, a cluster represents a set of glycan conformations that occupy a similar region of conformational space, defined primarily by their torsional angles (ϕ, ψ, ω). In this section, I will briefly summarise the workflow of GAP below for completeness.

The production trajectories for each glycan are merged into one dataset. Each frame from the dataset is then transformed into a graph (distance) matrix. The matrices are then transformed into a one-dimensional array by flattening its lower half. The dimensionality of the array is then reduced by principal component analysis (PCA) from the *sklearn* library(Pedregosa *et al*. 2012). The data is then clustered using a Gaussian Mixture Model (GMM) from the *sklearn* library and an optimal number of clusters are returned based on the silhouette score.. The 3D structures associated with these clusters correspond to the corresponding Kernel Density Estimate (KDE) max values and their corresponding torsion angles (ϕ, ψ, ω) can be sourced on GlycoShape (https://glycoshape.org).

2.3 Results

In this section, I will discuss the structure and dynamics of the free glycan structures I built and analysed as a contribution to the GlycoShape Glycan 3D Structures Database, namely the triantennary N-glycan structures, the ABH and Lewis blood group antigens and polysialic acids.

2.3.1 Triantennary *N*-Glycans

The triantennary *N*-glycans structures I analysed all shared a common starting structure which was then functionalised with additions of fucose, galactose and sialic acids residues, see **Figure 2.1**.

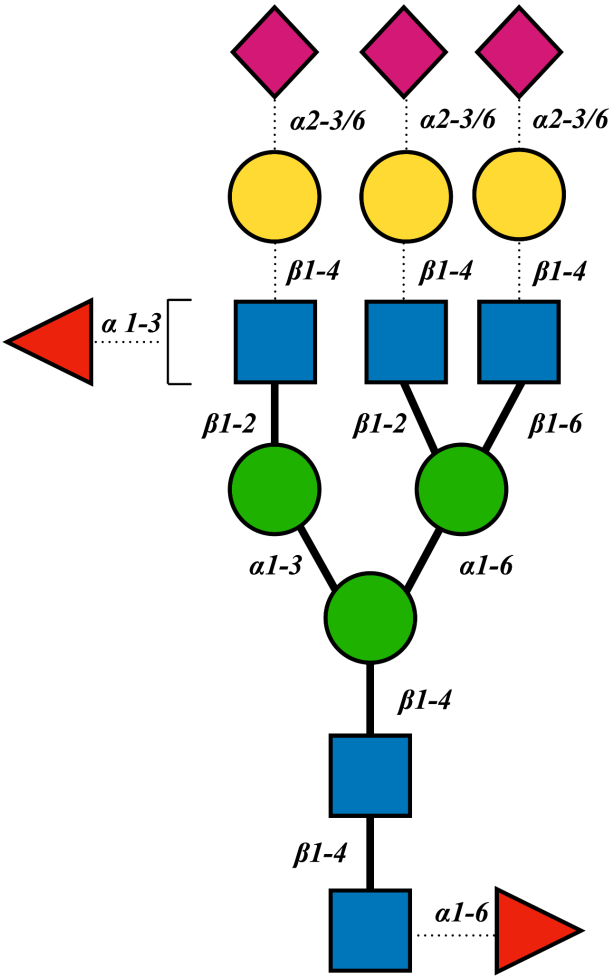


Figure 2.1 Triantennary *N*-glycan structures illustrated using the Symbol Nomenclature for Glycans (SNFG)(Neelamegham *et al*. 2019). The bonds represented with a solid line are common to all structures discussed in this Chapter, while the bonds represented by a dotted line indicate a functionalization that may or may not be present in all structures.

The structure of the triantennary *N*-glycan structures can be discussed as a sum of its parts, or in terms of "glycoblocks", which adopt precise structures with characteristic dynamics and flexibility(Fogarty *et al*. 2020; Fogarty *et al*. 2024), see **Figure 2.2**.

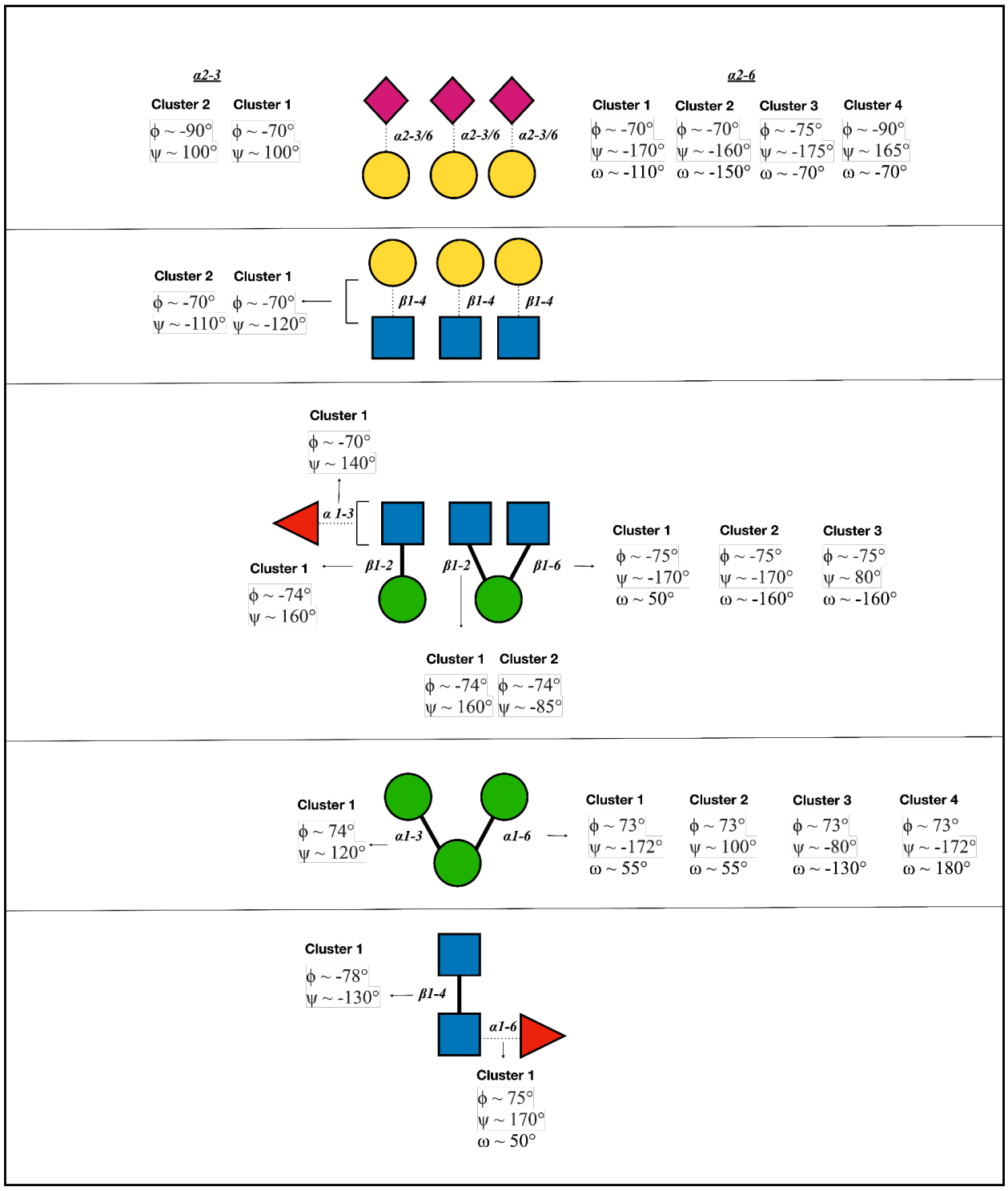


Figure 2.2: Triantennary *N*-glycan structures broken down into different blocks. The values of the torsion angles characterising the conformation of each bond in different clusters are indicated. The '~' symbol indicates an average standard deviation of 10°. The torsion angle values have been sourced from https://glycoshape.org.

The core GlcNAc-β(1-4)-GlcNAc, commonly referred to as chitobiose, is rigid with only minimal torsional flexibility around 10° (Harbison *et al*. 2019). The addition of fucose through an α(1-6) linkage to the reducing GlcNAc does not affect the structure of the

chitobiose(Harbison *et al*. 2019). The α(1-3) arm of the tri-mannosyl core is relatively rigid, while the α(1-6) arm is highly flexible leading to a wide variety of conformations as shown in **Figure 2.3.**

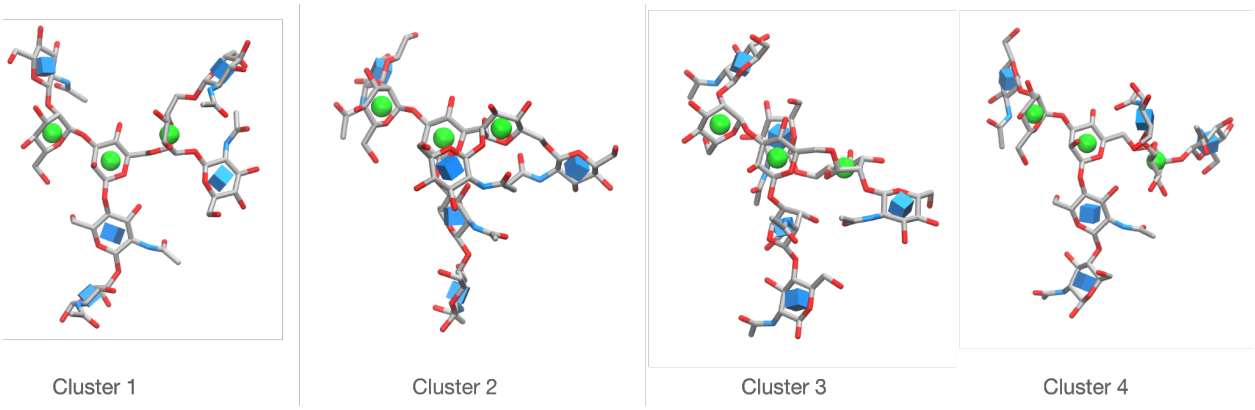


Figure 2.3. Representative triantennary *N*-glycans structures corresponding to the highest populated conformers of the α(1-6) arm, labelled by representative clusters see **Figure 2.2**. Graphical rendering with VMD(Humphrey *et al*. 1996)(https://www.ks.uiuc.edu/Research/vmd/).

For the triantennary structure with each arm terminating in β(1-2)-GlcNAc, the α(1-6) arm predominantly adopts two conformations. The most populated conformation is an open conformation(Harbison *et al*. 2019; Fogarty *et al*. 2024; Fogarty & Fadda 2021), characterised by torsion angle values around the α(1-6) bond of φ ~ 73.0°, ψ ~ -172°, ω ~ 55° (Cluster 1). The second most populated conformation is one in which the β(1-2)-linked GlcNAc on the α(1-6) arm is positioned in a plane-to-plane configuration with the central mannose, a configuration similar to the "front fold" conformation described in a previous analysis of the architecture of oligomannose *N*-glycans(Fogarty & Fadda 2021), with torsion angles φ ~ 73°, ψ ~ 100°, and ω ~ 55°(Cluster 2). We also note the presence of two minor confirmations. One corresponds to the 'back fold' conformation identified in the same study (Fogarty & Fadda 2021), with torsion angles φ ~ 73°, ψ ~ -80°, and ω ~ -130°(Cluster 3), where the β(1-6)-linked GlcNAc on the α(1-6) arm adopts a plane-to-plane orientation with the central mannose of the tri-mannosyl core. The second is a variation of the open fold structure, exhibiting torsion angles of φ ~ 73.0°, ψ ~ -172°, and ω ~ 179.0° (Cluster 4), where the ω angle deviates from the 50° seen from the canonical open conformation.

Similarly, the Man-β(1-6)-GlcNAc linkage extending from the α1-6 arm shows a high degree of flexibility owing to the nature of the 1-6 linkage exhibiting a range of distinct conformations, see **Figure 2.4.**

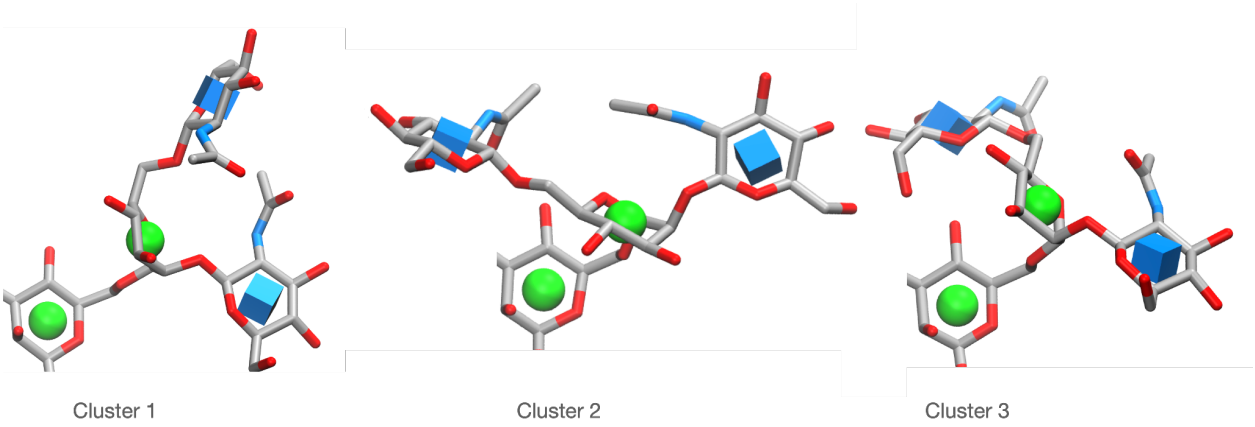


Figure 2.4. Populated conformations of the β6-linked GlcNAc on the α6 arm of A3, labelled by representative clusters see **Figure 2.2**. Rendering with VMD(Humphrey *et al*. 1996)(https://www.ks.uiuc.edu/Research/vmd/).

In the case of each arm terminating in β(1-2)-GlcNAc, the most populated conformation of the β(1-6) arm is characterised by torsion angles of φ ~ -75°, ψ ~ -170°, and ω ~ 50° (Cluster 1). In this arrangement, the arm resembles a 'claw-like' configuration where the β(1-6)-linked GlcNAc faces the β(1-2)-linked GlcNAc on the same arm. The second most populated confirmation has torsion values of φ ~ -75°, ψ ~ -170°, ω ~ -160°(Cluster 2) with the β(1-6)-linked GlcNAc pointing toward the α(1-3) arm. There is also low populated third conformation with torsion angle values of φ ~ -75°, ψ ~ 80°, ω ~ -160° (Cluster 3), in which the β(1-6)-linked GlcNAc is positioned behind the α(1-2)-linked GlcNAc on the same arm.

By contrast, the Man-β(1-2)-GlcNAc linkage on the same arm as well as the Man-β(1-2)-GlcNAc linkage on the α(1-3) arm are relatively rigid. The addition of fucose via an α(1-3) linkage to the GlcNAc residues extending from the tri-mannosyl group consistently exhibits a well-defined set of torsion angles across the different arms and structures. Extension of the GlcNAc arms with Gal through a β(1-4) linkage and a further extension via α(2-3) linkage to Neu5Ac also demonstrated very little flexibility occupying only a distinct set of torsional values, whereas the Neu5Ac linked through an α(2-6) bond is more flexible occupying a wider range of torsion angles.

The extension of the GlcNAc arms with β(1-4)-linked Gal did not alter the conformational preferences of the flexible α(1-6) arm extending from the tri-mannosyl core, nor did it affect the β(1-6)-linked GlcNAc on the α(1-6) arm. However, the addition of terminal Neu5Ac via both α(2-3) and α(2-6) linkages shifted the equilibrium of the α(1-6) arm, favoring the front-

fold conformation (Cluster 2) as the most populated, followed by the open conformation (Cluster 1).

For the β(1-6)-linked GlcNAc on the α(1-6) arm, the addition of terminal Neu5Ac via an α(2-6) linkage maintained the existing conformational preferences. However, the Neu5Ac via an α(2-3) linkage altered the conformational equilibrium, making the conformation corresponding to Cluster 2 the most populated, followed by the conformation corresponding to Cluster 3 and the 'claw-like' configuration (Cluster 1).

2.3.2 Blood Group Antigens

The ABH and Lewis structures were among the first polymorphisms identified on human red blood cells(Stowell & Stowell 2019). These oligosaccharide structures are expressed as part of glycoconjugates on the epithelial surface of various tissues, including the gastrointestinal, urinary, and reproductive tracts(Marionneau *et al*. 2001; Henry 2001). The structures of the ABH blood group antigens from types 1 to 4 I built and analysed for the GlycoShape database are shown in **Figure 2.5.**

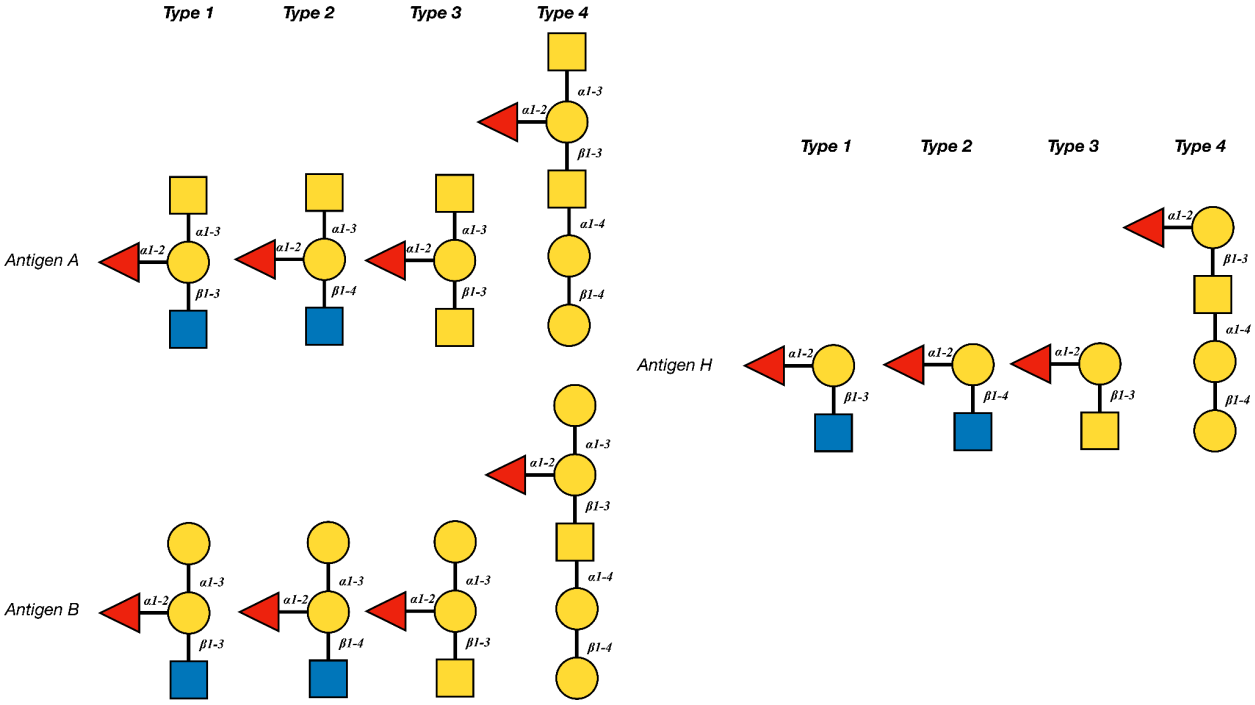


Figure 2.5: SNFG representation of the blood group antigens organised into ABH groups with their corresponding types, type 1 to 4.

The highest populated torsion angle values for the linkages in each cluster are shown in **Figure 2.6.** The most significant result of this analysis is that the blood group antigens are generally rigid 3D structures, occupying a very limited set of spatial conformations characterised by limited mobility across the different linkages.

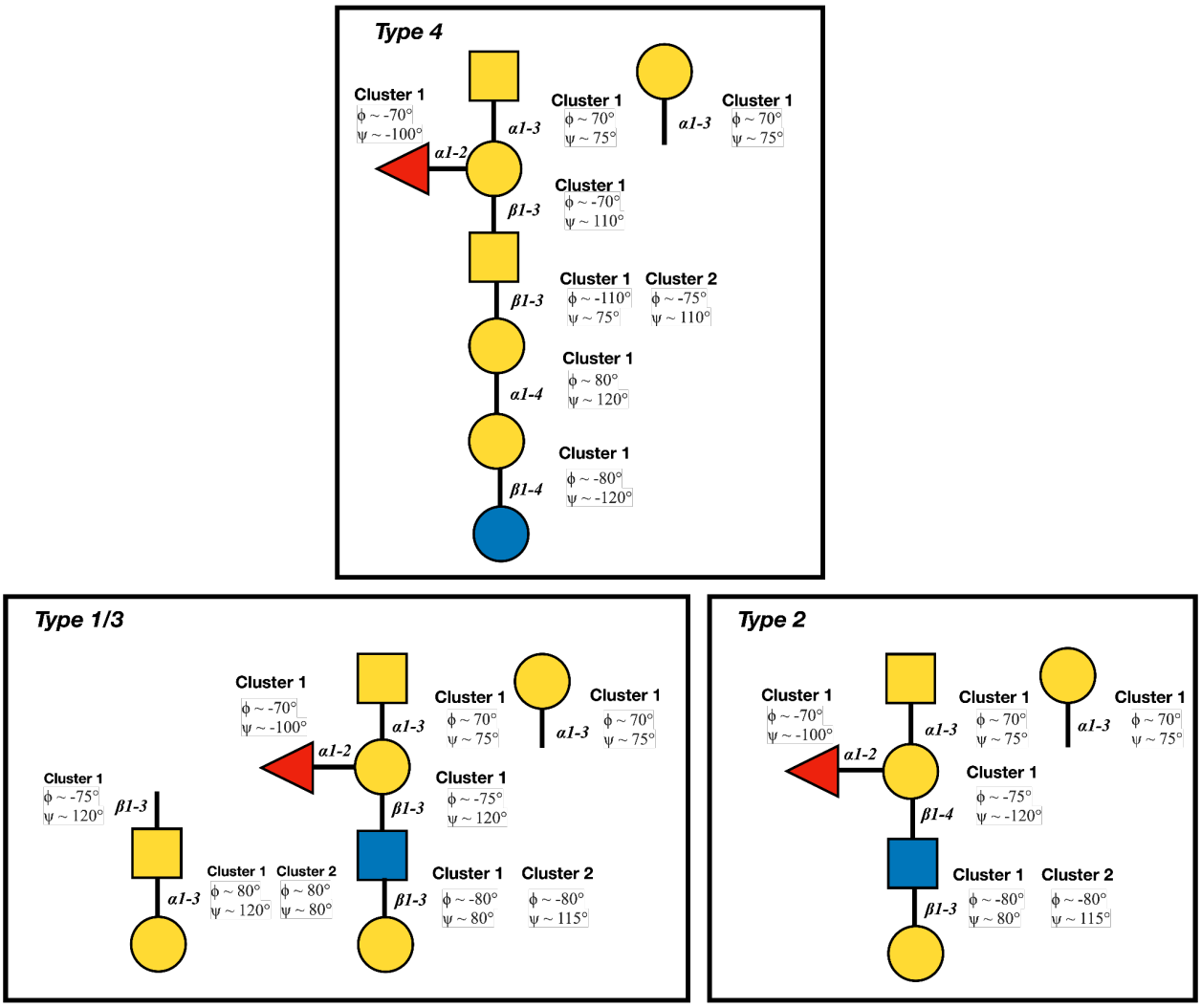


Figure 2.6: SNFG representation of the structure of blood group antigens broken down into different 'glycoblocks'. The highest populated values of the torsion angles for each linkage are shown for each cluster. The torsion angle values have been sourced from https://glycoshape.org.

2.3.3 Lewis Antigens

The highest populated torsion angles obtained for the Lewis antigens are shown in **Figure 2.7.** Similarly to the ABH blood group antigens, the glycosidic bonds within the Lewis antigen structures have very little flexibility with the exception of the α(2-3) linkage to Neu5Ac in both sialyl Lewis-A (sLeA) and sialyl Lewis-X (sLeX). This flexibility arises from the ability of Neu5Ac to interact with the adjacent Gal residue through hydrogen bonding involving its

hydroxyl group. The relatively small size of the Lewis antigens further facilitates this dynamic behavior. This flexibility is not observed in the α(2-3) linkage to the terminal Neu5Ac in the triantennary N-glycan structures due to steric constraints imposed by restricted spatial arrangement of the triantennary structure.

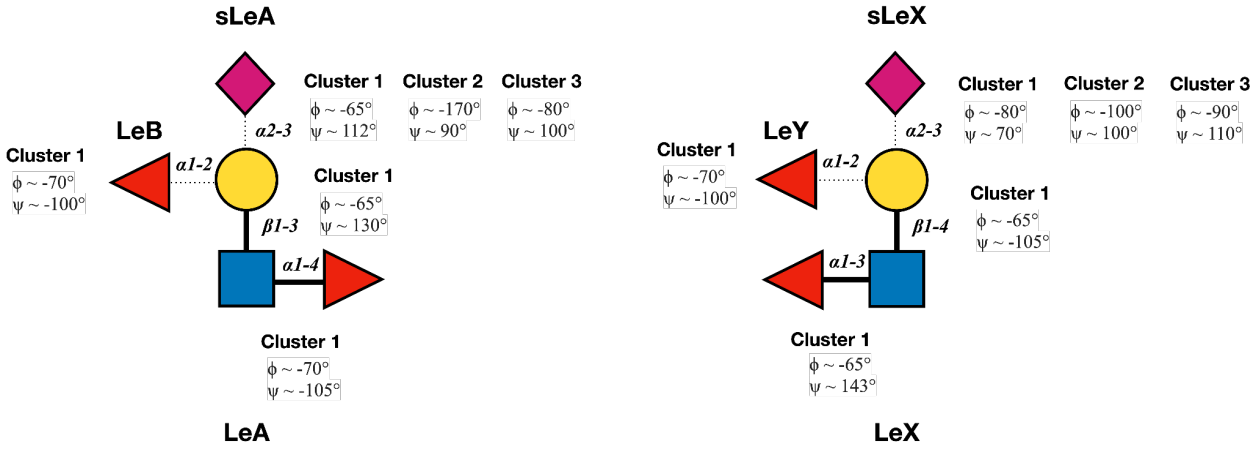


Figure 2.7: The highest populated values of the torsion angles for the linkages for the Lewis antigens are indicated for each cluster. The Lewis-A (LeA) and Lewis-X (LeX) antigens correspond to the SNFG structures connected with the solid bonds. The addition of Neu5Ac results in the sialylated version of the lewis antigen i.e sLeA and sLeX. The addition of an extra fucose via a a(1-2) linkage leads to the Lewis-B (LeB) from LeA and Lewis-Y (LeY) from the LeX. The torsion angle values have been sourced from https://glycoshape.org.

2.3.4 α2,8-linked polysialic acids

I analysed α(2-8)-linked polysialic acids of varying lengths, i.e. trimer, pentamer and octamer, see **Figure 2.8a.** The conformations of all these α(2-8)-linked polysialic acids are similar with remarkably small degrees of flexibility, see **Figure 2.8b**. Indeed, the structure of all α(2-8)-linked polysialic acids is a pseudo-helix with a trimeric unit as a minimal motif, structured as a 'hook', see **Figure 2.8c.** This shape is stabilised by a network of intramolecular hydrogen bond interactions between hydroxyl and amine groups of the different Neu5Acs units, see **Figure 2.8c.**

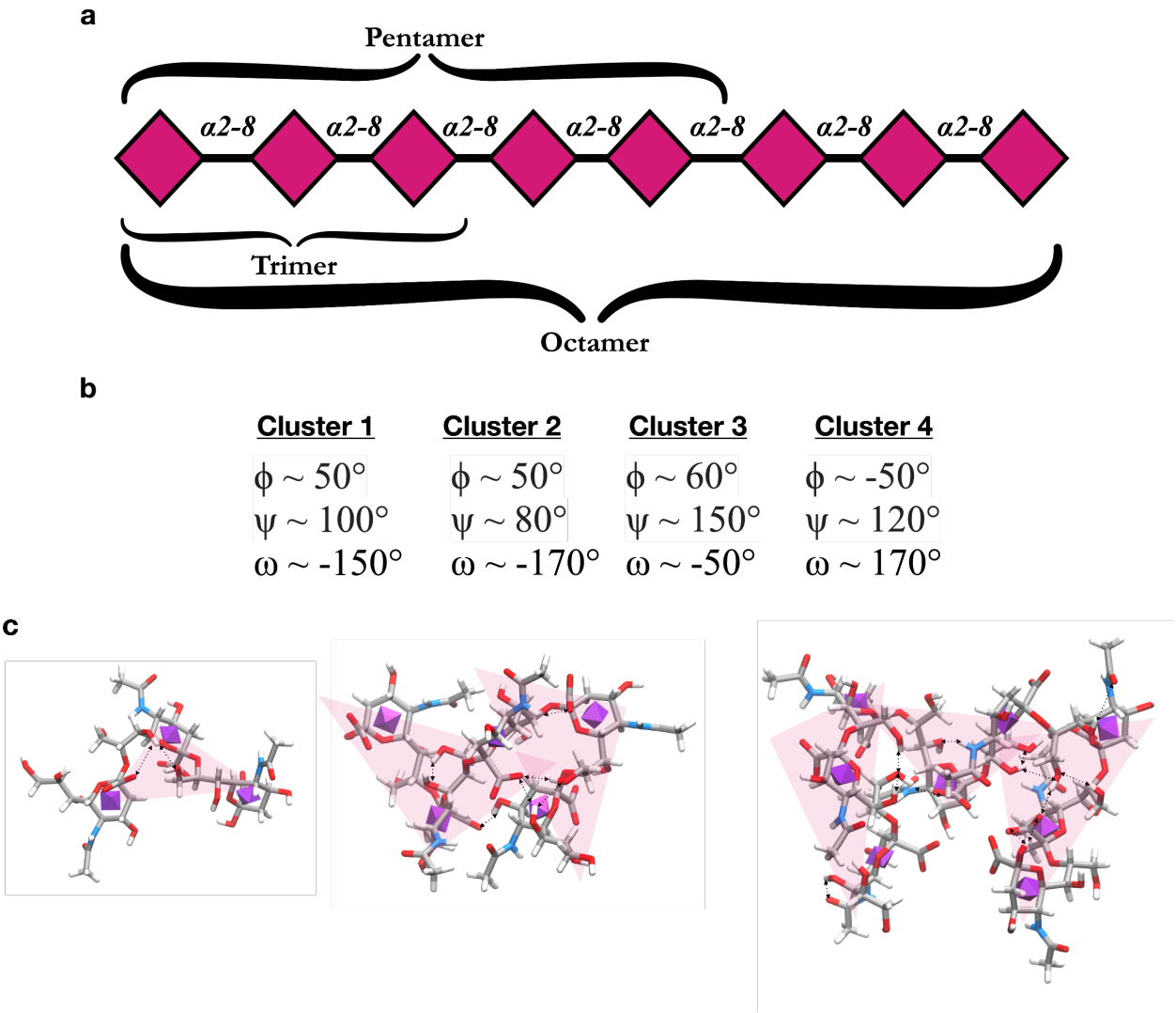


Figure 2.8. Panel a) SNFG representation of the polysialic acid analysed in this work with labels indicating the specific lengths. **Panel b)** Highest populated values of the torsion angles associated with Neu5Ac-α(2-8)-Neu5Ac linkage in each cluster. The torsion angle values have been sourced from https://glycoshape.org. **Panel c)** Highest populated 3D conformations of the trimer, pentamer and octamer of the α(2-8)-linked polysialic acids, from left to right, respectively. Hydrogen bond interactions are indicated with dotted lines. A purple transparent triangle highlights guides the eye to detect the "hook" motif. Graphical rendering with VMD(Humphrey *et al*. 1996)(https://www.ks.uiuc.edu/Research/vmd/).

2.4 Discussion

I analysed the structure and dynamics of triantennary *N*-glycan structures, gaining important insight into the organisation of their structure into "glycoblocks" (Fogarty & Fadda 2021), which allowed me to make direct comparisons with the architecture of biantennary (Harbison *et al*. 2019) and oligomannose (Fogarty & Fadda 2021) *N*-glycan structures. The conformation of the chitobiose in all *N*-glycan structures is similar regardless of the presence of core fucose.

The α(1-3) arm extending from the tri-mannosyl core shows a similar behaviour as in the biantennary system with a relatively contained flexibility. Meanwhile, the α(1-6) arm on biantennary and triantennary structures carries a greater degree of flexibility. In the triantennary system, where the α(1-6) carries two branches, the α(1-6) arm shows an even broader range of conformations ,as shown in **Figures 2.3 and 2.4,** relative to the biantennary *N*-glycans which predominantly occupied an open or closed conformation(Harbison *et al*. 2019) with some of the conformations occupied by the triantennary structure also seen in the study of oligomannose structures such as the "front-fold" and "back-fold" conformations(Fogarty & Fadda 2021). The conformational equilibrium can be modulated through the functionalisation of the arms as observed via the effects of terminal sialylation. In Chapter 4, I will revisit the conformational diversity of triantennary *N*-glycans in comparison to biantennary and bisected N-glycans and explore how these structural variations may influence binding to the enzymes FUT8 and B4GalT1.

The ABH and Lewis antigens were overall rigid 3D structures, with a limited flexibility associated with each of their glycosidic linkages. Across the different blood groups and types, I saw similar torsional values for the same linkages suggesting that much of the conformational variability of these antigens may come down to the size and arrangement of the glycan building blocks as opposed to the inherent flexibility of the linkages themselves.

The analysis of the structure and dynamics of α(2-8)-linked polysialic acids of various lengths, from trimer to octamer, shows that a distinct pseudo-helical structure characterised by a repeating 3D 'hook' motif is shared as structural unit across the different lengths. This 3D motif is stabilised by a network of intramolecular hydrogen bonds between the amide and hydroxyl groups of Neu5Ac units in a N+1 pattern. This behaviour has been seen in previous NMR studies of polysialic acids of various lengths where Neu5Ac units form hydrogen bonds with neighbouring units stabilising a pseudo-helical structure(Mindler *et al*. 2021; Henderson *et al*. 2003; Azurmendi *et al*. 2017). However, there is no consensus to a define conformation on free polysialic acids in solution and several models have been suggested from random coils to variants of a left-handed helical structures(Mindler *et al*. 2021). Interestingly, it has also been reported that polysialic acid structures expressed on the surface of bacterial cells, namely *Neisseria meningitidis*, demonstrate similar structural behaviour to free polysialic acid structures(Azurmendi *et al*. 2007). In **Chapter 4 and 5**, I will characterise the binding

specificities of the pilin protein of *Neisseria meningitidis* and *Neisseria gonorhoeae* for the α2,8-linked polysialic acids as well as the other related glycans.

2.5 Conclusion

In this chapter, I explored the structure and dynamics of triantennary *N*-glycans, ABH blood group antigens, Lewis antigens and α(2-8)-linked polysialic acid structures I built and analysed as a contribution to the of the GlycoShape Glycan 3D Structure database. This analysis provides insights to the conformational diversity of these glycan structures and its dependence on sequence and branching. For the triantennary N-glycan, I identified its distinct "glycoblocks" and highlighted structural similarities and differences compared to the previously studied biantennary and oligomannose N-glycan structures, particularly with respect to the rigidity of the chitobiose core and the flexibility of the α1-6 and β1-6 arms. My analysis of the ABH blood group and Lewis antigens demonstrated limited flexibility within the linkages whereas α2,8-linked polysialic acids showcased a greater flexibility within its linkages but distinct behavioural patterns across different lengths including adopting “hook”-like confirmations stabilised by inter hydrogen bonds. The structural analysis of these glycans provides the basis for the study of their recognition presented in the next Chapters.

References:

1. Abraham, M.J. *et al*., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, pp.19–25.
2. Alocci, D. *et al*., 2019. GlyConnect: Glycoproteomics goes visual, interactive, and analytical. *Journal of proteome research*, 18(2), pp.664–677.
3. Angulo, J. *et al*., 2022. Structural biology of glycan recognition. *Essentials of Glycobiology, 4th edition.*
4. Azurmendi, H. *et al*., 2007. Extracellular structure of polysialic acid explored by on cell solution NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 104(28), pp.11557–11561.
5. Azurmendi, H.F. *et al*., 2017. The β-reducing end in α(2-8)-polysialic acid constitutes a unique structural motif. *Glycobiology*, 27(9), pp.900–911.
6. Bernardi, A. *et al*., 2019. ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. *SoftwareX*, 10(100241), p.100241.
7. Esmail, S. & Manolson, M.F., 2021. Advances in understanding N-glycosylation structure, function, and regulation in health and disease. *European journal of cell biology*, 100(7-8), p.151186.
8. Fadda, E., 2022. Molecular simulations of complex carbohydrates and glycoconjugates. *Current opinion in chemical biology*, 69(102175), p.102175.
9. Fogarty, C.A. *et al*., 2024. Glycoprotein Analysis W. B. Struwe, ed. , 15, pp.315–328.
10. Fogarty, C.A. *et al*., 2020. How and why plants and human N-glycans are different: Insight from molecular dynamics into the “glycoblocks” architecture of complex carbohydrates. *Beilstein journal of organic chemistry*, 16, pp.2046–2056.
11. Fogarty, C.A. & Fadda, E., 2021. Oligomannose N-glycans 3D architecture and its

response to the FcγRIIIa structural landscape. *The journal of physical chemistry. B*, 125(10), pp.2607–2616.

12. Gimeno, A. *et al*., 2020. Glycan structures and their interactions with proteins. A NMR view. *Current opinion in structural biology*, 62, pp.22–30.
13. Harbison, A.M. *et al*., 2019. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology*, 29(1), pp.94–103.
14. He, M., Zhou, X. & Wang, X., 2024. Glycosylation: mechanisms, biological functions and clinical implications. *Signal transduction and targeted therapy*, 9(1), p.194.
15. Henderson, T.J., Venable, R.M. & Egan, W., 2003. Conformational flexibility of the group B meningococcal polysaccharide in solution. *Journal of the American Chemical Society*, 125(10), pp.2930–2939.
16. Henry, S.M., 2001. Molecular diversity in the biosynthesis of GI tract glycoconjugates. A blood-group-related chart of microorganism receptors. *Transfusion clinique et biologique: journal de la Societe francaise de transfusion sanguine*, 8(3), pp.226–230.
17. Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), pp.33–8, 27–8.
18. Ives, C.M. *et al*., 2024. Restoring protein glycosylation with GlycoShape. *Nature Methods*, pp.1–11.
19. Jorgensen, W.L. *et al*., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), pp.926–935.
20. Kirschner, K.N. *et al*., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry*, 29(4), pp.622–655.
21. Lee, T.-S. *et al*., 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. *Journal of chemical information and modeling*, 58(10), pp.2043–2050.

22. Maier, J.A. *et al*., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation*, 11(8), pp.3696–3713.
23. Malhotra, S. & Ramsland, P.A., 2020. Editorial overview: Carbohydrates - structural glycobiology catches the wave of rapid progress. *Current opinion in structural biology*, 62, pp.iii–v.
24. Marionneau, S. *et al*., 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie*, 83(7), pp.565–573.
25. Mindler, K., Ostertag, E. & Stehle, T., 2021. The polyfunctional polysialic acid: A structural view. *Carbohydrate research*, 507(108376), p.108376.
26. Munkley, J., 2022. Aberrant sialylation in cancer: Therapeutic opportunities. *Cancers*, 14(17), p.4248.
27. Nagae, M. & Yamaguchi, Y., 2012. Function and 3D structure of the N-glycans on glycoproteins. *International journal of molecular sciences*, 13(7), pp.8398–8429.
28. Neelamegham, S. *et al*., 2019. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*, 29(9), pp.620–624.
29. Pedregosa, F. *et al*., 2012. Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]*. Available at: http://arxiv.org/abs/1201.0490.
30. Pinho, S.S. & Reis, C.A., 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nature reviews. Cancer*, 15(9), pp.540–555.
31. Reily, C. *et al*., 2019. Glycosylation in health and disease. *Nature reviews. Nephrology*, 15(6), pp.346–366.
32. Schauer, R., 1985. Sialic acids and their role as biological masks. *Trends in biochemical sciences*, 10(9), pp.357–360.

33. Stowell, C.P. & Stowell, S.R., 2019. Biologic roles of the ABH and Lewis histo-blood group antigens Part I: infection and immunity. *Vox sanguinis*, 114(5), pp.426–442.

34. Tiemeyer, M. *et al*., 2017. GlyTouCan: an accessible glycan structure repository. *Glycobiology*, 27(10), pp.915–919.

35. Trbojević-Akmačić, I. *et al*., 2022. High-throughput glycomic methods. *Chemical reviews*, 122(20), pp.15865–15913.

36. Varki, A. *et al*. eds., 2022. *Essentials of glycobiology, fourth edition* 4th ed., New York, NY: Cold Spring Harbor Laboratory Press.

37. Varki, A. & Gagneux, P., 2017. Biological functions of glycans.

38. Varki, A. & Sharon, N., 2009. Historical background and overview. *Essentials of Glycobiology. 2nd edition*.

39. York, W.S. *et al*., 2020. GlyGen: Computational and informatics resources for glycoscience. *Glycobiology*, 30(2), pp.72–73.

40. Zacchi, L.F. & Schulz, B.L., 2016. N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconjugate journal*, 33(3), pp.359–376.

Chapter 3. Effect of bisection on the structure, dynamics, maturation and recognition of *N*-glycans

3.1 Introduction

N-glycosylation is post-translational modification (PTM) that hinges on a highly complex and sophisticated enzymatic machinery operating as proteins are translocated through the secretory pathway. *N*-glycosylation assists protein folding and quaternary assembly in the endoplasmic reticulum (ER), maturing through the Golgi to support trafficking and secretion(Aebi 2013; Stanley *et al*. 2017; Schjoldager *et al*. 2020). This process starts when specific Asn residues within Asn-X(not Pro)-S/T canonical consensus sequons, and to a lesser degree Asn-X-C/V non-canonical sequons(Sun & Zhang 2015), are functionalised by the oligosaccharyltransferase (OST) multiprotein complex(Mohorko *et al*. 2011; Schjoldager *et al*. 2020; Wild *et al*. 2018) with a preformed $(Glc)_3Man9$ structure delivered by a -PP-dolichol-linked donor(Mohorko *et al*. 2011; Wild *et al*. 2018). *N*-glycan maturation is the result of a highly concerted mechanism operated by a multitude of highly specific glycoside hydrolases (GHs) and glycosyltransferases (GTs) that selectively trim-down the highmannose arms through the ER and sequentially functionalise the core, to generate hybrid and/or multi-antennary complex structures through the Golgi(Schjoldager *et al*. 2020; Stanley *et al*. 2017), with varying levels of core and arms fucosylation and terminal sialylation(Pinho & Reis 2015).

The type of *N*-glycosylation occurring in any specific protein is strictly cell-dependent(Wildt & Gerngross 2005; Wilson 2002; Paschinger & Wilson 2019; Strasser 2014; Speciale *et al*. 2017) and tightly controlled as it is crucial to optimal cell function in health and disease(Taniguchi & Kizuka 2015; Varki 2017; Cobb 2020). Yet, as the process is not template-driven, the type of *N*-glycans at each site can be different, with a degree of heterogeneity that depends on the levels of expression of the enzymes responsible for specific branching and functionalisation, and on the degree of accessibility of the glycosylation site on the protein surface(Thaysen-Andersen & Packer 2012; Fogarty & Fadda 2021; Riley *et al*. 2019; Lee *et al*. 2014). Therefore, in principle, as a protein goes through the secretory pathway, an *N*-glycan in a highly accessible site on the protein surface will gradually mature to reach complete functionalisation, with populations of structures corresponding to a normal distribution, reflecting the stochastic nature of the process. The peak and width of such

distribution is determined primarily by the levels of expression of the GHs and GTs encoded in each specific cell line, among other environmental and operational parameters that can affect thermodynamics and kinetics.

Complex *N*-glycans are the product of the highest degree of functionalisation within the *N*-glycosylation pathway(Stanley *et al*. 2017). In mammalian cells complex *N*-glycans develop from a biantennary architecture, obtained from Man5 by the addition of a GlcNAc in β2 by *N*-acetyl-glucosamine transferase 1 (GlcNAc-T1 or MGAT1, as the encoding gene) on the α3 Man, see **Figure 3.1**, followed by the excision of the two Man on the α6 arm by α-mannosidases, and the subsequent addition of the GlcNAc in β2 by *N*-acetyl-glucosamine transferase 2 (GlcNAc-T2 or MGAT2) on the α6 Man in the medial Golgi(Schachter 1986). As GlcNAc-T1 initiates hybrid and complex *N*-glycan synthesis, its inhibition prevents further functionalisation, determining an homogeneous oligomannose-type *N*-glycosylation(Stanley *et al*. 2017; Gleeson & Schachter 1983). The addition of a third antenna requires the presence of GlcNAc-β2-Man on both arms of a biantennary (F)A2 system(Gleeson & Schachter 1983) and is initiated by the transfer of GlcNAc in β4 on the α3 arm by *N*-acetyl-glucosamine transferase 4 (GlcNAc-T4 or MGAT4), or by the transfer of GlcNAc in β6 on the α6 arm by *N*-acetyl-glucosamine transferase 5 (GlcNAc-T5 or MGAT5), see **Figure 3.1**. Both of these triantennary structures can evolve into tetra-antennary architectures through the sequential action of GlcNAc-T4 and T5. In mammals, all antennae in multi-antennary constructs are known to progress to full extension and capping(Schjoldager *et al*. 2020), with galactosylation by β4-galactosylransferase (b4Gal-T1)(Stanley *et al*. 2017; McDonald *et al*. 2014), sialylation by different substrate and linkage-specific sialyltransferases, such as ST3/6GalTs and ST3/6GalNAcTs(Harduin-Lepers *et al*. 2001; Bhide & Colley 2017), and fucosylation of the core in α6 by FUT8(Järvå *et al*. 2020; García-García *et al*. 2020) and in α3 of the GlcNAc on the antennae in a Lewis-X type termination(Mondal *et al*. 2018).

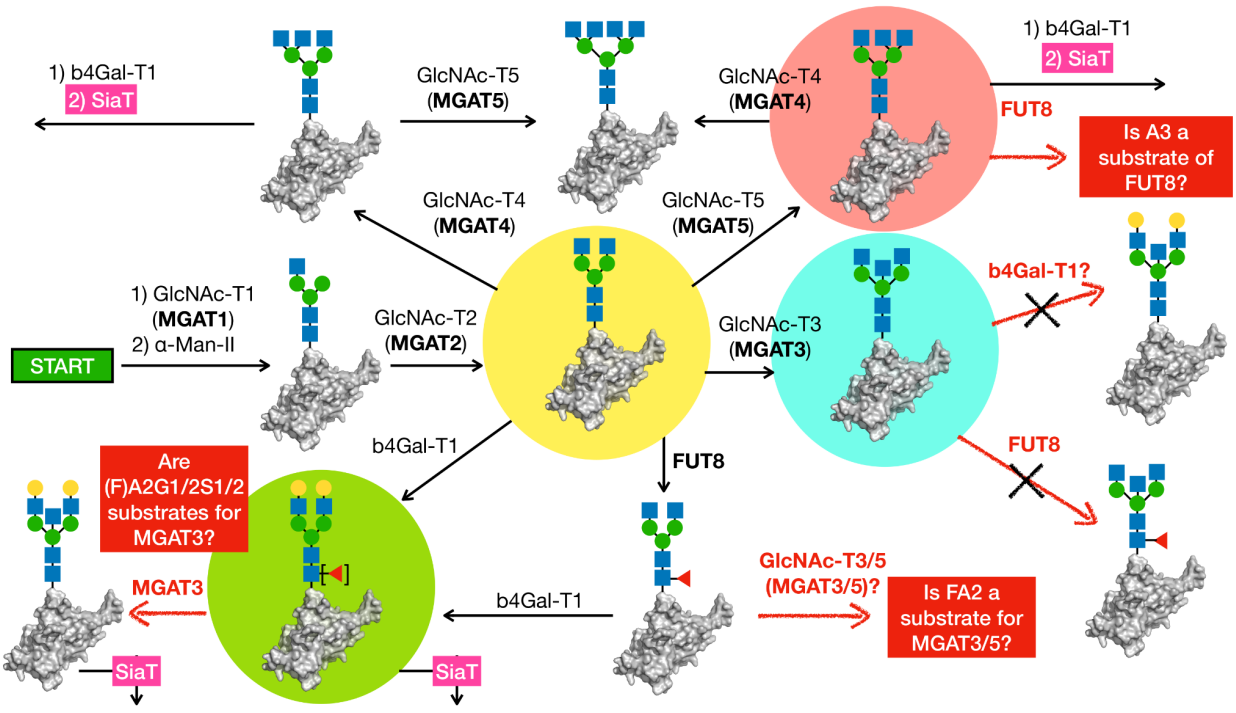


Figure 3.1. Schematic representation of the possible pathways leading to the maturation of a complex *N*-glycan. The biantennary A2 *N*-glycan is highlighted in yellow as a central node in the maturation. The bisected A2B system obtained as a product of MGAT3 on A2 is highlighted in cyan and tested in this work as substrates for galactosylation via b4GalT1 and for core fucosylation via FUT8. The triantennary A3 system with two antennae on the α6 arm is highlighted in pink and tested for core fucosylation by FUT8. Galactosylated (F)A2G1/2 *N*-glycans as potential substrates of MGAT3 are highlighted in green. Questions that remain to be addressed are shown in red boxes, and the corresponding pathways we are attempting to clarify are indicated by red arrows. Known pathways are shown with black arrows. Information used to build this scheme is based on the following references(Stanley *et al*. 2017; Schjoldager *et al*. 2020; Nakano *et al*. 2019).

The complexity and precision of the *N*-glycosylation machinery leads to an enormous degree of possibilities in the elaboration of *N*-glycans structures, which specific recognition contributes to a highly sophisticated cell-communication strategy known as the "glycocode"(Pilobello & Mahal 2007; Rodríguez *et al*. 2018; Ambrosi *et al*. 2005). Within this context, *N*-acetyl-glucosamine transferase 3 (GlcNAc-T3 or MGAT3) is responsible for a very important type of modification, namely the transfer of a GlcNAc in β4 onto the central Man of the *N*-glycan pentasaccharide core, also known as bisecting GlcNAc, see **Figure 3.1**. This type of functionalisation is not common, yet is highly expressed in specific tissues(Kizuka & Taniguchi 2018), and it has been shown to be linked to several disease states(Nakano *et al*. 2019; Chen *et al*. 2020), e.g. to cancer progression(Song *et al*. 2010; Yoshimura *et al*. 1995;

Miwa *et al*. 2013; Song *et al*. 2001; Kohler *et al*. 2016). The tissue specialisation and critical role in disease suggests that *N*-glycan bisection may act like a 'stop codon' within the glycocode, silencing messaging normally delivered by multiantennary *N*-glycan structures *via* recognition of their functionalised arms. Indeed, bisection was found to inhibit the maturation of tri- and tetra-antennary systems(Nakano *et al*. 2019; Schachter 1986).

Earlier work suggested that MGAT3 operates after GlcNAcylation of the α3 arm by GlcNAc-T1, but before GlcNAc-TII/IV and V and thus using primarily hybrid and biantennary *N*-glycans as substrates(Priatel *et al*. 1997; Schachter 1986), see **Figure 3.1**. Moreover, while bisected *N*-glycans have been shown to carry highly functionalised α3 and α6 arms in IgGs and tumour cells(Bondt *et al*. 2014; Link-Lenczowski *et al*. 2018), the occurrence of bisection has also been linked to inhibition of arms maturation and growth(Nakano *et al*. 2019), a conclusion that is also supported by a reduced cell surface binding by galectins(Miwa *et al*. 2013). This uncertainty about the consequences of *N*-glycans bisection in the structure and maturation of *N*-glycans may be indicative of the difficulties inherent to the experimental characterisation of these structures, where bisected and triantennary *N*-glycans have the same mass, and specific structures can be identified by lectin blotting, or by GC-MS or by multi-stage MS(Chen *et al*. 2020) and of the lack of knowledge around the substrate preference and mechanism of action of MGAT3.

In this work, I used extensive sampling through deterministic MD simulations run based on independent replicas to provide structural insight on the effect of bisection on the architecture of *N*-glycans and how does bisection affects recognition by human b4GalT1 and FUT8. In particular, I highlighted structural differences relative to fully functionalised triantennary *N*-glycans carrying two antennae on the α6 arm, and to biantennary *N*-glycans described in an earlier study(Harbison *et al*. 2019). This information lead me to understand if and how bisection can hinder elongation of the arms and at what specific steps in the maturation pathway. Furthermore, in collaboration with Prof Daniel Kolarich and Dr Andrea Maggioni in Prof Mark von Itzstein's group at the Institute for Biomedicine and Glycomics at Griffith University in Gold Coast, QLD, Australia, we are also seeking to characterise experimentally the affinity of a soluble construct of MGAT3 for mature complex *N*-glycans, as a strategy to understand the pathway leading to bisection in fucosylated, galactosylated and sialylated N-glycans often found in IgGs and also more elaborated, yet likely to be less common, multi-antenna structures(Klisch *et al*. 2008).

3.2 Methods

3.2.1 Computational Methods

The starting structures of the bisected and triantennary *N*-glycans were generated using the carbohydrate builder tool on GLYCAM-WEB (https://glycam.org/). For each *N*-glycan, I generated nine structures using different combinations of the α(1–6) torsions values. The topology files for each structure were obtained from the *tleap* program included in the AMBER 18(Lee *et al*. 2018) software package. The carbohydrate atoms were represented using the GLYCAM06j-1version of the GLYCAM06 force field(Kirschner *et al*. 2008) while the counterions in the simulations were represented using the AMBER ff14SB(Maier *et al*. 2015). Water molecules were represented using the TIP3P water model (Jorgensen *et al*. 1983). The MD simulations for the *N*-glycan structures were run in 200 mM NaCl concentration. All MD simulations were run following a similar protocol, starting with an energy minimization stage through 500,000 steps of steepest descent. During minimization, only hydrogen atoms, water molecules, and counterions were left unrestrained, while the position of all other atoms were kept restrained with a potential weight of 5 kcal/mol Å^2 with the exception of water molecules and counterions. Following minimisation, the system was brought up to standard temperature through a heating phase in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During the first stage, the temperature was raised from 0 to 100 K where the volume was kept constant and from 100 to 300 K during the second stage where the pressure was kept constant. The system was then equilibrated for 1 ns in the NPT ensemble to a 1 atm pressure with the Berendsen barostat. The same restraints used for the minimisation stage were kept for these equilibrations. Conformational equilibration was carried out for 10 ns without restraints. Production phase followed and analysis was carried out on the 500 ns trajectories for each starting structure. The torsion angles and RMSD values were calculated using VMD(Humphrey *et al*. 1996). A density-based clustering method, DBSCAN(Ester *et al*. 1996), was used to calculate the populated conformations occupied by each *N*-glycan based on the torsion angles calculated from all trajectories for that *N*-glycan.

The highest populated structures of A2B and A3 were structurally aligned to the b4GalT1 structure co-crystallised with the tetrasaccharide from lacto-N-neohexose, PDB 4EE4 (Ramakrishnan *et al*. 2012). To generate the fit, the terminal β2-GlcNAc of arms A2B and A3 were superimposed onto the terminal β6-GlcNAc of the co-crystallised Lacto-N-neohexose.

Another colleague from my research group, Silvia D'Andrea, assessed the potential recognition of the bisected and triantennary N-glycans by the FUT8 using as a template the protein co-crystallised with a biantennary structure PDB 6VLD (Järvå *et al*., 2020). In this case the A2B and A3 glycans structures were superimposed onto the biantennary structure, using the shared chitobiose core for alignment. All structural alignments were done with PyMOL(www.pymol.org). All non-clashing structures of the complexes obtained by alignment of A2B and A3 with b4GalT1 and FUT8 were tested for stability using MD where the simulations were run using the same protocol described above for the free glycans. We used the AMBER ff14SB force field (Maier *et al*., 2015) to represent protein atoms. While I ran the simulations of the b4GalT1 system, Silvia D'Andrea handled the FUT8 simulations. Production runs for each N-glycan-protein complex were run up to 1 μs.

3.3 Results

3.3.1 Biantennary *N*-glycans

In earlier work(Harbison *et al*. 2019) my research group showed through extensive sampling by MD simulations that the 3D structural propensity of biantennary *N*-glycans is clearly sequence-dependent. The fucosylated biantennary structure , (F)A2, with both arms terminating with β2-GlcNAc, which I will consider here as the principal node for complex *N*-glycan maturation, is relatively rigid with the only significant degree of freedom associated with the orientation of the α6 arm. This can adopt two conformations that we defined as 'open' (φ=70°, ψ=180°) and 'closed' (φ=70°, ψ=80°), with ω angle values predominantly around 50°(Harbison *et al*. 2019). In the open conformation the α6 arm is extended and accessible to transferases for further functionalisation, meanwhile in the closed conformation the α6 arm is found stacking the chitobiose core and interacting with the core fucose where present(Harbison *et al*. 2019). In (F)A2 structures the α6 arm open conformation is only slightly more populated than the closed conformation, with 55% vs. 45% relative populations over the sampling time. The addition of galactose on the α6 arm producing (F)A2G or (F)A2G2 structure, shifts this open/close equilibrium, where the closed conformation becomes dominant (75% population). This equilibrium is slightly shifted to favour even more the closed conformer by terminal sialylation(Harbison *et al*. 2019).

3.3.2 Bisected and triantennary *N*-glycans

The trajectories of the bisected (A2B) and triantennary (A3) *N*-glycans from MD simulations were analysed using VMD (Humphrey *et al*., 1996), and the torsion angles associated with their linkages were clustered using DBSCAN (Ester *et al*., 1996). The populations of the different torsion angles for the α6 arm in the A2B system, as well as the torsional values for both the α6 arm and the β6 torsion angles of the third branch in the A3 system, are reported in **Table 3.1**.

Table 3.1. Conformations of the α6 arm and relative populations obtained from clustering analysis of the MD results in function of the *N*-glycan architecture, namely bisected (A2B) or triantennary (A3), and sequence. The values of the β6 torsion angle in the A3 system, regulating the orientation of the third branch, are also shown. For each cluster, the values (in degrees) of the torsion angles φ ($O_1C_1O_xC_x$)(*sd*)/ψ ($C_1O_xC_xC_{x+1}$) (*sd*)/ω ($O_6C_6C_5C_4$))(*sd*) correspond to the median calculated for each cluster with standard deviations shown in parenthesis and relative populations (%) in red. The α6 arm sequences are represented according to the SNFG convention(Neelamegham *et al*. 2019).

A2B	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
GlcNAc-β2-Man-	71.8 (*9.4*) / 167.9 (*9.3*) / 54.5 (*9.4*) / **72**	74.2 (*13.1*) / 82.6 (*15.2*) / 44 (*6.8*) / **28**		
Gal-β4-GlcNAc-β2-Man-	80.3 (*14.7*) / 89.2 (*14.2*) / 46 (*7.1*) / **50**	71.9 (*8.6*) / 167.1 (*9.2*) / 54.6 (*8.9*) / **41**	73.7 (*6.2*) / 65.9 (*7.0*) / 166.2 (*5.6*) / **9**	
Sia-α3-Gal-β4-GlcNAc-β2-Man-	77.6 (*15.1*) / 85.6 (*15.0*) / 45.2 (*7.6*) / **70**	71.6 (*8.3*) / 167.0 (*8.7*) / 54.4 (*8.6*) / **30**		
A3	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
GlcNAc-β2-[GlcNAc-β6]-Man-**α6**-Man	70.0 (*9.0*) / -172.0 (*12.9*) / 55.2 (*9.5*) / **49**	72.3 (*11.0*) / 89.3 (*13.9*) / 48.5 (*9.1*) / **44**	77.7 (*4.9*) / -73.9 (*6.0*) / -139.0 (*6.1*) / **4**	62.4 (*5.2*) / -177.0 (*5.9*) / 179.0 (*5.1*) / **3**
Gal-β4-GlcNAc-β2[Gal-β4-GlcNAc-β6]-Man-**α6**-Man	71.4 (*9.5*) / -172.0 (*14*) / 55.6 (*9.2*) / **46**	75.3 (*12.4*) / 92.9 (*10*) / 51.6 (*7.9*) / **37**	68.4(10.3)/ -173(12.9)/ -173(14.2)/ **10**	60.8(9.5)/ 54.3(11.3)/ 171(7.8)/ **6**

Sia-α6-Gal-β4-GlcNAc-β2-[Sia-α6-Gal-β4-GlcNAc-β6]-Man-**α6**-Man	73.3 (11.9) / 89.3 (16.9) / 48.3 (11.0) / **47**	68.3 (9.51) / -175 (13.8) / 53.3 (10.2) / **36**	68.3 (11.6) / 61.6 (11.5) / 170 (8.28) / **12**	90 (6.26) / -48.1 (7.68) / -145 (6.93) / **3**
Sia-α3-Gal-β4-GlcNAc-β2-[Sia-α3-Gal-β4-GlcNAc-β6]-Man-**α6**-Man	71.9 (12.1) / 87.2 (16.6) / 48.3 (9.5) / **46**	69.4 (8.2) / -173 (12.2) / 54.5 (8.6) / **23**	67.8 (6.32) / -166 (10.7) / -65.2 (8.7) / **15**	78 (12.1) / -67.9 (8.5) / -143.0 (8.4) / **15**
GlcNAc-**β6**-Man-	-78.7(12.6)/ -178.8(18.7)/ 52.7(9.7)/ **52**	-75.0(10.3)/ -170.6(15.9)/ -168.0(11.0)/ **42**	-81.0(7.9)/ -89.0(7.9)/ -160.5(9.3)/ **6**	
Gal-β4-GlcNAc-**β6**-Man-	-81.2(11.6)/ 171.4(14.4)/ 53.1(8.9)/ **56**	-74.68(9.35)/ -173.15(15.08)/ -168.72(9.82)/ **32**	-72.23(8.21)/ -85.42(9.77)/ -157.94(9.14)/ **12**	
Sia-α6-Gal-β4-GlcNAc-**β6**-Man-	-71.0(9.4)/ -163.8(10.8)/ 53.1(9.9)/ **68**	-66.9(6.5)/ -145.5(9.6)/ -161.3(7.3)/ **17**	-68.8(6.5)/ 175.3(8.6)/ -169.2(6.9)/ **15**	
Sia-α3-Gal-β4-GlcNAc-**β6**-Man-	-74.0(10.2)/ -174.7(17.0)/ -166.8(12.4)/ **49**	-70(8.4)/ -87.6(10.6)/ -161.9(10.5)/ **24**	-76.3(9.3)/ -179.2(18.6)/ 54.4(8.8)/ **19**	-87.9(7.1)/ -84.0(7.26)/ -88.1(8.0)/ **5**

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

The presence of a bisecting GlcNAc does not affect the structure of the *N*-glycan core and the conformation of the α3 arm, relative to the equilibrium values determined for the biantennary systems(Harbison *et al*. 2019). Yet, as shown in **Table 3.1**, the conformational equilibrium of the α6 arm is highly affected by bisection. Indeed, in a A2B structure with both terminated by a β2-GlcNAc, the most populated conformations are the open (φ = 71.8°, ψ = 167.9°, ω = 54.5°) and closed (φ = 74.2°, ψ = 82.6°, ω = 44.0°) states, as shown in **Figure 3.2**. However, relative to the (F)A2 structure, the addition of the bisecting GlcNAc shifts the equilibrium toward the open conformation which during the MD simulations represents 72% of the conformational population. Adding a terminal Gal to both arms rebalances the equilibrium

between open and closed conformations, bringing them closer to an even distribution, ie. 41% open and 50% closed Terminal sialylation shifts this equilibrium further favouring the closed conformation, reflected by the relative population reaching 70%.

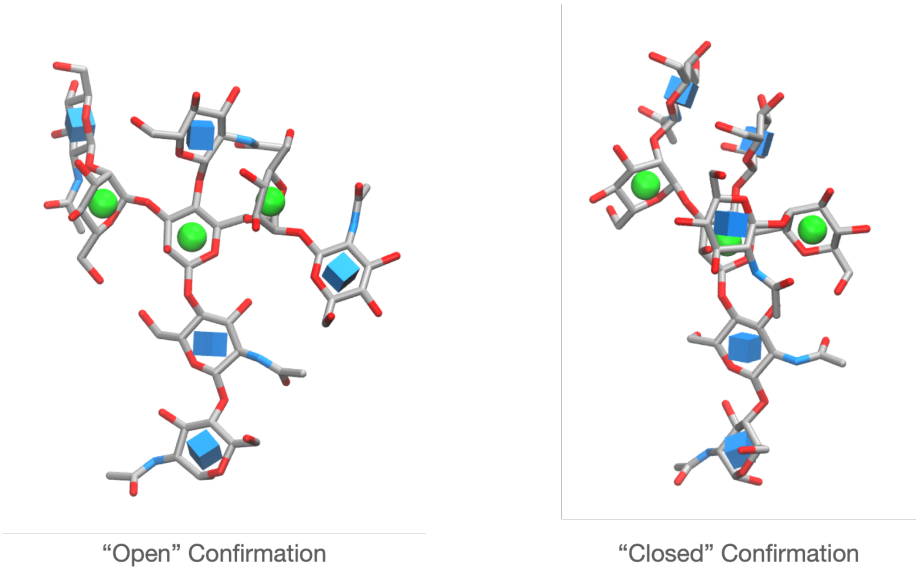


Figure 3.2. Representative structures of the "open" (left) and "closed" (right) conformations of the α6 arm in the bisected A2B *N*-glycan. Graphical rendering with VMD(Humphrey *et al*. 1996)(https://www.ks.uiuc.edu/Research/vmd/).

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

As described in **Chapter 3**, the triantennary structure exhibits greater flexibility around its α(1-6) arm compared to biantennary and bisected structures, allowing it to sample a broader range of conformations. For triantennary structures with each arm terminating in β(1-2)-GlcNAc, the most populated conformations are the open conformation (φ = 70.0°, ψ = -172.0°, ω = 55.2°), which occurs 49% of the time, followed by the front fold (φ = 72.3°, ψ = 89.3°, ω = 48.5°) at 44%. Less frequently observed are the back fold (φ = 77.7°, ψ = -73.9°, ω = -139.0°) at 4% and a variant of the open conformation (φ = 64.5°, ψ = -177.0°, ω = 179.0°) at 3%, as shown in **Table 3.1**.

The addition of terminal Gal to the α(1-6) arm has little effect on the conformational distribution, with the open conformation (46%) still dominant, followed by the front-fold (37%). The back-fold (10%) and a variant of the open conformation (6%) are less populated. In contrast, terminal sialylation shifts the equilibrium toward the front-fold conformation, which becomes the most populated,47% for α(2-6)-linked and 46% for α(2-3)-linked sialylation, and the open conformation decreases to 36% and 23%, respectively, while the other conformations emerge at lower populations.

The β(1-6)-linked GlcNAc extending from the α(1-6) arm also exhibits a range of distinct conformations, also described in **Chapter 3**. These conformations include one in which the β(1-6)-linked GlcNAc adopts a 'claw-like' configuration, facing the β(1-2)-linked GlcNAc on the same arm. For triantennary structures with each arm terminating in β(1-2)-GlcNAc this conformation is the most populated, with an occupancy of 52% (φ = -78.7°, ψ = -178.8°, ω = 52.7°). A second conformation, observed 42% of the time (φ = -75.0°, ψ = -170.6°, ω = -168.0°), positions the β(1-6)-linked GlcNAc oriented toward the α(1-3) arm, while a third, lower-occupancy conformation at 6% (φ = -81.0°, ψ = -89.0°, ω = -160.5°) aligns the β(1-6) arm behind the α(1-2)-linked GlcNAc on the same arm.

The addition of β(1-4)-linked galactose does not significantly alter this distribution, with the 'claw-like' conformation remaining dominant at 56% (φ = -81.2°, ψ = 171.4°, ω = 53.1°), followed by the α(1-3)-oriented conformation at 32% (φ = -74.7°, ψ = -173.2°, ω = -168.7°) while the third conformation, in which the β(1-6)-linked GlcNAc aligns behind the α(1-2)-linked GlcNAc, is observed at 12% (φ = -72.2°, ψ = -85.4°, ω = -157.9°).

Terminal α(2-6) sialylation only slightly increases the prevalence of the 'claw-like' conformation to 68% (φ = -71.0°, ψ = -163.8°, ω = 53.1°) whereas the terminal α(2-3) sialylation on the other hand shifts this conformational equilibrium. The most populated conformation now becomes the one where the β(1-6)-linked GlcNAc is oriented toward the α(1-3) arm, increasing to 49% (φ = -74.0°, ψ = -174.7°, ω = -166.8°). Meanwhile, the conformation in which the β(1-6)-linked GlcNAc aligns behind the α(1-3)-linked GlcNAc also rises in prevalence to 24% (φ = -70.0°, ψ = -87.6°, ω = -161.9°). The prevalence of the 'claw-like' state decreases to 19% (φ = -76.3°, ψ = -179.2°, ω = 54.4°).

3.3.5 Assessing the suitability of A2B and A3 as substrates for b4GalT1

The β4-galactosyltransferase 1 (b4GalT1, EC 2.4.1.38) is a transmembrane protein located in the Golgi responsible for the transfer of galactose from a UDP-α-D-galactose donor to a terminal *N*-acetyl-β-D-glucosamine. I selected the b4GalT1 structure from PDB 4EE4(Ramakrishnan *et al*. 2012) to build a recognition complex using the highest populated conformers of A2B and A3 we identified through MD. The b4GalT1 structure from 4EE4 was resolved at 1.95 Å and was crystallised in complex with a tetrasaccharide from lacto-N-neohexose(GlcNAcβ1–3 (GlcNAcβ1-6)Galβ1–4Glc). I aligned the highest populated

conformers I obtained for A2B and A3 by pair-fitting the terminal β2-GlcNAc on their arms onto the terminal β6-GlcNAc of the lacto-N-neohexose. Each alignment using the terminal β2-GlcNAc on different arms would lead to clashes between the residue of the b4GalT1 and different parts of the bisected structure, see **Figure 3.3.**

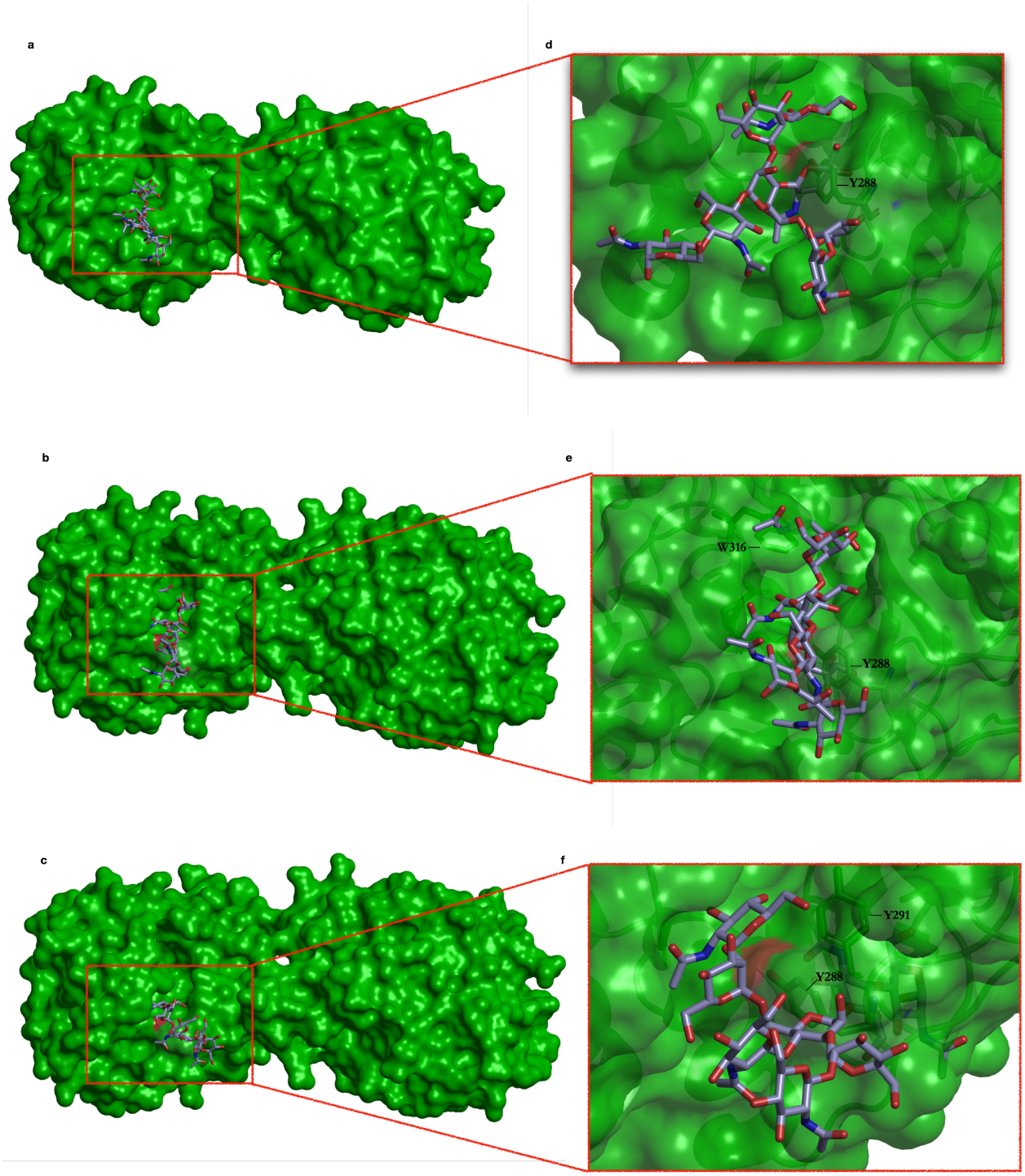


Figure 3.3. **Panel a-c)**b4GalT1 in complex with A2B generated from structural alignment with the β2-GlcNAc on different arms. b4GalT1 rendered as surface (green) and A2B rendered as sticks (violet). **Panel d-f)** Close-up of b4GalT1/A2B complex highlighting steric clash of the bisecting GlcNAc with the different residues. Labels provided for residues Y288, Y291 and W316. Renderings generated using PyMOL(www.pymol.org)

In contrast, the structural alignment of the terminal β2-GlcNAc on the third antenna the 'open' A3 conformer with the co-crystallised ligand produced a good fit wherein the arms were able to exist in complex with the b4GalT1 structure without any steric clashes. From this alignment, I generated structures of the b4GalT1-A3 recognition complex and assessed its stability by MD simulations. Results from the 1μs of MD simulation are shown in **Figure 3.4.** The A3 structure remained bound for the duration of the simulation with an average RMSD of 4.56 Å and a standard deviation of 0.48 Å. The A3 structure was able to interact with residues lining the binding site, namely Q173, D203, D204 and R244, see **Figure 3.4b** and **Figure 3.4d**

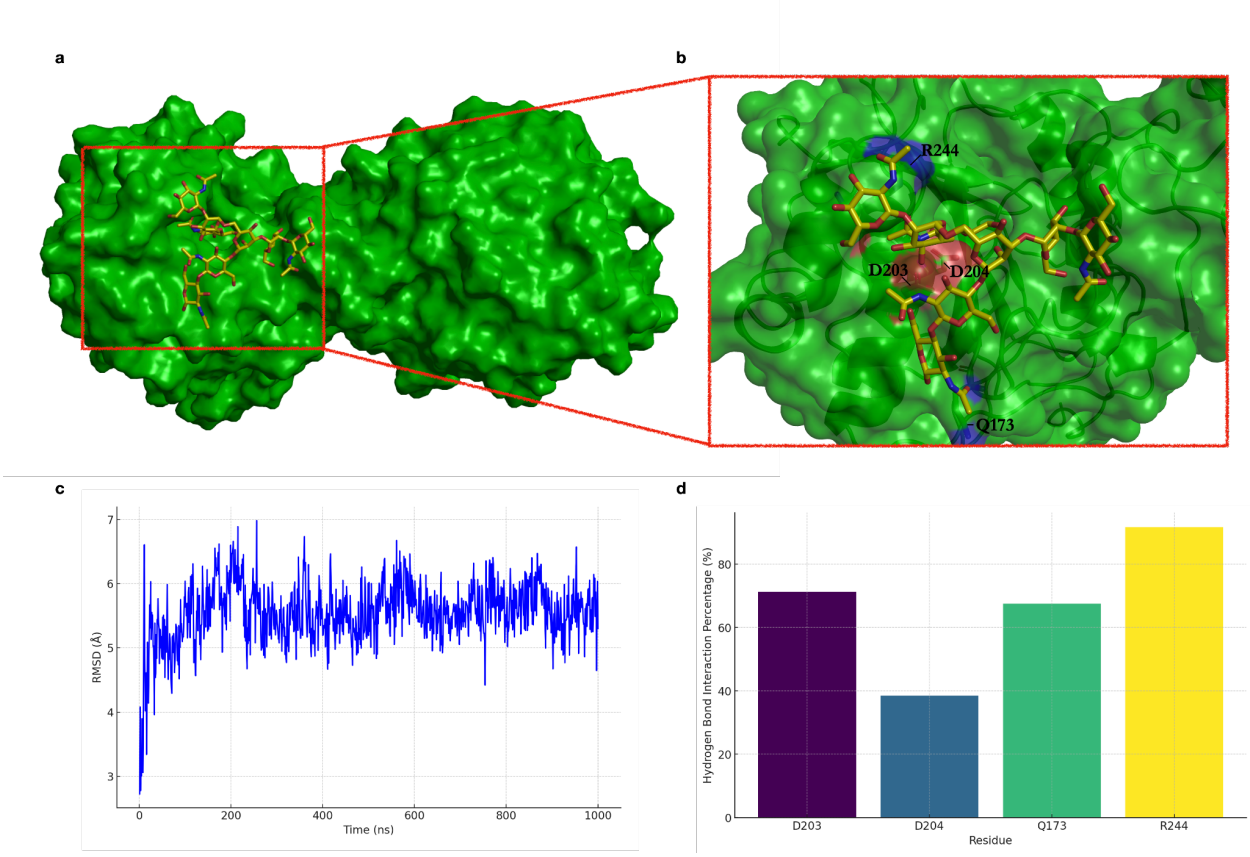


Figure 3.4. Panel a) Representative snapshot from the MD trajectory of the recognition complex between A3 and b4GalT1. b4GalT1 rendered as surface (green) and A3 rendered as sticks (purple). **Panel b)** Atomistic view of the bound A3 on the protein surface. Key binding residues, namely Q173, D203, D204 and R244 are labelled. **Panel c)** Root Mean Square Deviation (RMSD) values (Å) of the bound A3 structure plotted against simulation time(ns). Graphical representation with PyMOL (www.pymol.org). **Panel d)** Barplot indicating the persistence of the hydrogen bonding interactions between the A3 structure and specific binding site residues calculated during the MD trajectory.

3.3.6 Assessing recognition of A2B and A3 by FUT8

The α-1,6-fucosyltransferase (FUT8, EC 2.4.1.68) is an enzyme located in the Golgi that catalyses the transfer of fucose from a GDP-beta-L-fucose to the innermost N-acetylglucosamine of complex and hybrid N-glycan in a process known as core-fucosylation. FUT8 functions as a dimer, stabilised by a four-helix bundle formed by its N-terminal coiled-coil domains. This dimerisation is crucial for maintaining the structural integrity of the SH3 domain, which is responsible for recognising the acceptor substrate and enabling precise substrate binding (Järvå et al. 2020). My colleague, Silvia D'Andrea, followed a similar protocol using PDB 6VLD which contains the resolved dimer structure at 2.28 Å of FUT8 co-crystallised with GDP and an asialo-agalacto-biantennary glycopeptide (A2SGP) acceptor substrate to build structures of A2B and A3 in complex with FUT8. The co-crystallised A2SGP as a template for the structural alignments of the A2B and A3 potential substrates using the shared chitobiose core to superimpose the N-glycan structures. The results of her analysis are summarised below.

Analogously to the case discussed above for b4GalT1, the structural alignment of A2B structures indicated that bisection precludes recognition by FUT8, see **Figure 3.5**.

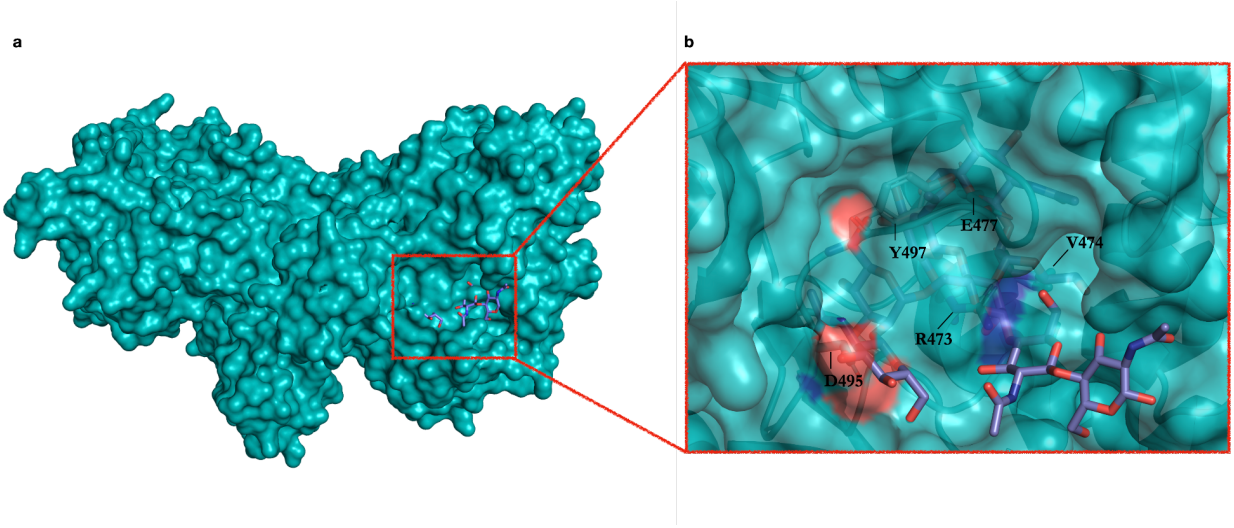


Figure 3.5. **Panel a)** Structural alignment of A2B with FUT8. FUT8 is rendered as surface (cyan) and A2B is rendered with sticks (C atoms in violet, O atoms in red and N atoms in blue). **Panel b)** Atomistic view of FUT8/A2B complex highlighting the incompatibility of the A2B structure with the FUT8 proteins with steric clashes of the A2B with residues R473, V474, E477, D495 and Y497(labelled). Renderings generated using PyMOL(www.pymol.org).

Meanwhile, the structural alignment of the open A3 structure with the co-crystallised A2SGP generated a viable fit. The structures of the complex obtained from this alignment were tested for stability by MD simulations. Results from the 1μs of MD simulation of the A3 structure in

complex with the FUT8 structure are shown in **Figure 3.6.** The A3 structure remained stable within the complex throughout the simulation, with an average RMSD of 4.57 Å and a standard deviation of 0.75 Å.

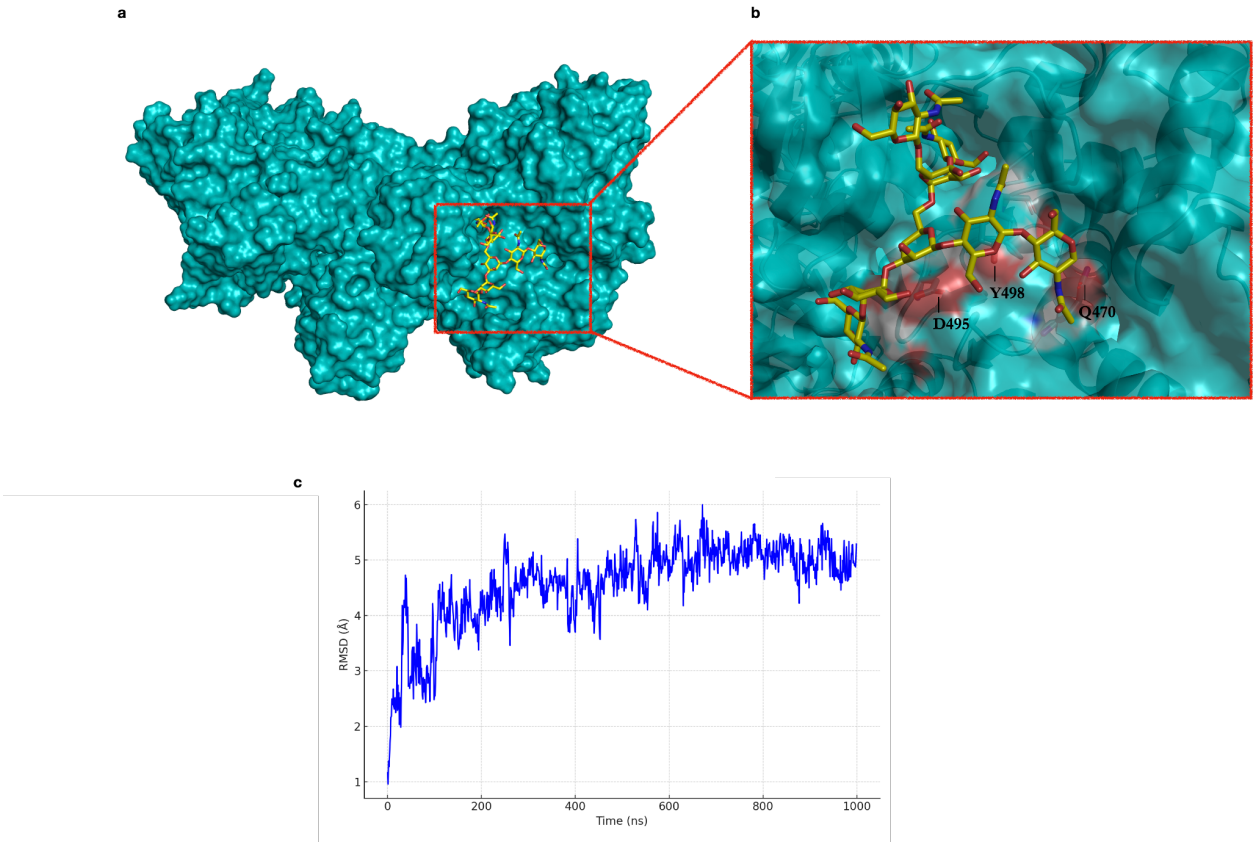


Figure 3.6. Panel a) Graphical representation of representative trajectory of A3 in complex with FUT8 taken from MD simulation. FUT8 rendered as surface (cyan) and A3 rendered as sticks (yellow). **Panel b)** Atomistic view of the bound A3 on the protein surface. Key binding residues, namely Q470, D495 and D Y498 are labelled. **Panel c)** RMSD values values (Å) of the bound A3 structure plotted against simulation time(ns).

3.4 Discussion

The results of the structural alignments and MD simulations gives us some important insight into how different N-glycan architectures, i.e. biantennary, triantennary and bisected structures, modulate recognition and thus maturation. Previous work on biantennary structures (Harbison *et al.*, 2019) showed that the conformation of the α6 arm shifts progressively toward a closed state upon addition of terminal galactose and terminal sialic acid. I have shown here that *N*-glycan bisection stabilises the open conformation of A2B structures. This shift may be due to steric hindrance introduced by the bisecting GlcNAc, which restricts the α6 arm rotational

degrees of freedom, making it more difficult for it to transition to a closed conformation. The addition of the terminal Gal and Neu5Ac pushes the equilibrium back in favour of the closed state, showing how the elongation of the arms helps to maintain the closed conformation. I found that the addition of a third antenna on the α6 arm, leading to the triantennary A3 structure, introduces greater conformational complexity to the *N*-glycan architecture. The α6 arm in A3 can occupy a broader range of states, including the 'front-fold' and 'back-fold' conformations previously observed in oligomannose structures (Fogarty & Fadda, 2021). The flexible β6 linkage of the third antenna also adopts a wide array of torsion angles, adding to the conformational diversity of the structure. This diversity can be further varied by the elongation of the arms.

The in-depth conformational analysis of the bisected and triantennary structures unbound in solution, led me to investigate if these could be recognised and bound by key enzymes responsible for the functionalisation of the core (FUT8) and of the arms (b4GalT1). Based on my analysis and that of my colleague, Silvia D'Andrea, the A2B structure could not be recognised by either FUT8 or B4GalT1, where the structural alignment of the highest populated conformers from MD to co-crystallised ligands lead to unsolvable steric clashes with the proteins. These findings align with previous research, which reported that A2B is incompatible with FUT8 (Järvå *et al*., 2020) and inhibits FUT8 activity, preventing core fucosylation (Tseng *et al*. 2017; Ferrara *et al*. 2006; Schuster *et al*. 2005). Similarly, it has been shown that bisection of the biantennary *N*-glycan generally suppresses arm elongation (Nakano *et al*., 2019), consistent with our results, thereby preventing the further addition of Gal.

While our results demonstrate why bisection hinders recognition of A2B structures by FUT8 and B4GalT1, effectively limiting *N*-glycan elongation, we still observe bisected biantennary structures with core fucosylation and highly functionalised arms in IgGs and in tumor cells (Bondt *et al*., 2014; Link-Lenczowski *et al*., 2018). The presence of the bisecting GlcNAc has also been reported as part of core-fucosylated tetraantennary structures carrying terminal SDa(Neu5Acα2-3[GalNAcβ1-4]Galβ1-4GlcNAc) epitopes, characterised as part of pregnancy-associated glycoproteins in bovines(Klisch *et al*. 2008). This suggests an alternative pathway leading to mature bisected *N*-glycan structures. One such alternative could entail functionalisation of already mature forms by MGAT3, see **Figure 3.1.** There are currently no available PDB structures of MGAT3 and the AF structure (UniProt Q09327) does not provide sufficient insight for us to understand recognition through molecular docking analysis. To

address this shortcoming, we set up a collaboration with Dr Maggioni and Prof Kolarich at the Institute for Biomedicine and Glycomics at Griffith University in QLD, Australia, who (to date) have successfully expressed a soluble construct of human MGAT3 and are currently investigating its activity through glycomic profiling by mass spectrometry.

Our MD-led structural alignment approach indicated that triantennary A3 structures can be recognised by both FUT8 and b4GalT1, suggesting that they can function as substrates. Previous studies have shown that a knockout of the activity of b4GalT1 protein leads to a decrease in galactosylation in triantennary *N*-glycans(Bydlinski *et al*. 2018; Zhao *et al*. 2023) supporting the possibility that triantennary structures could serve as substrates for b4GalT1. In the case of FUT8, earlier studies indicate a higher activity toward triantennary structures synthesised by MGAT4, where a β4-linked GlcNAc is added to the α3 mannose arm. However, triantennary structures synthesised by MGAT5—with a β6-linked GlcNAc on the α6 mannose arm i.e the triantennary structure we tested—showed lower activity toward FUT8 than both biantennary and MGAT4-synthesised triantennary structures(Tseng *et al*. 2017; Tomida *et al*. 2024). The structural differences between these triantennary forms are illustrated in **Figure 3.7**. This suggests that the triantennary structure synthesised by MGAT5 is not the preferred substrate for FUT8.

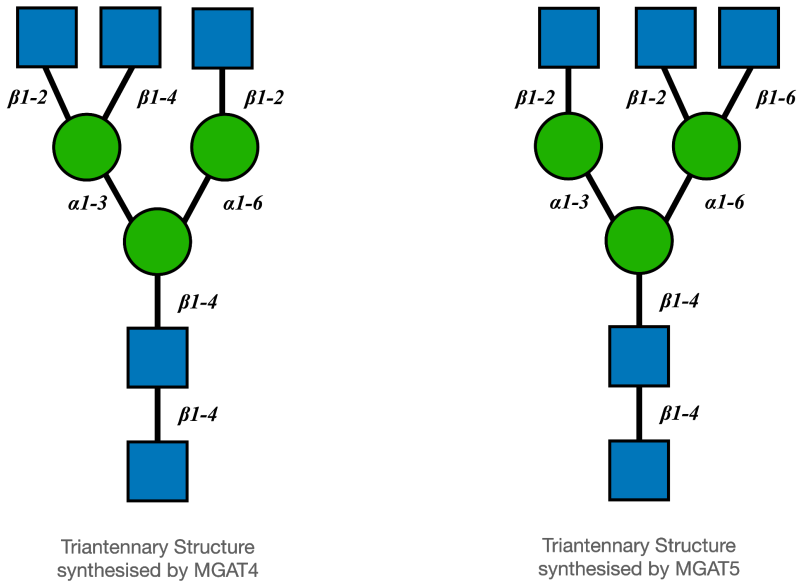


Figure 3.7: Triantantennary *N*-glycan structures synthesised by MGAT4 and MGAT5 illustrated using the Symbol Nomenclature for Glycans (SNFG)(Neelamegham *et al*. 2019).

3.5 Conclusion

In this work, I explored the architecture of the bisected and triantennary *N*-glycan structures in comparison to the biantennary structure using MD simulations. My results show how the presence of a bisecting GlcNAc leads to a shift in the conformational equilibrium and how the addition of a third antenna leads to greater conformational diversity. I further investigated how these modifications may impact *N*-glycan maturation and substrate compatibility using the equilibrium structures we identified for A2B and A3 in solution as potential recognition substrates for FUT8 and b4GalT1. My approach based on structural alignment of the equilibrium N-glycans structures in combination with additional MD simulations of the viable complexes, indicates that bisection precludes substrate recognition by both FUT8, and b4GalT1, while the A3 structure is a suitable substrate for both. My findings provide a structural rationale that complements experimental earlier findings, highlighting how *N*-glycan bisection can act as a suppressor of *N*-glycan maturation.

References:

1. Aebi, M., 2013. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(11), pp.2430–2437.
2. Ambrosi, M., Cameron, N.R. & Davis, B.G., 2005. Lectins: tools for the molecular understanding of the glycocode. *Organic & biomolecular chemistry*, 3(9), pp.1593–1608.
3. Bhide, G.P. & Colley, K.J., 2017. Sialylation of N-glycans: mechanism, cellular compartmentalization and function. *Histochemistry and cell biology*, 147(2), pp.149–174.
4. Bondt, A. *et al*., 2014. Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Molecular & cellular proteomics: MCP*, 13(11), pp.3029–3039.
5. Bydlinski, N. *et al*., 2018. The contributions of individual galactosyltransferases to protein specific N-glycan processing in Chinese Hamster Ovary cells. *Journal of biotechnology*, 282, pp.101–110.
6. Chen, Q. *et al*., 2020. The Essential Functions and Detection of Bisecting GlcNAc in Cell Biology. *Frontiers in chemistry*, 8, p.511.
7. Cobb, B.A., 2020. The history of IgG glycosylation and where we are now. *Glycobiology*, 30(4), pp.202–213.
8. Ester, M. *et al*., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Knowledge Discovery and Data Mining*, pp.226–231.
9. Ferrara, C. *et al*., 2006. Modulation of therapeutic antibody effector functions by glycosylation engineering: influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnology and bioengineering*, 93(5), pp.851–861.
10. Fogarty, C.A. & Fadda, E., 2021. Oligomannose N-Glycans 3D Architecture and Its Response to the FcγRIIIa Structural Landscape. *The journal of physical chemistry. B*,

125(10), pp.2607–2616.

11. García-García, A. *et al*., 2020. Structural basis for substrate specificity and catalysis of α1,6-fucosyltransferase. *Nature communications*, 11(1), pp.1–9.
12. Gleeson, P.A. & Schachter, H., 1983. Control of glycoprotein synthesis. *The Journal of biological chemistry*, 258(10), pp.6162–6173.
13. Harbison, A.M. *et al*., 2019. Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology*, 29(1), pp.94–103.
14. Harduin-Lepers, A. *et al*., 2001. The human sialyltransferase family. *Biochimie*, 83(8), pp.727–737.
15. Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), pp.33–8, 27–8.
16. Järvå, M.A. *et al*., 2020. Structural basis of substrate recognition and catalysis by fucosyltransferase 8. *The Journal of biological chemistry*, 295(19), pp.6677–6688.
17. Jorgensen, W.L. *et al*., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), pp.926–935.
18. Kirschner, K.N. *et al*., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry*, 29(4), pp.622–655.
19. Kizuka, Y. & Taniguchi, N., 2018. Neural functions of bisecting GlcNAc. *Glycoconjugate journal*, 35(4), pp.345–351.
20. Klisch, K. *et al*., 2008. A tetraantennary glycan with bisecting N-acetylglucosamine and the Sd(a) antigen is the predominant N-glycan on bovine pregnancy-associated glycoproteins. *Glycobiology*, 18(1), pp.42–52.
21. Kohler, R.S. *et al*., 2016. Epigenetic activation of MGAT3 and corresponding bisecting GlcNAc shortens the survival of cancer patients. *Oncotarget*, 7(32), pp.51674–51686.
22. Lee, L.Y. *et al*., 2014. Differential site accessibility mechanistically explains subcellular-

specific N-glycosylation determinants. *Frontiers in immunology*, 5, p.404.

23. Lee, T.-S. *et al*., 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. *Journal of chemical information and modeling*, 58(10), pp.2043–2050.
24. Link-Lenczowski, P. *et al*., 2018. The glycomic effect of N-acetylglucosaminyltransferase III overexpression in metastatic melanoma cells. GnT-III modifies highly branched N-glycans. *Glycoconjugate journal*, 35(2), pp.217–231.
25. Maier, J.A. *et al*., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation*, 11(8), pp.3696–3713.
26. McDonald, A.G. *et al*., 2014. Galactosyltransferase 4 is a major control point for glycan branching in N-linked glycosylation. *Journal of cell science*, 127(Pt 23), pp.5014–5026.
27. Miwa, H.E. *et al*., 2013. Bisected, complex N-glycans and galectins in mouse mammary tumor progression and human breast cancer. *Glycobiology*, 23(12), pp.1477–1490.
28. Mohorko, E., Glockshuber, R. & Aebi, M., 2011. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *Journal of inherited metabolic disease*, 34(4), pp.869–878.
29. Mondal, N. *et al*., 2018. Distinct human α(1,3)-fucosyltransferases drive Lewis-X/sialyl Lewis-X assembly in human cells. *The Journal of biological chemistry*, 293(19), pp.7300–7314.
30. Nakano, M. *et al*., 2019. Bisecting GlcNAc Is a General Suppressor of Terminal Modification of N-glycan*, [S]. *Molecular & cellular proteomics: MCP*, 18(10), pp.2044–2057.
31. Neelamegham, S. *et al*., 2019. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology*, 29(9), pp.620–624.

32. Paschinger, K. & Wilson, I.B.H., 2019. Comparisons of N-glycans across invertebrate phyla. *Parasitology*, 146(14), pp.1733–1742.

33. Pilobello, K.T. & Mahal, L.K., 2007. Deciphering the glycocode: the complexity and analytical challenge of glycomics. *Current opinion in chemical biology*, 11(3), pp.300–305.

34. Pinho, S.S. & Reis, C.A., 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nature reviews. Cancer*, 15(9), pp.540–555.

35. Priatel, J.J. *et al*., 1997. Isolation, characterization and inactivation of the mouse Mgat3 gene: the bisecting N-acetylglucosamine in asparagine-linked oligosaccharides appears dispensable for viability and reproduction. *Glycobiology*, 7(1), pp.45–56.

36. Ramakrishnan, B., Boeggeman, E. & Qasba, P.K., 2012. Binding of N-acetylglucosamine (GlcNAc) β1-6-branched oligosaccharide acceptors to β4-galactosyltransferase I reveals a new ligand binding mode. *The Journal of biological chemistry*, 287(34), pp.28666–28674.

37. Riley, N.M. *et al*., 2019. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nature communications*, 10(1), p.1311.

38. RodrÍguez, E., Schetters, S.T.T. & van Kooyk, Y., 2018. The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nature reviews. Immunology*, 18(3), pp.204–211.

39. Schachter, H., 1986. Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 64(3), pp.163–181.

40. Schjoldager, K.T. *et al*., 2020. Global view of human protein glycosylation pathways and functions. *Nature reviews. Molecular cell biology*, 21(12), pp.729–749.

41. Schuster, M. *et al*., 2005. Improved effector functions of a therapeutic monoclonal Lewis Y-specific antibody by glycoform engineering. *Cancer research*, 65(17), pp.7934–7941.

42. Song, E.Y. *et al*., 2001. Expression of bisecting N-acetylglucosaminyltransferase-III in

human hepatocarcinoma tissues, fetal liver tissues, and hepatoma cell lines of Hep3B and HepG2. *Cancer investigation*, 19(8), pp.799–807.

43. Song, Y. *et al*., 2010. The bisecting GlcNAc on N-glycans inhibits growth factor signaling and retards mammary tumor progression. *Cancer research*, 70(8), pp.3361–3371.
44. Speciale, I. *et al*., 2017. Structure of the N-glycans from the chlorovirus NE-JV-1. *Antonie van Leeuwenhoek*, 110(11), pp.1391–1399.
45. Stanley, P., Taniguchi, N. & Aebi, M., 2017. N-Glycans. In A. Varki *et al*., eds. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
46. Strasser, R., 2014. Biological significance of complex N-glycans in plants and their impact on plant physiology. *Frontiers in plant science*, 5, p.363.
47. Sun, S. & Zhang, H., 2015. Identification and Validation of Atypical N-Glycosylation Sites. *Analytical chemistry*, 87(24), pp.11948–11951.
48. Taniguchi, N. & Kizuka, Y., 2015. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Advances in cancer research*, 126, pp.11–51.
49. Thaysen-Andersen, M. & Packer, N.H., 2012. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology*, 22(11), pp.1440–1452.
50. Tomida, S., Nagae, M. & Kizuka, Y., 2024. Distinctive domains and activity regulation of core fucosylation enzyme FUT8. *Biochimica et biophysica acta. General subjects*, 1868(4), p.130561.
51. Tseng, T.-H. *et al*., 2017. Substrate Preference and Interplay of Fucosyltransferase 8 and N-Acetylglucosaminyltransferases. *Journal of the American Chemical Society*, 139(28), pp.9431–9434.
52. Varki, A., 2017. Biological roles of glycans. *Glycobiology*, 27(1), pp.3–49.

53. Wild, R. *et al*., 2018. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science*, 359(6375), pp.545–550.

54. Wildt, S. & Gerngross, T.U., 2005. The humanization of N-glycosylation pathways in yeast. *Nature reviews. Microbiology*, 3(2), pp.119–128.

55. Wilson, I.B.H., 2002. Glycosylation of proteins in plants and invertebrates. *Current opinion in structural biology*, 12(5), pp.569–577.

56. Yoshimura, M. *et al*., 1995. Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection. *Proceedings of the National Academy of Sciences*, 92(19), pp.8754–8758.

57. Zhao, Y. *et al*., 2023. In-depth mass spectrometry analysis reveals the plasma proteomic and N-glycoproteomic impact of an Amish-enriched cardioprotective variant in B4GALT1. *Molecular & cellular proteomics: MCP*, 22(8), p.100595.

Chapter 4. Structure, dynamics, and glycan-binding specificities of the *N. meningitidis* Type IV Pili

4.1 Introduction

The genus *Neisseria* includes 55 members, largely populated by innocuous commensal Gram-negative bacteria that colonise the mucosa and dental surfaces of humans(Bennett *et al*., 2012). Only two members of the *Neisseria* genus are pathogenic, namely *Neisseria meningitidis* (*Nm*) and *Neisseria gonorrhoeae* (*Ng*). *Nm* is known to cause meningococcal disease, the main cause of meningitis and septicemia. Humans are the only natural host of *Nm*, and an estimated 5-10% of the population carries it in the nasopharyngeal region without consequences to their health. In the rare cases where *Nm* enters the bloodstream, it can breach the blood-brain barrier(Coureuil *et al.*, 2009) causing severe morbidity and mortality. Although treatable with widely available antibiotics, such as penicillin and ampicillin, meningococcal disease has a very rapid onset and progression with fatalities still reaching an 8–15% ratio and the highest incidence in children and young adults (https://www.ecdc.europa.eu/en/meningococcal-disease/factsheet).

The majority of pathogenic *Nm* strains are encapsulated and have been categorised into serogroups based on the nature of their capsular polysaccharide. Serogroups A, B, C, W, X, and Y have been reported as the predominant causes of the disease globally with variations in prevalence based on geographical location(Jafri *et al.*, 2013). In recent years, capsule polysaccharide (CPS)-conjugate vaccines have seen successes in the prevention of meningococcal disease from Serogroups A, C, W, Y(Helena De Oliveira *et al.*, 2017). However, the development of a conjugate vaccine for serogroup B has been particularly challenging due to the poor immunogenicity of its (α2,8)-linked polysialic acid capsule, which closely resembles human neural cell adhesion molecules (NCAM), thus raising concerns over autoimmunity (Finne, Leinonen and Mäkelä, 1983). Although resistance to antibiotics within strains of *Nm* remains rare, recent reports have shown an increase in incidence of these resistant strains in the UK (Willerton *et al.*, 2021) and the US (McNamara *et al.*, 2020) populations, calling for advancements in the characterisation of *Neisseria* host-pathogen adhesion mechanism for leading to the development of new strategies in the prevention and treatment of infection.

One of the primary adhesion factors of *Nm* is its Type IV pili (T4P)**,** filamentous structures that extend past the bacterial outer membrane enabling both mobility and adhesion to host cells (Craig, Forest and Maier, 2019). The pilus is primarily composed of the major pilin subunit, pilE, and is built in a helical arrangement with its conserved hydrophobic N-terminal α-helices forming a tightly packed core, leaving its variable C-terminal domains forming the outer layer contributing to flexibility and surface interactions. Initially, the pilE subunits are anchored in the inner membrane and assembled into the growing filament through the action of pilus assembly machinery, which includes ATPases and inner membrane platform proteins(Craig, Forest and Maier, 2019), see **Figure 4.1**. During assembly, the N-terminal α1 helix of pilE interacts with adjacent subunits, stabilising the core via charge complementarity, while a central segment of α1 unravels to allow proper packing. As more subunits are added, the pilus extends through the outer membrane secretin channel and is displayed on the bacterial surface. This construct enables T4P to perform multiple functions, including twitching motility, biofilm formation, and, most importantly, host-cell adhesion, a key step in colonisation(Exley *et al.*, 2009).

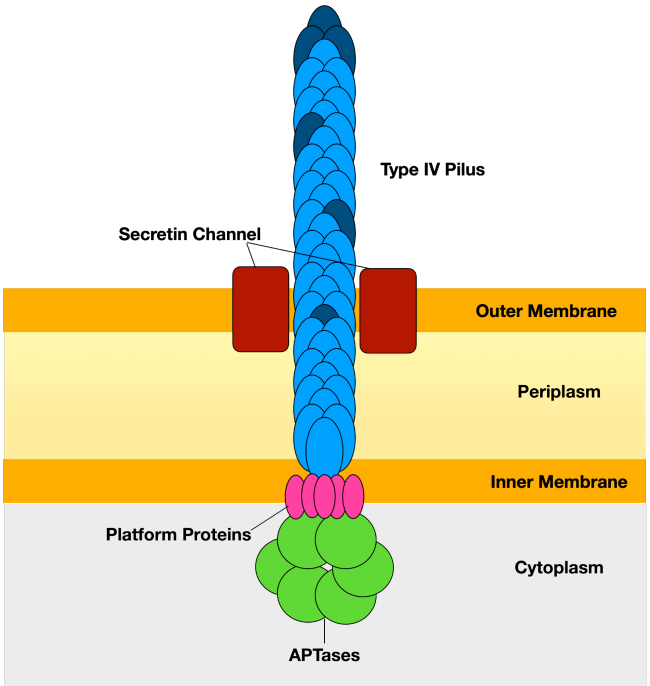


Figure 4.1: Framework of the Type IV pilus machinery, with labels provided for the APTases(green), platform proteins(pink), secretin channel(wine), and the Type IV pilus, where the major pilin subunits are represented in light blue and minor pilin subunits represented in dark blue.

Minor pilins, present in smaller quantities, are also a crucial component of the T4P. Minor pilin subunits can be categorised into core and non-core minor pilins. Core minor pilins, which include pilH, pilI, pilJ and pilK, form complexes near the tip of the pilus and promote the initiation of the helical pilus assembly. Deletion of the genes encoding the core minor pilins have been reported to result in non-piliated bacteria (Carbonnelle *et al.*, 2005). Non-core minor pilin subunits are distributed throughout the pilus and have been reported to promote additional functions such DNA binding and adherence(Roux, Spagnolo and de Bentzmann, 2012). In particular, pilV has been reported to promote host cell adherence and is suggested to play a role in binding to host receptors(Bernard *et al.*, 2014). T4P are not exclusive to *Nm* and are common to other bacteria within the *Neisseria* species, including *Ng* as well as other bacterial species such as *Pseudomonas aeruginosa* and *Myxococcus xanthus*. T4P has been reported to be important for initial adhesion to the host cell with literature identifying contact with the platelet-activating factor receptor (PAFr) on the surface of endothelial cells lining the airway system (Jen *et al.*, 2013). They may also be an important factor in colonisation and thus targeting the pili has been shown to reduce the ability of the bacterium to form stable colonies (Denis *et al.*, 2019).

The glycan landscape of the *Nesseria* bacterium has been very well documented (Mubaiwa, Semchenko, *et al.*, 2017) and is characterised by a vast array of surface-exposed carbohydrate structures such as capsular polysaccharides, lipo-oligosaccharides (LOS) and glycans extending from glycoproteins (Mubaiwa, Hartley-Tassell, *et al.*, 2017). These glycan structures aid the bacterium's pathogenesis and immune evasion (Schneider *et al.*, 2007))(Gault *et al.*, 2015). In addition to that, the bacteria can interact with the various glycan structures on the host cell affecting immune suppression and immune evasion, and facilitating adhesion(Gasparini *et al.*, 2015). For example, the capsular polysaccharide of Serogroup B allows *N. meningitidis* to evade complement-mediated killing(Jarvis and Vedros, 1987), while LOS undergoes sialylation to mimic host cell surfaces(Kahler *et al.*, 1998), reducing immune detection. The ability of *Nm* to modify its surface glycans through phase variation and antigenic variation, where the bacterium alters the expression of antigen on its surfaces associated virulence factors including its T4P, further complicates immune responses and enhances its ability to persist within the host(Davidsen and Tønjum, 2006).

In this study, I explored the architecture of *Nm*'s T4P system by building a model of the system using the major pilin subunit, pilE, and analysing its structure and dynamics through Molecular

Dynamics (MD) simulations. I examined the differences in the structure and sequence between pilE and the minor pilin subunit, pilV, and assessed how incorporating the pilV subunits into my T4P model would affect the dynamics of the whole system.

I then investigated the glycan binding specificity of the T4P system by first using a combination of docking coupled with MD to identify viable binding pockets within our system. Using those results, I built an MD-based glycan array to simultaneously evaluate the T4P's binding specificity for key glycan epitopes.

I collaborated with Dr Freda Jen and Dr Chris Day in Prof Micheal Jennings' group at the Institute for Biomedicine and Glycomics at Griffith University in Gold Coast, QLD, Australia to conduct Surface Plasmon Resonance(SPR) experiments and determine the binding affinity of the *Nm* pilin for the glycan epitopes.

4.2 Methods

4.2.1 Computational Methods.

The single pilE subunit structure was generated using AlphaFold2 (Jumper *et al.*, 2021) using the genomic data of the Serogroup B strain C311(Jen *et al.*, 2021). The starting structures for the bacterial glycans and all glycan epitopes tested were generated using the carbohydrate builder tool on GLYCAM-WEB(https://glycam.org/). I performed MD simulations for each glycan structure, and the resulting trajectories were uploaded to the GlycoShape Glycan Structure Database (Ives *et al.*, 2024), from which the most populated structures for each glycan were obtained. I ran a minimum of 3 replicas for each glycan structure. For glycan structures that produced 2 energetically available conformers from the carbohydrate builder tool, I ran 2 replicas for each conformer. For structures, that produced 3 or more energetically available conformers, I ran a replica for each conformer. The parameters for the phosphoryl-choline(ChoP) were generated using AMBER18(Lee *et al.*, 2018) software package detailed in **Chapter 1**.

I used the *tleap* program from AMBER18(Lee *et al.*, 2018) to bond the *O*-linked bacterial trisaccharide, Galβ1-4Galα1-3DATDH, and the two ChoP groups to S63, S157 and S160(Jen *et al.*, 2013) respectively. I arranged the pilE subunits to make the 26-mer pilE system using

the cryo-EM structure PDBid 5KUA(Kolappan *et al*., 2016) as a reference. I superimposed the pilE subunits onto the pilin subunits of the 5KUA structure using the backbone of the alpha-helix as my alignment basis I then analysed the dynamics of this system using MD simulation.

I screened potential binding sites by molecular docking on the pilE monomer and dimer using HADDOCK (Dominguez, Boelens, and Bonvin, 2003) with Neu5Acα2-3Galβ1-4GlcNAc and Neu5Acα2-6Galβ1-4GlcNAc as glycan epitopes. Promising 3D models for the complexes were only found in the case of pilE dimers, and I tested these for conformational stability by extensive deterministic MD sampling. Through this approach, I identified stable conformations in which the glycan epitopes occupied the proposed binding site with the same pose for over 70% of the MD production trajectory, encompassing over 1.5 μs of simulation time.

To populate the *Nm* pilin array, I superimposed the terminal substituent of each glycan epitope onto the terminal sialic acids of both α2,6- and α2,3-linked Sia-LacNAc bound to the *Nm* pilE dimer. I then superimposed this construct onto pilin subunits of the 26-mer *Nm* pilE system using the backbone of the alpha-helix as the alignment basis. I then removed the *Nm* pilE dimer, leaving the glycan epitope positioned between two adjacent *Nm* pilin subunits. I assessed the stability of all these epitopes within the binding sites using MD simulations.

All MD simulations were run in 200 mM NaCl concentration using the AMBER18(Lee *et al.*, 2018) software package. Protein atoms and counterions were represented by the AMBER ff14SB(Maier *et al.*, 2015) parameter set, while the carbohydrate atoms were represented using the GLYCAM06j-1version of the GLYCAM06 force field(Kirschner *et al.*, 2008). The ChoPs were represented using GAFF2(Wang *et al.*, 2004). Water molecules were represented using the TIP3P water model(Jorgensen *et al.*, 1983).

All MD simulations were run following a similar protocol, starting with an energy minimization stage through 500,000 steps of steepest descent. During minimization, only hydrogen atoms, water molecules, and counterions were left unrestrained, while the position of all other atoms were kept restrained with a potential weight of 5 kcal/mol $Å^2$. Following minimisation, the system was brought up to standard temperature through a heating phase in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During phase one the temperature was raised from 0 to 100 K where the volume was kept constant and from 100 to 300 K during phase two where the pressure was

kept constant. The system was then equilibrated for 1 ns in the NPT ensemble to a 1 am pressure with the Berendsen barostat. During these equilibration stages, we used the same position restraints as for minimisation.

For the pilE dimer complexes generated with molecular docking, conformational equilibration was initiated with restraints applied to both the glycan epitopes and the alpha helices of the pilin subunits. Following equilibration, the production phase began with the removal of restraints on the glycan epitopes.

For the full 26-subunit pilE system and 26 pilE/pilV system, both with and without the glycan epitopes, a stepwise equilibration process was employed. Restraints were initially put on the glycan epitopes and pilin subunits 1-7 and 19-26 which were then gradually relaxed to the glycan epitopes and subunits 1-3 and 24-26, and subsequently narrowed down to the the glycan epitopes and alpha helices of subunits 1-3 and 24-26. Production phases followed with all restraints on the heavy atoms removed except for those for the alpha helices of the pilin subunits 1-3 and 24-26.

4.2.2 Surface Plasmon Resonance (SPR).

As part of this project in February 2024 I visited Prof Micheal Jennings' research group at the Institute of Biomedicine and Glycomics at Griffith University in Gold Coast, QLD, Australia as part of my PhD international placement. During this internship, I worked with Dr Freda Jen and Dr Chris Day who ran SPR experiments designed to test the results I obtained from the pilin array screening by MD simulations. The experiments were run using the method described below.

Vortex released pilin from *Nm* with differing glycosylation modifications were immobilised on a CDH chip (Sartorius) and run using the Pioneer FE SPR system (Sartorius). Immobilisation was carried out at pH 4.5 and the protein flowed at 10 µL per minute for 5 minutes over an EDC/NHS activated surface. The surface was quenched with ethanolamine at 10 µL per minute for 7 minutes. A pilin deletion mutant bacteria vortexed and treated the same as the piliated bacteria was used as the negative surface control for double reference subtraction. Glycans were flowed over immobilised protein at 50 µL per minute at a maximum concentration of 100 µM using the OneStep kinetics with regeneration. Data from each run was exported to QDat analysis software (Sartorius) and analysed using the standard K_A/K_D model. The calculated binding affinity (K_D) are from a minimum of two biological replicates.

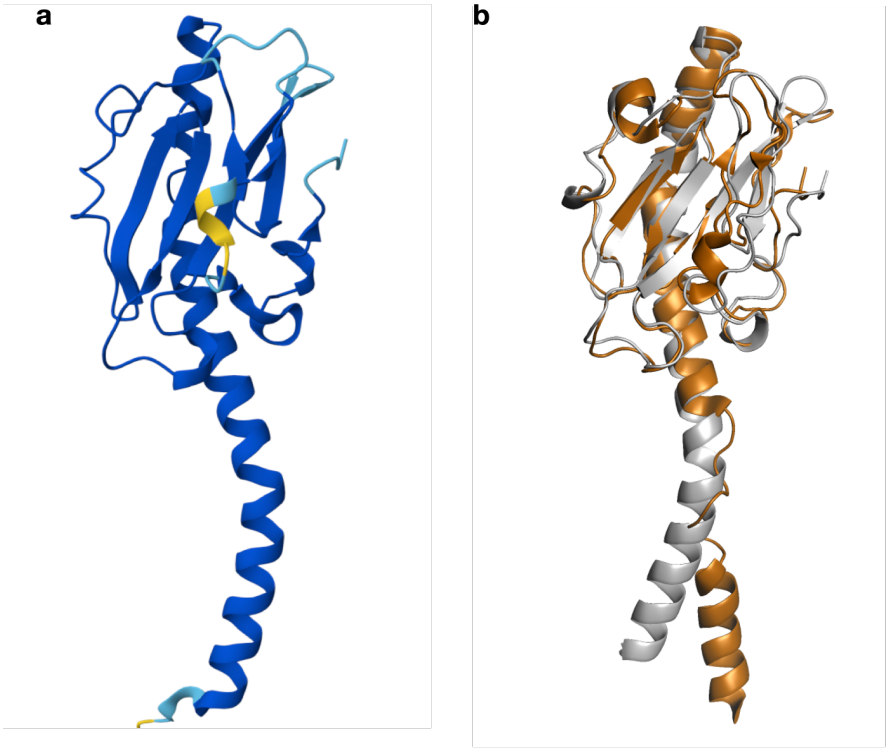
4.3 Results

In this section, I will present the results of the molecular dynamics (MD) simulations on a 26-mer pilE T4P system, which I will refer to from now on as pilE systems. I will show how the simulations highlight the functional role of the pilE monomeric units and of their PTMs in the pilE system's architecture and stability. In addition to studying the pilE-only system, I also investigated a mixed pilE/pilV system, incorporating pilV subunits into the pilus structure to assess their impact on stability and function. We will then present the results of the MD-based glycan-binding screening I performed using the pilE-only system through an approach we called 'Pilin Array'.

4.3.1 Structure and Dynamics of the *Nm* T4P 26-mer: The PilE system

To ensure my study focused on the immunoevasive Serogroup B and to keep consistent with the posttranslational modifications detailed later, I decided to build the pilE system using pilE structure from strain C311(Jen *et al.*, 2013). As there was no crystal structure available for this strain, I used AlphaFold (Jumper *et al*., 2021) to generate the pilE structure based on its sequencing data (Jen *et al*., 2021), see **Figure 4.2a**.

I constructed the starting structure of the pilE system using the cryo-EM structure 5KUA (Kolappan *et al*., 2016) as a reference. I superimposed the pilE subunits from strain C311 onto the pilin subunits of 5KUA, using the backbone of the alpha-helix of each subunit as the alignment basis. To assess structural differences, I compared the 3D structures of the pilE subunits from 5KUA, which originate from strain 8013 (Serogroup C), with the AlphaFold-predicted model of strain C311. This comparison was performed via structural alignment of the backbone atoms using PyMOL (www.pymol.org), see **Figure 4.2b**. The alignment returned an RMSD score of 1.3 Å and sequence analysis of both structures revealed a similarity score of 82.5% and identified several mutations between the two strains, as shown in **Figure 4.2c-d**.



c

```
Strain 8013     FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGEWPGD 60
Strain C311     FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGEWPGN 60
                ************************************************************:

Strain 8013     NSSAGVATSADIKGKYVQSVTVANGVITAQMASSNVNNEIKSKKLSLWAKRQNGSVKWFC 120
Strain C311     NTSAGVASSSTIKGKYVKEVTVANGVITATMLSSGVNKEIQGKKLSLWAKRQNGSVKWFC 120
                *:*****:*: ******:.********** * **.**:**:.******************

Strain 8013     GQPVTRTTATATDVAAANGKTDDKINTKHLPSTCRDDSSAS    161
Strain C311     GQPVTRNDTDDT-VAAVAADNTGNINTKHLPSTCRDASDAS    160
                ******. :  * ***. ... .:************ *.**
```

d

	Strain 8013	Strain C311
Strain 8013	100.00	82.5
Strain C311	82.5	100

Figure 4.2. Panel a) Structure of N*m* pilE subunit from strain C311(GenBank accession number QXZ29465.1) generated using AlphaFold. The structure is coloured based on the per-residue model confidence score(pLDDT) where dark blue represents very high confidence(pLDDT > 90), cyan represents high confidence(90 > pLDDT > 70) and yellow represents low confidence(70 > pLDDT > 50). **Panel b)** Structural alignment of the AlphaFold model(white) with the pilE subunits from the PDB structure 5KUA(orange). **Panel c)** Sequence alignment of the pilE AlphaFold sequence (Strain C311), against the sequences of the pilE subunits from PDB 5KUA(strain 8013). Sequence alignment was conducted using Clustal Omega(https://www.ebi.ac.uk/jdispatcher/msa/clustalo)(Sievers *et al.*, 2011). **Panel d)** Percentage identity scores for the sequence alignment. Molecular rendering was done with pymol (www.pymol.org).

The pilE system counts 26 pilE subunits assembled in a helical arrangement, see **Figures 4.3a-c**. Within this architecture, the alpha helices of the pilE subunits (aa 1-55) comprise the core of the system, while the globular domain (aa 56-160) is exposed on the surface. The pilE subunits

are packed in such a way that the globular domain of each subunit aligns with the globular domain of the adjacent subunit creating a small gap between them, see **Figure 4.3b**. Each pilE subunit corresponding to strain C311#2(Jen *et al.*, 2013) contains the following post-translational modification (PTM): an *O*-linked trisaccharide, Galβ1-4Galα1-3DATDH, at S63 and two phosphorylcholines (ChoP) at S157 and S160, see **Figure 4.3d**. I incorporated these post-translational modifications into the pilE structure using the 'bond' functionality in *tleap*. The two ChoP groups are located at the C-terminus of each pilE subunit while the trisaccharide extends towards the pilE located immediately above in the preceding helical turn.

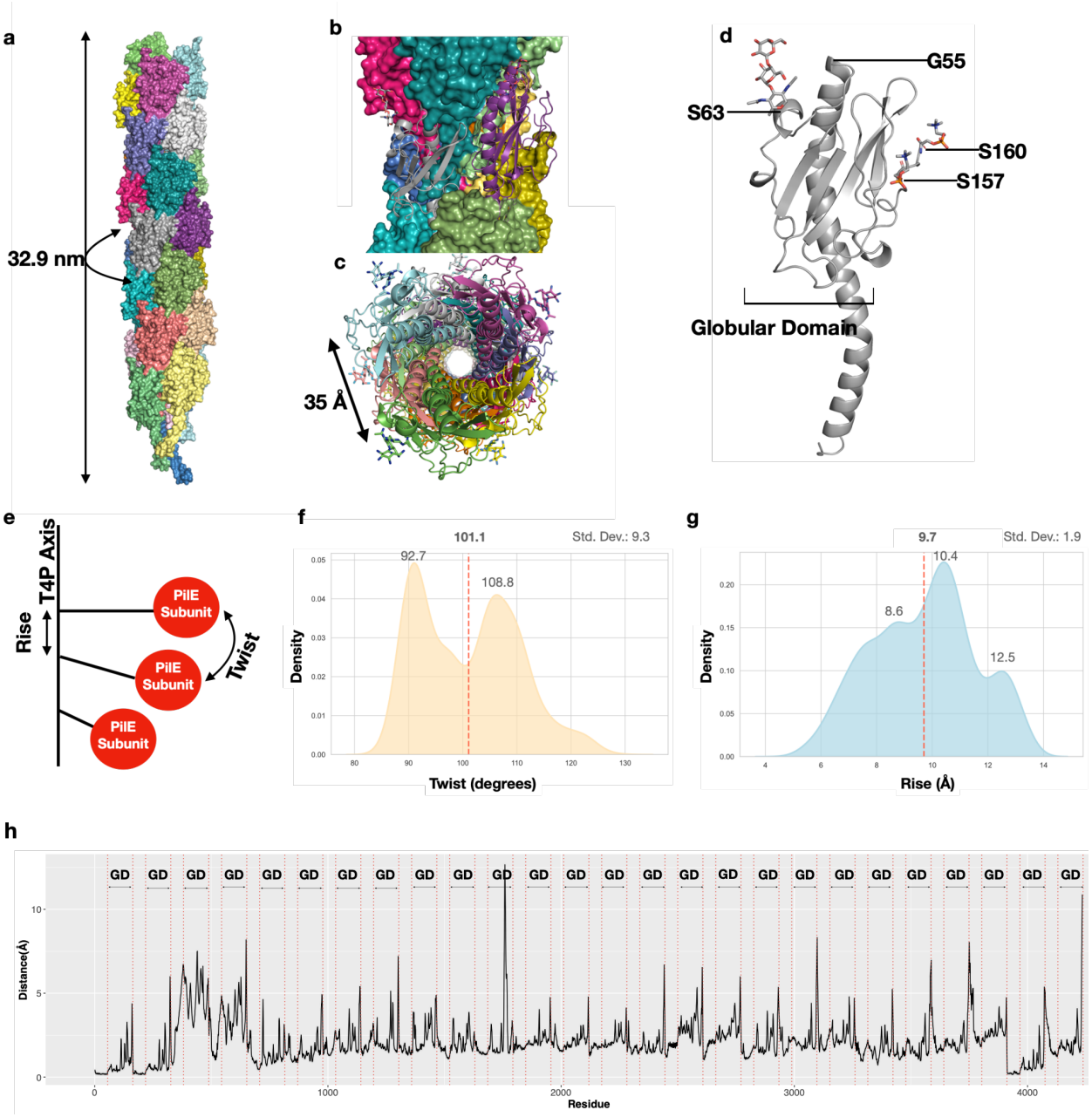


Figure 4.3. **Panel a)** Graphical representation of the reconstructed pilE-only T4P structure in this study. Each pilE subunit is represented by a different colour. **Panel b)** Close-up of the pilE-only T4P structure highlighting the conformation of two adjacent pilE subunits rendered in cartoon representation **Panel c)** Top-view of the pilE-only T4P structure showing the helices

association stabilising the T4P core. **Panel d)** Graphical representation of a single pilE subunit rendered in cartoons, with the position of the amino acids G55, S63, S157, S160 highlighted. **Panel e)** Illustration of the rise and twist parameters used to define the structure of the T4P. **Panel f)** Kernel density estimates (KDE) distribution of twist values measured for the pilE subunits throughout the MD simulation.**Panel g)** KDE distribution of rise values measured for the pilE subunits throughout the simulation. The average values for each parameter are labelled and indicated with red dotted lines. The major peaks for each parameter are also labelled. The standard deviation(Std. Dev.) for the rise and twist parameters are listed at the top of the plots. **Panel h)** Line plot of the RMSF values calculated from the MD trajectories of the backbone atoms of the pilE-only T4P structure. Red dotted lines and labels identify the residues corresponding to the globular domains(GD) of the pilEs. Molecular rendering done with Pymol (www.pymol.org).

Results from the molecular dynamics (MD) simulation of the pilE system demonstrate a general trend across all the pilE subunits within our framework. The helical rise and twist values of the subunits in the system shown in **Figures 4.3f-g**, are consistent with the cryo-EM structure we used as reference (Kolappan *et al.*, 2016), indicating the relative stability of the helical packing at the core. The RMSF plot, see **Figure 4.3e**, shows that the globular domain of the pilE subunits has a greater flexibility relative to the core region. The compact arrangement of the system restricts the movement of the helical regions located within the core. The globular domain of the pilE subunits consistently engage in hydrogen bonding and dispersion interaction with the residues on adjacent pilin subunits as they move within proximity of each other. In particular, the globular domain facing each other across the gap between adjacent pilin subunits are rich in lysine residues which can form strong, dynamic interactions with the side chains and backbones of residues nearby.

4.3.2 Structure and Dynamics of the *Nm* T4P 26-mer: The PilE/PilV system

PilV has been reported to play a significant role in adhesion alongside pilE, with evidence showing that ΔpilV mutants exhibited adhesin defects (Barnier *et al.*, 2021). Unlike pilE, which contains a hypervariable region contributing to antigenic variation and immune evasion, pilV is more conserved across strains(Cehovin *et al.*, 2010). Although distributed throughout the pilus filament in smaller quantities, the conserved nature of the minor pilV makes it a potential target for therapeutic intervention. To investigate the dynamics and glycan specificity of a T4P pilus system containing pilV, I remodeled the 26 pilE system and substituted 5 pilV subunits across the system as shown in **Figure 4.4a**. The pilV subunits were distributed across the filament to ensure reproducibility. I generated the pilV structure, as shown in Figure 5.4b, using Alphafold (Jumper *et al*., 2021) with the sequence data for the pilV protein from strain C311 (Jen *et al*., 2021).

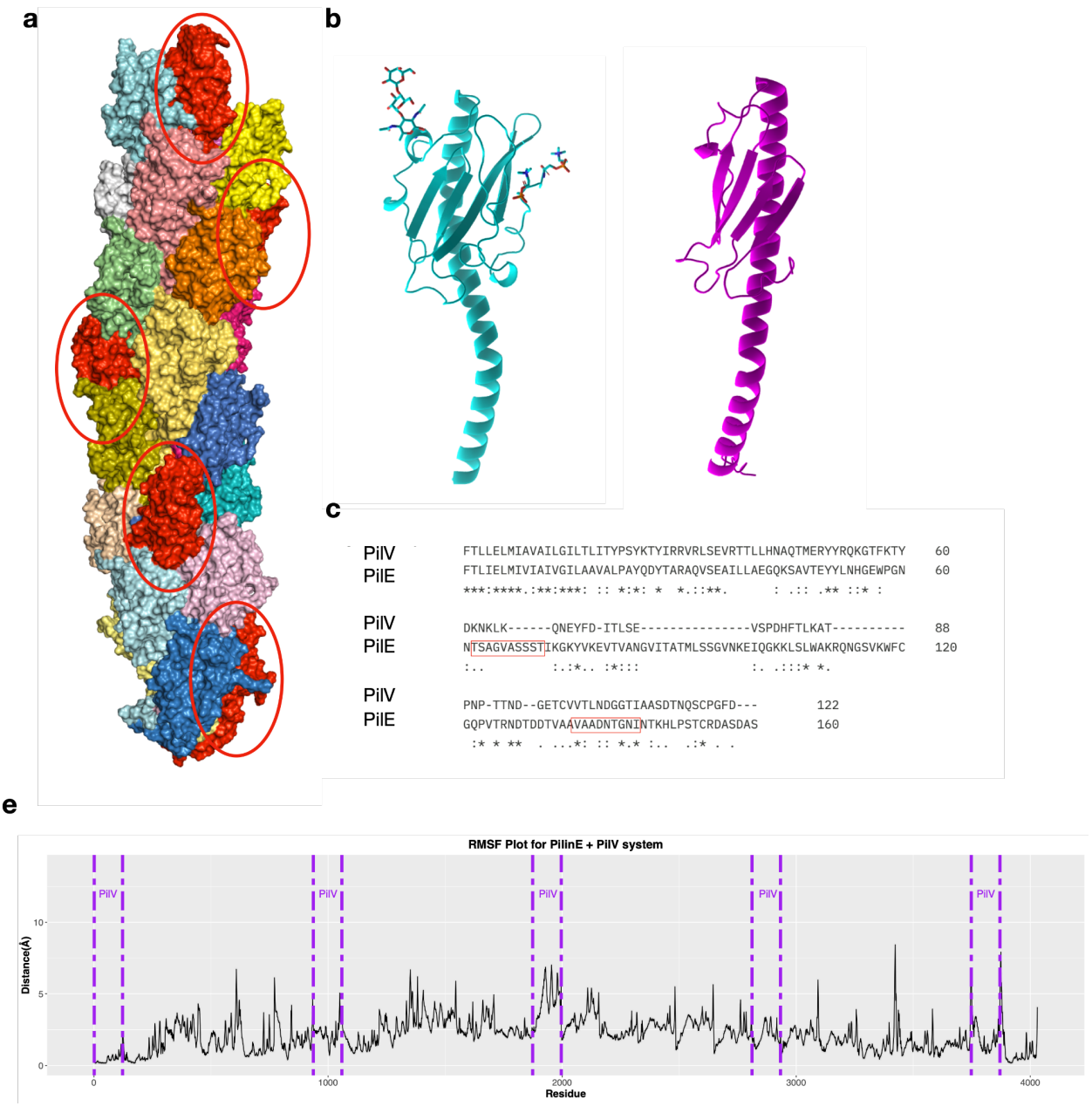


Figure 4.4. Panel a) Structure of a T4P carrying both pilE and pilV subunits. PilV subunits are rendered in red, highlighted within circles. **Panel b)** Graphical representation of isolated pilE (left) and pilV (right) subunits, rendered with cartoons in cyan and in purple, respectively. **Panel c)** Sequence alignment of pilE and pilV. The hypervariable regions of the pilE are highlighted in a red box. **Panel d)** Line plot of the Root Mean Square Fluctuation (RMSF) values from the MD trajectories of the backbone atoms of the pilE/pilV system. The dotted lines highlight the RMSF values corresponding to the pilV subunits. Molecular rendering with pymol (www.pymol.org).

The pilV is a shorter protein than pilE, see **Figures 4.4c-d**, and lacks both the PTMs and the larger globular domain characteristic of pilE. Due to the smaller globular domain of the pilV, there is a larger gap between adjacent pilE and pilV subunits compared to the tighter arrangement between pilE subunits.

Molecular dynamics simulations of this system displayed similar behavior to the 26 pilE system. However, the gap between pilE and pilV subunits slightly expanded throughout the simulation. PilV was able to establish electrostatic contacts on one side of its globular domain, but due to its smaller size, it could not fully interact with the globular domain of adjacent pilE subunits, leaving a persistent gap. Despite this, the pilE-pilV system maintained its structural integrity for over 1 μs. However, once glycan epitopes were introduced, the structure became unstable. The glycans either exited the pilus construct entirely or fell through the gap between pilV and pilE subunits, disrupting the core. This caused the pilin subunits to protrude outward, ultimately leading to the collapse of the overall structure.

4.3.3 Finding the glycan binding site: Molecular Docking and Dynamics on a PilE Dimer

Earlier work reports that the *Nm* T4P interacts with the CD147 receptor by binding to a triantennary *N*-glycans present on its surface(Le Guennec *et al.*, 2020). Binding affinity was found to be enhanced by terminal sialylation, while fucosylation significantly hinders binding. To investigate these findings, I initially used molecular docking using the glycan epitopes Neu5Acα2-6Galβ1-4GlcNAc and Neu5Acα2-3Galβ1-4GlcNAc as ligands, to screen for potential binding sites within the pilE monomers. The 3D structures of these ligands were determined separately in solution by MD simulations according to a protocol described in Methods and now adopted for the generation of the structures in the GlycoShape database(Ives *et al.*, 2024). This test did not prove to be successful as I was unable to identify any stable binding pose on single pilE subunits. My next step was inspired by earlier results by Wimmerova *et al.* (Wimmerova *et al.*, 2003), showing a singular multivalent glycan binding mode in the *Aleuria aurantia* lectin, where fucose is specifically recognised and binds between adjacent propellers. Accordingly, I decided to rebuild pilE dimers following the architecture of the cryoEM structure PDBid 5KUA(Kolappan *et al.*, 2016) as a template, and through molecular docking we identified a viable binding pocket, rich in polar residues, located between the globular domains of two adjacent pilE subunits within the same helical turn. I ran MD simulations to evaluate the stability of the highest scoring and most promising poses, see **Figure 4.5**.

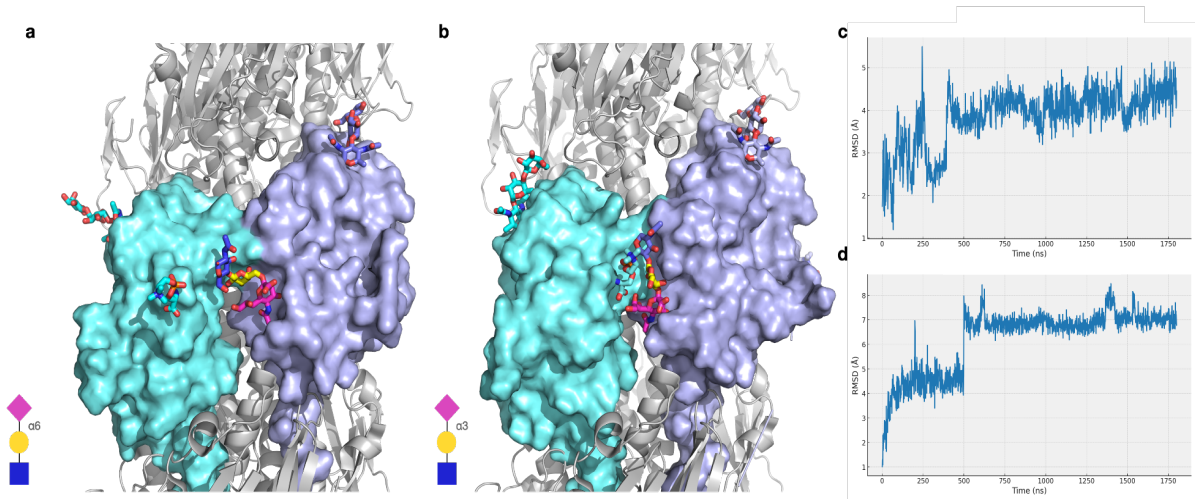


Figure 4.5. **Panel a)** Structure of the bound Neu5Acα2-6Galβ1-4GlcNAc to the pilE dimer. **Panel b)** Structure of the bound Neu5Acα2-3Galβ1-4GlcNAc to the pilE dimer. **Panel c)** Root Mean Square Deviation (RMSD) values obtained from the MD simulation of the pilE dimer in complex with Neu5Acα2-3Galβ1-4GlcNAc. **Panel d)** RMSD values obtained from the MD simulation of the pilE dimer in complex with Neu5Acα2-3Galβ1-4GlcNAc. Molecular rendering with pymol (www.pymol.org)

As shown by the RMSD values, during the MD simulation both glycan epitopes readjusted their conformation within the binding pocket, relative to the starting structure obtained from docking. In both cases, the optimised conformations remained stable for over 1 μs. Following this step, I decided to investigate if these poses were consistent with the recognition and binding of a whole triantennary *N*-glycan structure as experimentally determined(Le Guennec *et al.*, 2020). I also wanted to assess whether core fucosylation, which is located away from the epitope, would hinder binding.

As for the epitopes, the structure of the whole *N*-glycans were determined separately by MD simulations of the isolated systems in solution(Ives *et al.*, 2024). I then superimposed the whole structure of the sialylated and core-fucosylated triantennary *N*-glycan to the bound pose using the shared SialylLacNAc as the alignment basis, **see Figure 4.6a**. MD simulation of this structure showed that the sialylated and core-fucosylated triantennary *N*-glycans remained stably bound to the pilE dimer throughout the 1 μs MD simulation, see **Figure 4.6b**.

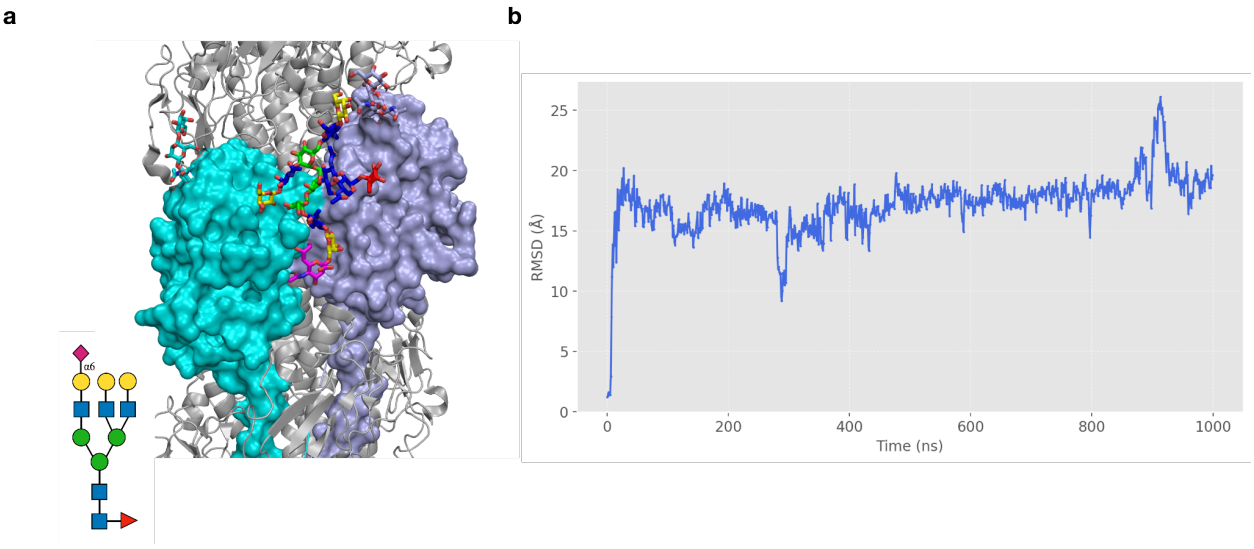


Figure 4.6 Panel a) Structure of the bound triantennary *N*-glycan to the pilE dimer. **Panel b)** Root Mean Square Deviation (RMSD) values obtained from the MD simulation of the pilE dimer in complex with the triantennary structure. Molecular rendering with pymol (www.pymol.org).

To investigate the effect of branch fucosylation, I occupied the identified binding pocket with a fucosylated LacNAc using the same docking method with HADDOCK and ran additional MD simulations. While the fucosylated LacNAc remained bound for the duration of the simulation, when the full triantennary structure was superimposed, the structure would sterically clash with the pilE protein leading to an incompatible fit see **Figure 4.7**.

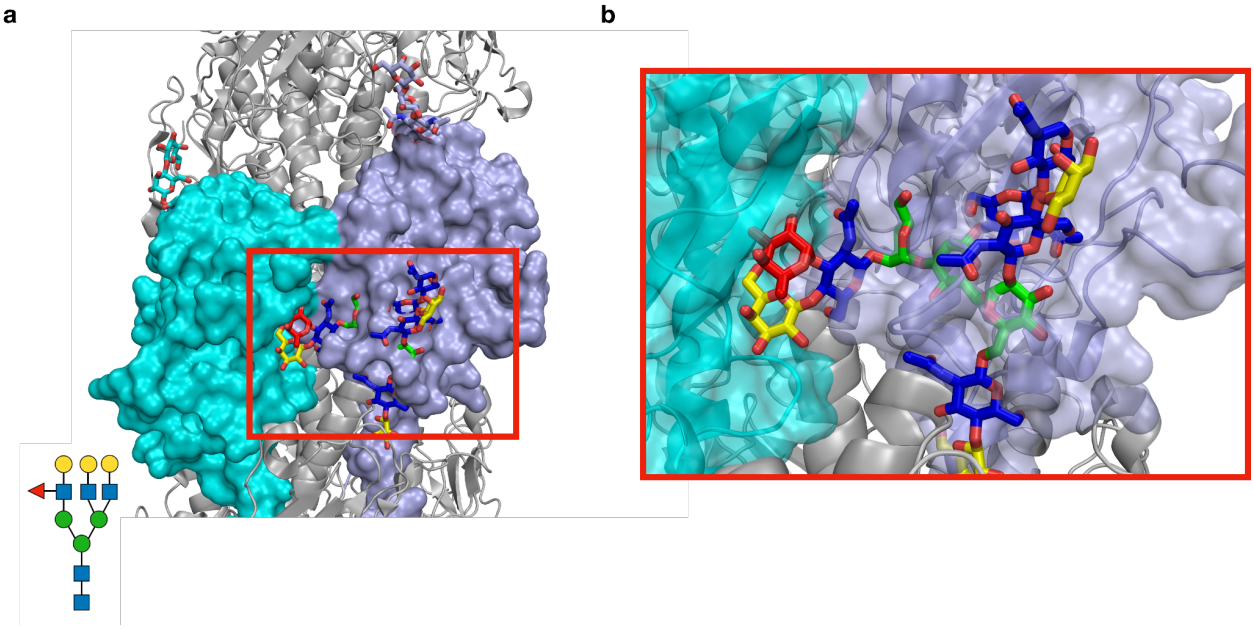


Figure 4.7 Panel a) Structure of the superimposed branch fucosylated triantennary *N*-glycan to the pilE dimer. **Panel b)** Close-up of the branch fucosylated triantennary N-glycan highlighting the clash with the pilE protein.

4.3.5 'Pilin Array' Screening by MD simulations

I built the Pilin array by populating the 26-mer pilE system with a chosen subset of glycan epitopes, see **Figure 4.8**. For greater sampling, I used two replicas of every glycan epitope, except for LacNAc and SialylLacNAc epitopes, for which I used only one replica. I had extensively analysed the binding motif of the SialylLacNAcs when studying the binding pocket between the isolated dimer in **Section 4.3.2.** I superimposed the terminal substituent of each glycan epitope onto the terminal sialic acids of both α2,6- and α2,3-linked SialylLacNAc bound to the *Nm* pilE dimer. This glycan-bound construct was then aligned onto the pilin subunits of the 26-mer Nm pilE system using the backbone of the alpha-helix as the alignment basis. After alignment, the pilE dimer was removed, leaving the glycan epitope positioned between two adjacent pilin subunits. To ensure the stability of the pilin array and test whether the pilE system could accommodate multiple glycan epitopes effectively, I only populated 13 of the 26 available binding pockets, leaving an empty pocket between each occupied dimer such that the pilin subunits would have space to adjust and maintain the structural integrity of the overall Pilin Array.

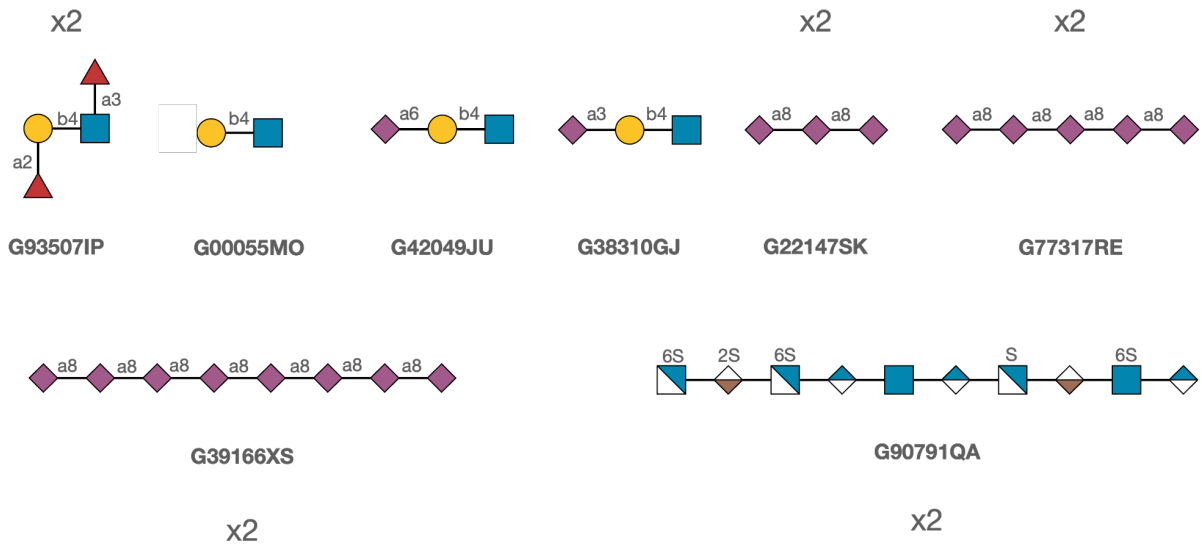


Figure 4.8. Graphical representations of glycan epitopes used to populate the 26 pilE subunit system. Each glycan is labelled with its GlyTouCan(Tiemeyer *et al.*, 2017) identifier (https://glytoucan.org/).

I ran the MD simulation of the pilin Array for over 1.5 μs. The structural integrity of the array was maintained throughout the simulation time. There were some deviations in the rise and twist parameters in the pilin array compared to the unoccupied pilE system associated with the readjustment of the pilin subunits to accommodate the glycan epitopes, see **Figure 4.9,** but

overall the pilin array was stable throughout the simulation maintaining an average rise and twist values of 9.1 Å and 101.4°, respectively, compared to 9.7 Å and 101.4° in the unoccupied pilE system. The binding analysis of the different glycan epitopes to the pilin array is described below.

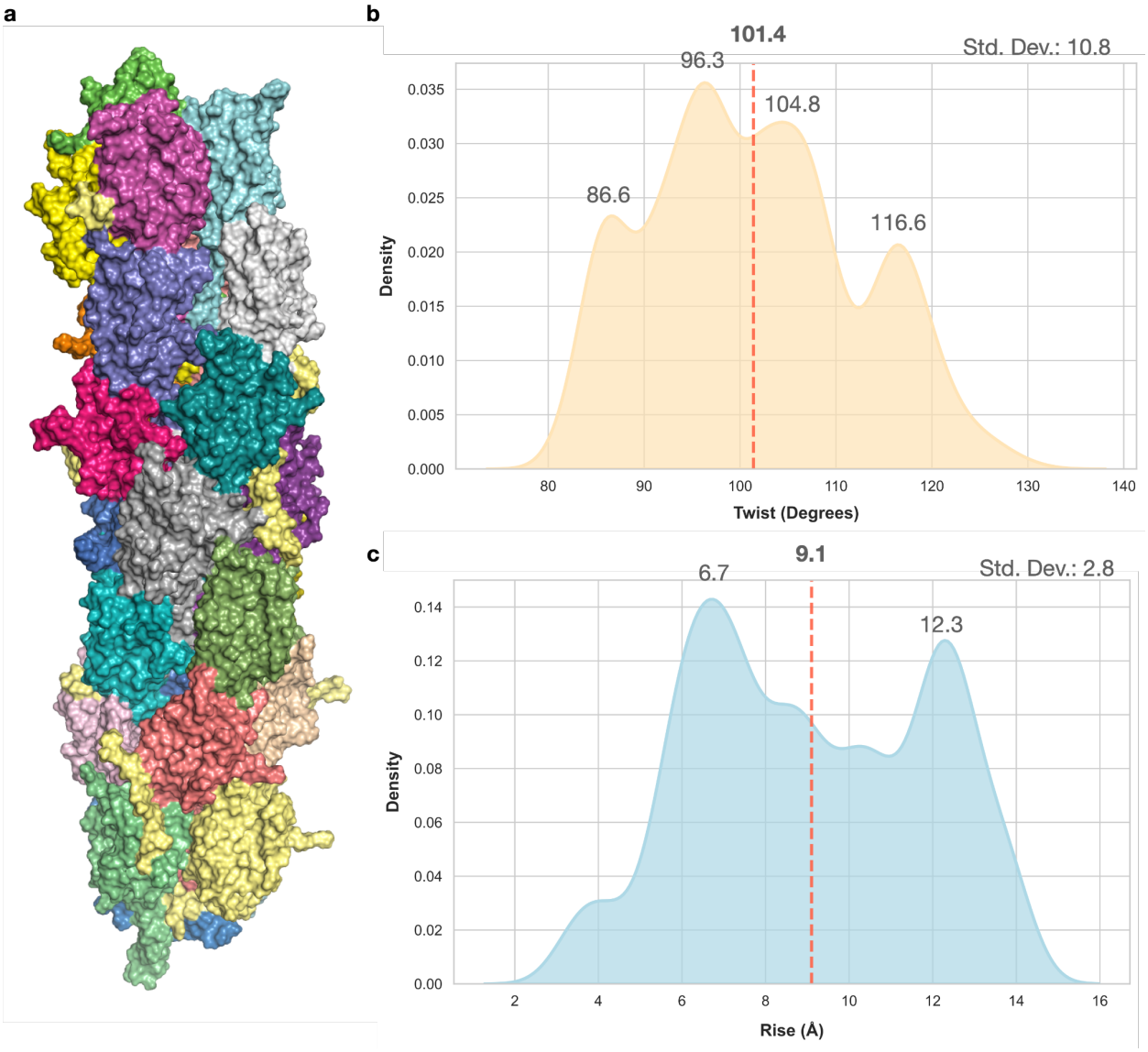


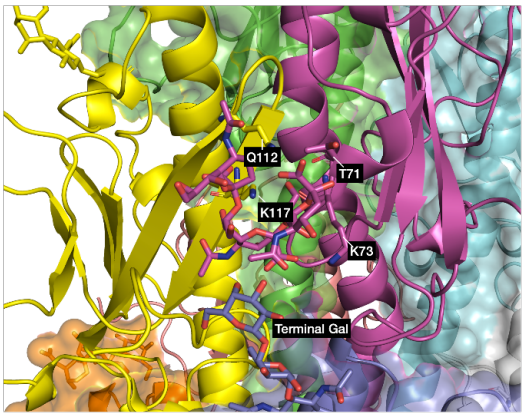
Figure 4.9. Panel a) Representative structure of the Pilin Array taken from the 1.5 µs MD simulation. Each pilE subunit is represented by a different colour. **Panel b)** Kernel density estimates (KDE) distribution of twist values measured for the pilE subunits throughout the MD simulation.**Panel c)** KDE distribution of rise values measured for the pilE subunits throughout the simulation. The average values for each parameter are labelled and indicated with red dotted lines. The major peaks for each parameter are also labelled. The standard deviation(Std. Dev.) for the rise and twist parameters are listed at the top of the plots. Molecular rendering done with Pymol (www.pymol.org).

4.3.6 Binding of Polysialic acids

The meningococcal capsule of *Nm* serogroup B strains carry α2-8 linked polysialic acids and could be a potential target for adhesion during the colonisation phase, where the pili would contribute to the interaction between different bacteria, contributing to the stability of the colony. To evaluate the potential for recognition and binding of polysialic acid motifs by T4P, I added to the stable structure of the pilE system three different lengths of α2-8 linked polysialic acids, namely a trisaccharide, a pentasaccharide, and an octasaccharide using the conformation of the bound terminal sialic acid we obtained from the MD of the pilE dimer in complex with sialyl-a(2-3/6)-LacNAc as a template for structural alignment.

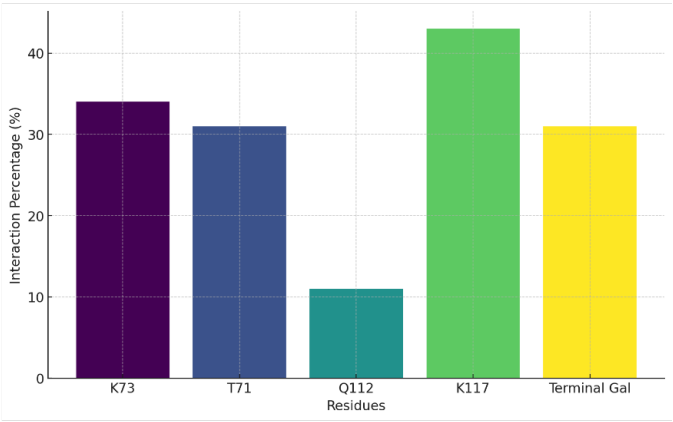
The analysis of the 1.5 μs MD simulation of the pilin array shows common features of the binding of the three polysialic acids. All polysialic acids remained in the binding site throughout the MD trajectory, see **Figure 4.10**, with poses stabilised by hydrogen bonds with specific polar residues in the binding pocket, and more persistent interactions with lysine residues lining the binding site pocket, see **Figure 4.10e**. The PTMs of the pilE subunits also contributed to binding stability; more specifically, the ChoP groups located at the C-terminus of the pilE subunits and the terminal Gal of the *O*-linked trisaccharide PTM from the pilE subunit located below the binding site, part of the next helical turn. As a common denominator across all polysialic acid lengths and simulation replicas, the polysialic acids consistently retained a highly stable "hook" structure characteristic of the trisaccharide minimum motif, see **Figure 4.10g**, which corresponds to the most populated conformation observed by MD simulation for the unbound polysialic acids in solution under physiological conditions (Ives *et al.*, 2024).

a

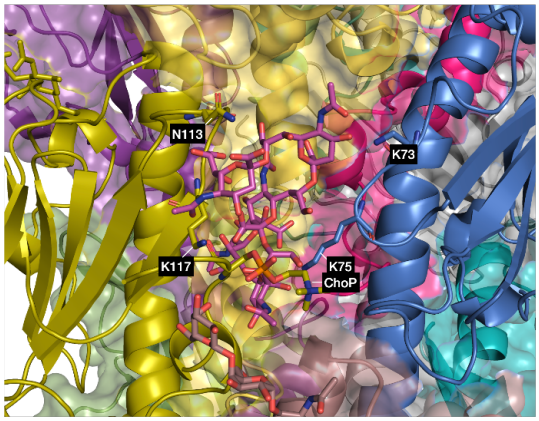


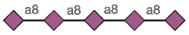
G22147SK

d



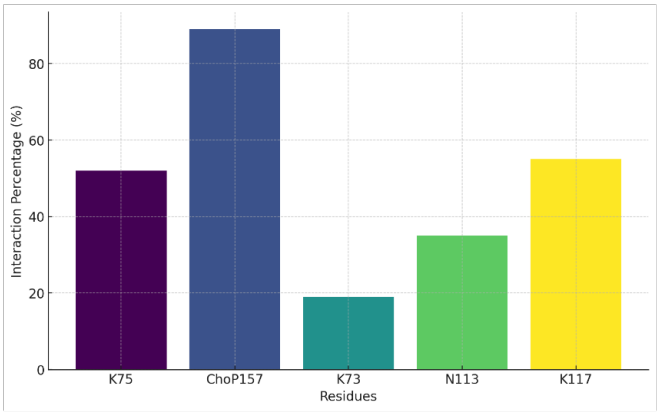
b



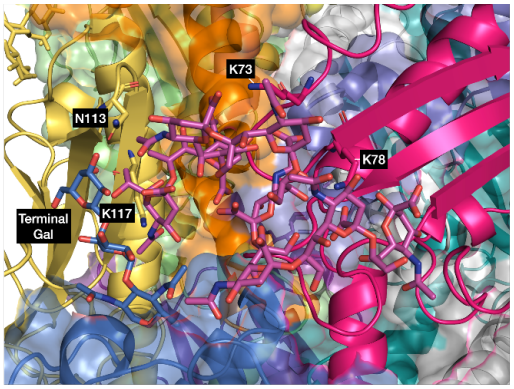


G77317RE

e

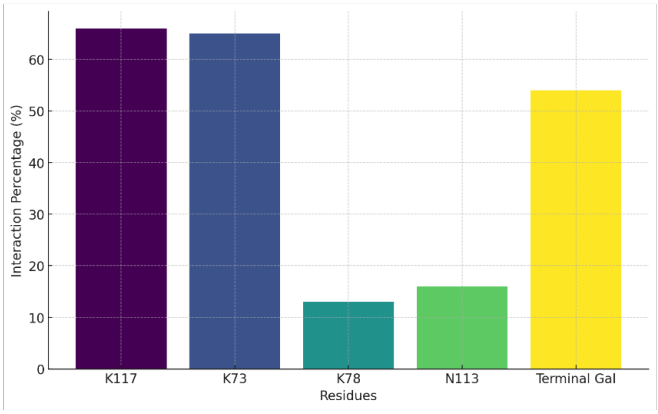


c



G39166XS

f



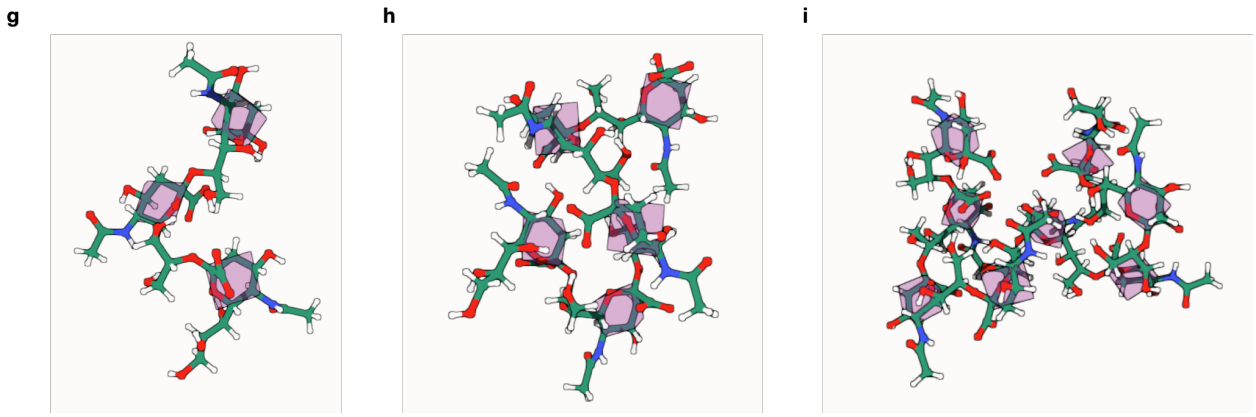


Figure 4.10. Panels a) to c) Stable binding poses identified for three different polysialic acid epitopes to the pilE system, namely Neu5Aca(2-8)- trimer in panel a), pentamer in panel b) and octamer in panel c). Highlighted are the key binding residues, namely key residues for trisaccharide, K73, T71, Q112, K117, Terminal Gal (from *O*-linked trisaccharide PTM); key residues for pentasaccharide, K75, ChoP157, K73, N113, K117; and key residues for octasaccharide, K117, K73, K78, N113, Terminal Gal (from *O*-linked trisaccharide PTM) **Panels d) to f)** Barplot of the hydrogen bond interactions stabilising the conformation of the polysialic acid epitopes. The values represent the percentage of time each residue maintained a hydrogen bond interaction throughout the MD simulations. **Panels g) to i)** Dominant conformations adopted by the polysialic acids when unbound in solution, i.e trisaccharide in panel g), pentasaccharide in panel h) and octasaccharide in panel i). Molecular rendering with Pymol (www.pymol.org)

4.3.7 Binding of Sialylated and non-sialylated LacNAc

Sialic acid (Neu5Ac) is abundantly present in the environment where *Nm* colonies grow *(Mubaiwa, Hartley-Tassell, et al., 2017)*. The capsular polysaccharide of *Nm* serogroup B carries α2,8-linked polysialic acid, which closely mimic the sialylated structures present on NCAM. In addition to this capsule, the lipooligosaccharide (LOS) structures lining the *Nm* outer membrane can also carry sialylated glycans. In particular, the L3 immunotype LOS contains a terminal Neu5Acα2-3Galβ1-4GlcNAc- structure and the expression of this sialylated LOS has been linked to invasive disease (Jones *et al*., 1992). Sialylated structures on the triantennary complex *N*-glycan of the CD147 receptor were found to enhance bacterial interaction (Le Guennec *et al*., 2020). To assess the effect of sialylation, I compared the recognition and binding of α2,6- and α2,3-sialylated LacNAc epitopes to non-sialylated LacNAc counterparts. Based on the results of the molecular docking followed by molecular dynamics on the pilE dimer construct, I already had insight on the stable binding poses of a(2-3/6) sialylated LacNAc. Yet, interestingly the binding results I obtained from the 26-mer Pilin Array show important differences. Notably, the binding of the α2,6-sialylated LacNAc was somewhat weakened within the larger system, despite its stability within the pilE dimer model. The α2,6-sialylated LacNAc was positioned between pilin subunits at the end of the T4P

structure, i.e. with no pilE subunits above it to provide additional stabilisation, analogously to the dimeric model. Yet, due to the complexity of the larger systems' dynamics and despite the restraints enforced on the helices of the four pilE at the T4P terminal turns, above and below, the distance between the globular domains of the two adjacent pilE subunits forming the binding site where the glycan was located broadened and the glycan was displaced.

The α2,3-sialylated LacNAc was positioned near the middle of the Pilin Array adopting a stable conformation, see **Figure 4.11**, stabilised by interactions with charged residues and PTMs. Because these interactions, in addition to those highlighted in the simulation of the pilE dimer, involve primarily the sialic acid, rather than the LacNAc, unsurprisingly we found that the LacNAc did not engage in any effective interactions with residues or PTMs in the binding site and was not able to occupy the binding pocket. As a note, the LacNAc was positioned in a non-terminal binding site near the upper-half of the Pilin Array.

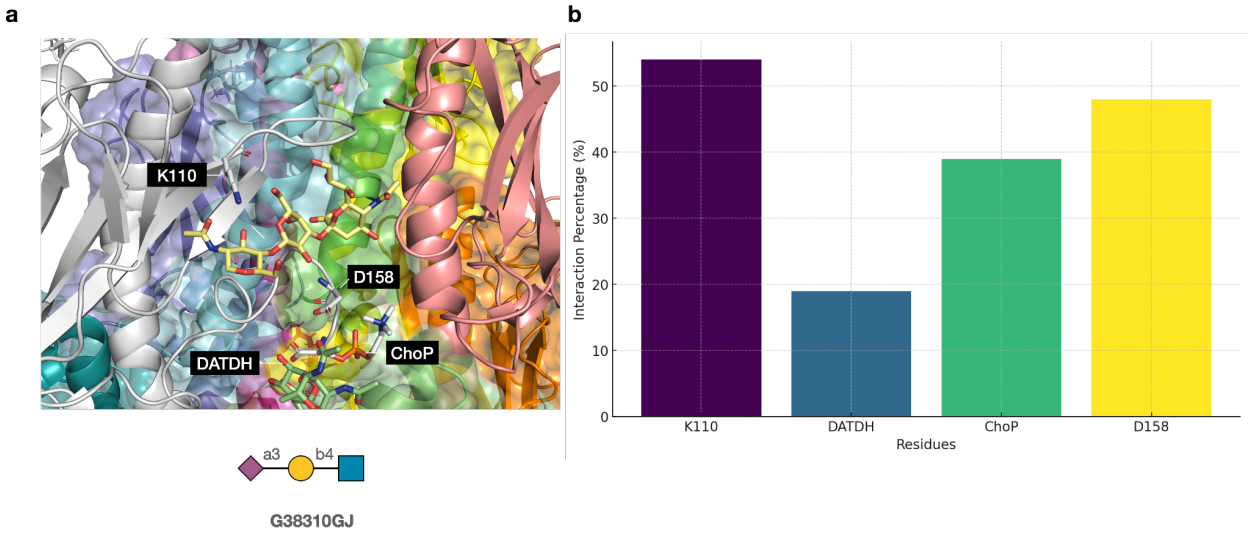


Figure 4.11. Panel a) Binding of the α2,3-sialylated LacNAc(yellow) on the Pilin Array. The key binding residues, namely K110, D158, DATDH(from *O*-linked trisaccharide PTM) and ChoP are labelled. **Panel b)** Barplot of the hydrogen bonding interactions indicating the persistence (%) of each specific interaction calculated along the MD trajectory. The values represent the percentage of time each residue maintained a hydrogen bond with the different antigens. Molecular rendering with Pymol (www.pymol.org)

4.3.8 Binding to Heparan Sulfate

Glycosaminoglycans (GAGs) are abundantly present in the epithelial cell environment where *Nm* builds colonies. Heparan sulfate (HS) is expressed on the surface of nasopharyngeal epithelial cells and has been shown to interact with several *Nm* proteins, including the outer membrane opacity proteins Opc and Opa (de Vries *et al*., 1998), as well as the surface-exposed

lipoprotein, Neisserial Heparin Binding Antigen (NHBA) (Vacca *et al*., 2016). Given HS abundance and established role in host-pathogen interactions, I tested whether HS could be a potential target for the T4P adhesion. The HS sequence I tested, see **Figure 4.12**, was found to bind effectively to the Pilin Array system with a different binding mode compared to the polysialic acids. Due to its larger size and sulfated nature, HS was not confined to the primary binding pocket it was placed initially, but during the MD sampling it extended beyond the binding pocket to interact with pilin subunits in the following helical turns, i.e. located immediately below the binding site. Analogously to the polysialic acid systems, HS was observed to establish stable interactions with the charged lysine residues in the pocket.

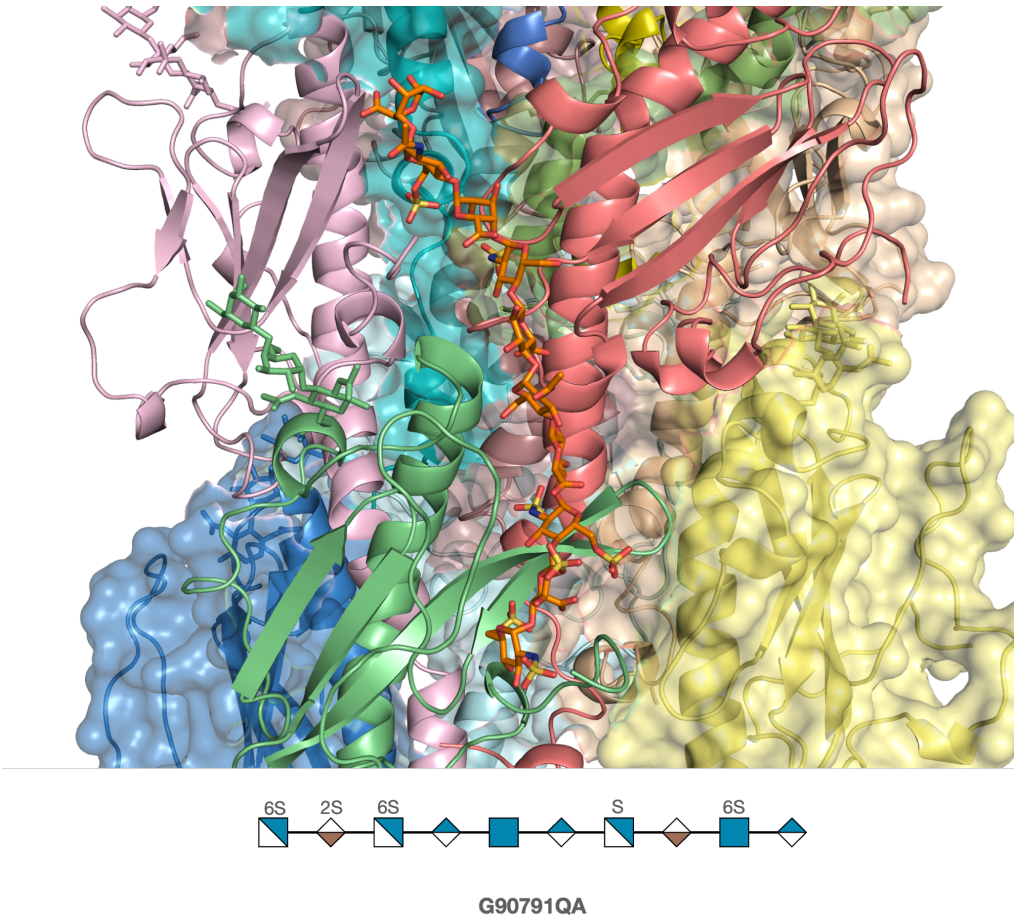


Figure 4.12. Representative bound pose of an heparan sulfate (HS) glycan to the Pilin Array. PilE directly in contact with HS are shown in cartoon representations. Molecular rendering with Pymol (www.pymol.org)

4.3.9 Binding of Lewis B antigen

To gain more insight on the T4P ability to bind fucosylated epitopes, I tested binding to the Lewis B antigen (LeB), which is commonly expressed in the respiratory mucosa (Daniels, 2013). LeB adopted a stable conformation in the binding site, see **Figure 4.13**, through specific interactions with charged residues, analogously to other stable glycans tested. Notably, LeB was found to occupy a deeper position towards the core of the T4P compared to the other glycans tested. This suggests that its stability was not driven only by interactions with the residues in the pocket, but also by steric factors, which may also be determined by a localised weakening or fraying of the T4P helical structure.

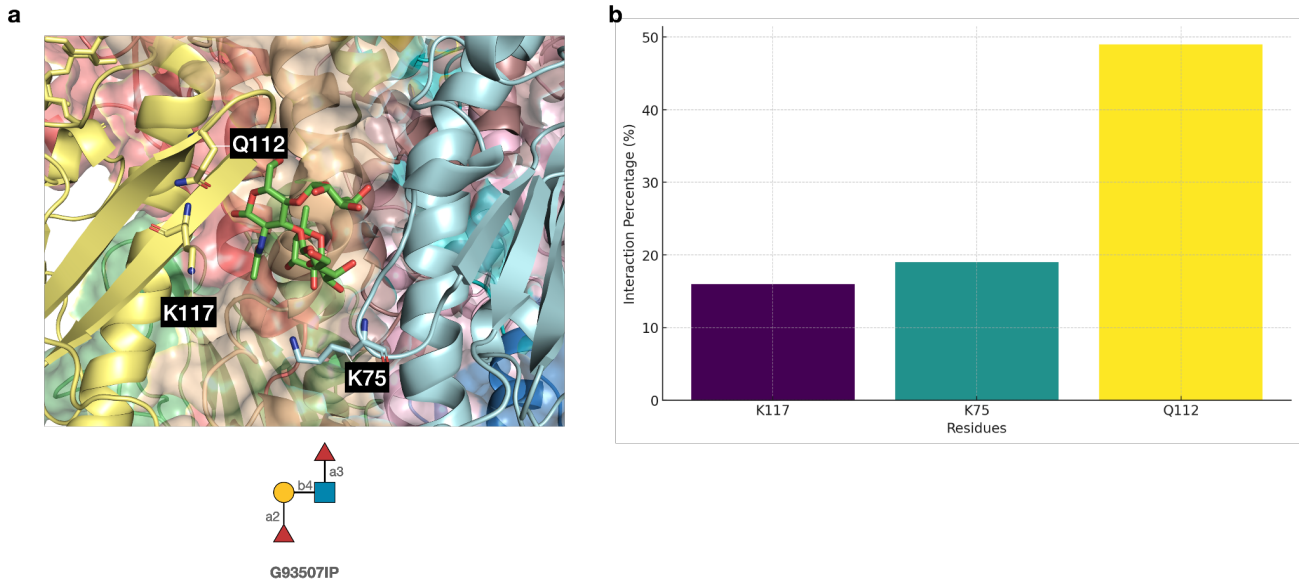


Figure 4.13. Panel a) Bound conformation of the LeB in the Pilin Array. Key binding residues, namely K117, K75, Q112, are labelled. **Panel b)** Barplot indicating the persistence of the hydrogen bonding interactions between LeB and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time. Molecular rendering with Pymol (www.pymol.org) .

4.3.10 Experimental Binding Assays

To further investigate the binding specificity of the T4P, I performed SPR assays in collaboration with Dr Freda Jen and Dr Chris Day from Prof Micheal Jennings'group at the Institute for Biomedicine and Glycomics at Griffith University in Gold Coast, QLD. To assess the impact of bacterial glycosylation on binding affinity, a ΔPglL mutant, which expresses pilin without bacterial glycan due to the knockout of the PglL oligosaccharyltransferase, was also tested. Despite analysing a variety of non-sialylated glycans, including blood group antigens, Lewis antigens, and mannose-containing glycans, no binding was observed at the maximum

concentration of 100 µM. The results of the SPR analysis of the bacterial glycans interacting with sialylated glycans are shown in **Figure 4.14**.

Table 4.1. Calculated binding affinities (K_D in µM) from a minimum of two biological replicates for WT *Nm* pilin and ΔPglL *Nm* pilin where α2-3SL = Neu5Acα2-3Galβ1-4Glc; α2-8α2-3SL=Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;α2-8α2-8α2-3SL= Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;(α2-8Neu5Ac)n~15 = Neu5Aα2-8Neu5Ac with an average of 15 repeats (colominic acid)

Glycan	WT Nm pilin	ΔPglL Nm pilin
α2-3SL	No binding	No binding
α2-8α2-3SL	No binding	No binding
α2-8α2-8α2-3SL	1.6µM±0.02	2.9µM±0.1
$(\alpha 2\text{-}8Neu5Ac)_{n\sim15}$	7.8µM±0.7	12.9µM±0.2

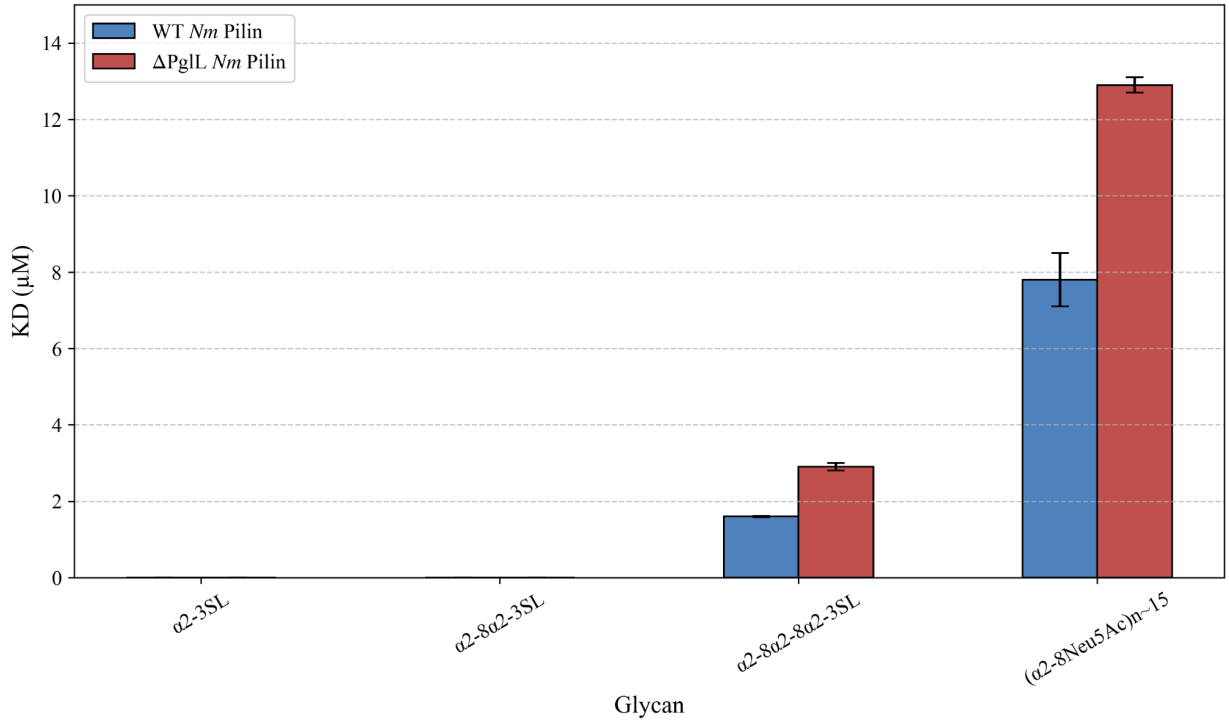


Figure 4.14. panel a) Barplot of the calculated binding affinity (K_D in µM) from a minimum of two biological replicates for WT *Nm* pilin and ΔPglL *Nm* pilin where α2-3SL = Neu5Acα2-3Galβ1-4Glc; α2-8α2-3SL=Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;α2-8α2-8α2-3SL= Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;(α2-8Neu5Ac)n~15 = Neu5Aα2-8Neu5Ac with an average of 15 repeats (colominic acid)

4.4 Discussion:

Using MD simulations, I was able to analyse the structure and dynamics of the *Nm* T4P system populated by pilE subunits. My results revealed how the tightly packed architecture of the pilE system creates potential multi-subunit carbohydrate binding pockets between the globular domains of adjacent subunits. This unique arrangement facilitates high avidity for glycan interactions, as there are four pilin subunits per helical turn, providing the T4P with four potential binding sites per turn. Given the significant length of the T4P, which can span several micrometers(Skerker and Berg, 2001), this structure makes the T4P an effective and dynamic adhesin.

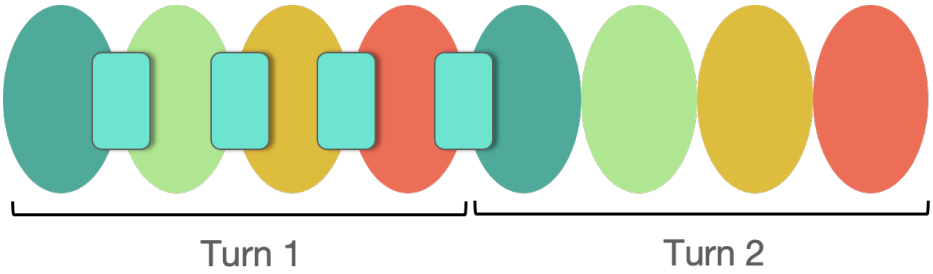


Figure 4.15: Illustration of the binding site availability between subunits per helical turn of pilin system. The ovals represent each subunits and the cyan rectangles represent the binding pocket between subunits

When incorporating pilV subunits into the pilE system, I observed that the shorter globular domain of the pilV subunits hindered binding capabilities due to the large gap left between the pilV and adjacent pilE subunits on one side. The collapse of the structure upon the addition of bound glycans may have been a result of this gap, as well as the number of pilV subunits introduced. While I added multiple pilV subunits to ensure reproducibility in our array, our results suggest that pilV does not integrate well into the pilE system under the current configuration. This highlights the need for careful consideration when incorporating pilV into the framework of the T4P system, as its structural differences can disrupt stability and function.

Results from my pilin array demonstrated that pilE was particularly effective at binding sialic acids, with polysialic acids showing especially stable interactions. Across multiple replicas, the α2-8 linked polysialic acids consistently established effective binding, with the same residues recurring in our binding affinity analysis, as shown in **Figure 4.10**. The binding site, located

between the globular domains of two adjacent subunits, was rich in polar residues, and the PTMs on the pilin subunits further enhanced binding stability. The key residues identified in the binding site, namely K73, K75, K110, Q112, N113, K117, are mostly conserved in pile proteins across various strains of Serogroup B and are also present in similar pilE structures from other serogroups, including Serogroups A and C, see **Figures 4.16 and 4.17.** To validate the importance of these residues in glycan binding, our collaborators, Dr. Freda Jen and Dr. Chris Day, are currently conducting mutagenesis experiments. Specifically, they are mutating K73, K75, and K117 to Ala (A73, A75, and A117, respectively) to assess how the removal of these polar residues will affect binding.

Figure 4.16: Sequence alignment of the pilE protein from different strains of Serogroup B with the key residues involved in the binding of polysialic acid highlighted with a red box and labelled. A legend is provided to match each sequence ID with the corresponding strain. Sequence alignment was done using Clustal Omega(Sievers *et al.*, 2011)

Figure 4.17: Sequence alignment of the pilE protein from different serogroups of *Neisseria meningitidis(Nm)* with the key residues involved in the binding of polysialic acid highlighted with a red box and labelled. A legend is provided to match each sequence ID from which the corresponding serogroups are sourced. Sequence alignment was done using Clustal Omega(Sievers *et al.*, 2011)

SPR analysis of the *Nm* pilin revealed a preference for binding sialylated glycans over non-sialylated glycans, with a minimum motif of three Neu5Ac residues. This finding is particularly

notable as the capsular polysaccharide of serogroup B consists of repeating α2,8-linked Neu5Ac, suggesting that T4P may interact with the capsular polysaccharide aiding in colonisation. Interestingly, the SPR results also indicated that when bacterial glycosylation was knocked out (ΔPglL mutant), binding to sialylated glycans increased. Our computational results indicated that the bacterial trisaccharide was able to contribute to the binding of the sialic acid. However, the removal of the trisaccharide may have improved binding site accessibility, allowing for more effective interactions, which could compensate for the loss of direct contributions from the trisaccharide in the binding process.

4.5 Conclusion

In this work, I analysed the structure and dynamics of an *Nm* T4P system composed of the major pilE subunits using MD simulations. My analysis revealed a dynamic and tightly packed system whereby the globular domain of adjacent subunits interact with each other, while the helical regions form a stable core. I observed how important this compact arrangement is, whereby substituting smaller pilV subunits led to gaps within the system and eventually the collapse of the T4P structure. Using a combination of docking and MD protocols, I identified a potential binding site for glycan epitopes between the globular domains of adjacent major pilE subunits with the T4P system. Using a combination of MD simulations coupled with SPR analysis to validate my results, I analysed the binding specificity of the *Nm* pilE system for glycan epitopes. The results of the pilin array and SPR analysis, showed that the *Nm* T4P had a strong preference for sialylated glycans, particularly α2,8-linked polysialic acids with the binding being facilitated by polar residues lining the binding pocket which are conserved across different strains and serogroups of the *Nm* pilE protein. The MD results also showed how the post-translational modifications can contribute to this binding while the SPR results found that the deletion of the *O*-linked trisaccharide can enhance binding.

References:

1. Barnier, J.-P. *et al*., 2021. The minor pilin PilV provides a conserved adhesion site throughout the antigenically variable meningococcal type IV pilus. Proceedings of the National Academy of Sciences of the United States of America, 118(45).
2. Bennett, J.S. *et al*., 2012. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria. Microbiology, 158(Pt 6), pp.1570–1580.
3. Bernard, S.C. *et al*., 2014. Pathogenic Neisseria meningitidis utilizes CD147 for vascular colonization. Nature medicine, 20(7), pp.725–731.
4. Carbonnelle, E. *et al*., 2005. Type IV pilus biogenesis in Neisseria meningitidis: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function: Type IV pilus biogenesis in pathogenic Neisseria. Molecular microbiology, 55(1), pp.54–64.
5. Cehovin, A. *et al*., 2010. Sequence conservation of pilus subunits in Neisseria meningitidis. Vaccine, 28(30), pp.4817–4826.
6. Coureuil, M. *et al*., 2009. Meningococcal type IV pili recruit the polarity complex to cross the brain endothelium. Science, 325(5936), pp.83–87.
7. Craig, L., Forest, K.T. & Maier, B., 2019. Type IV pili: dynamics, biophysics and functional consequences. Nature reviews. Microbiology, 17(7), pp.429–440.
8. Davidsen, T. & Tønjum, T., 2006. Meningococcal genome dynamics. Nature reviews. Microbiology, 4(1), pp.11–22.
9. Denis, K. *et al*., 2019. Targeting Type IV pili as an antivirulence strategy against invasive meningococcal disease. Nature microbiology, 4(6), pp.972–984.

10. Dominguez, C., Boelens, R. & Bonvin, A.M.J.J., 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society, 125(7), pp.1731–1737.
11. Exley, R.M. *et al*., 2009. Identification of meningococcal genes necessary for colonization of human upper airway tissue. Infection and immunity, 77(1), pp.45–51.
12. Finne, J., Leinonen, M. & Mäkelä, P.H., 1983. Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis. Lancet, 2(8346), pp.355–357.
13. Gasparini, R. *et al*., 2015. How the Knowledge of Interactions between Meningococcus and the Human Immune System Has Been Used to Prepare Effective Neisseria meningitidis Vaccines. Journal of immunology research, 2015, p.189153.
14. Gault, J. *et al*., 2015. Neisseria meningitidis Type IV Pili Composed of Sequence Invariable Pilins Are Masked by Multisite Glycosylation. PLoS pathogens, 11(9), p.e1005162.
15. Helena De Oliveira, L. *et al*., 2017. Impact and effectiveness of meningococcal vaccines: a review. Revista panamericana de salud publica [Pan American journal of public health], 41, p.e158.
16. Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. Journal of molecular graphics, 14(1), pp.33–8, 27–8.
17. Ives, C.M. *et al*., 2024. Restoring protein glycosylation with GlycoShape. Nature Methods, pp.1–11.
18. Jafri, R.Z. *et al*., 2013. Global epidemiology of invasive meningococcal disease. Population health metrics, 11(1), p.17.
19. Jarvis, G.A. & Vedros, N.A., 1987. Sialic acid of group B Neisseria meningitidis regulates alternative complement pathway activation. Infection and immunity, 55(1), pp.174–180.

20. Jen, F.E.-C. *et al*., 2021. Complete Genome Sequence of Serogroup B Neisseria meningitidis Strain C311. Microbiology resource announcements, 10(42), p.e0078821.
21. Jen, F.E.-C. *et al*., 2013. Dual pili post-translational modifications synergize to mediate meningococcal adherence to platelet activating factor receptor on human airway cells. PLoS pathogens, 9(5), p.e1003377.
22. Jorgensen, W.L. *et al*., 1983. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics, 79(2), pp.926–935.
23. Jumper, J. *et al*., 2021. Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), pp.583–589.
24. Kahler, C.M. *et al*., 1998. The (α2→8)-linked polysialic acid capsule and lipooligosaccharide structure both contribute to the ability of serogroup B Neisseria meningitidis to resist the bactericidal activity of normal human serum. Infection and immunity, 66(12), pp.5939–5947.
25. Kirschner, K.N. *et al*., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. Journal of computational chemistry, 29(4), pp.622–655.
26. Kolappan, S. *et al*., 2016. Structure of the Neisseria meningitidis Type IV pilus. Nature communications, 7, p.13015.
27. Lee, T.-S. *et al*., 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. Journal of chemical information and modeling, 58(10), pp.2043–2050.
28. Le Guennec, L. *et al*., 2020. Receptor recognition by meningococcal type IV pili relies on a specific complex -glycan. Proceedings of the National Academy of Sciences of the United States of America, 117(5), pp.2606–2612.

29. Maier, J.A. *et al*., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of chemical theory and computation, 11(8), pp.3696–3713.
30. McNamara, L.A. *et al*., 2020. Detection of Ciprofloxacin-Resistant, β-Lactamase-Producing Neisseria meningitidis Serogroup Y Isolates - United States, 2019-2020. MMWR. Morbidity and mortality weekly report, 69(24), pp.735–739.
31. Mubaiwa, T.D., Hartley-Tassell, L.E., *et al*., 2017. The glycointeractome of serogroup B Neisseria meningitidis strain MC58. Scientific reports, 7(1), p.5693.
32. Mubaiwa, T.D., Semchenko, E.A., *et al*., 2017. The sweet side of the pathogenic Neisseria: the role of glycan interactions in colonisation and disease. Pathogens and disease, 75(5). Available at: http://dx.doi.org/10.1093/femspd/ftx063.
33. Roux, N., Spagnolo, J. & de Bentzmann, S., 2012. Neglected but amazingly diverse type IVb pili. Research in microbiology, 163(9-10), pp.659–673.
34. Schneider, M.C. *et al*., 2007. Interactions between Neisseria meningitidis and the complement system. Trends in microbiology, 15(5), pp.233–240.
35. Sievers, F. *et al*., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7, p.539.
36. Skerker, J.M. & Berg, H.C., 2001. Direct observation of extension and retraction of type IV pili. Proceedings of the National Academy of Sciences of the United States of America, 98(12), pp.6901–6904.
37. Tiemeyer, M. *et al*., 2017. GlyTouCan: an accessible glycan structure repository. Glycobiology, 27(10), pp.915–919.
38. Wang, J. *et al*., 2004. Development and testing of a general amber force field. Journal of computational chemistry, 25(9), pp.1157–1174.

39. Willerton, L. *et al*., 2021. Increase in penicillin-resistant invasive meningococcal serogroup W ST-11 complex isolates in England. Vaccine, 39(19), pp.2719–2729.
40. Wimmerova, M. *et al*., 2003. Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin. The Journal of biological chemistry, 278(29), pp.27059–27067.

Chapter 5. Structure, dynamics, and glycan-binding specificities of the *N. gonorhoeae* Type IV Pilus (T4P)

5.1 Introduction

Neisseria gonorrhoeae (*Ng*) is another pathogenic member of the *Neisseria* genus, alongside *Neisseria meningitidis* (*Nm*), discussed in **Chapter 4**. *Ng* is the causative agent of the sexually transmitted infection (STI) gonorrhoea, which identified by the the World Health Organization (WHO) as a significant public health concern. *Ng* is one of the most common causes of STIs globally and in 2020 it was reported to affect an estimated 82.4 million adults worldwide (www.who.int/news-room/fact-sheets/detail/gonorrhoea-(neisseria-gonorrhoeae-infection)). Like *Nm*, humans are the exclusive natural hosts of this pathogen, which forms colonies mainly on the urogenital mucosa, but it has been also detected on the ocular, nasopharyngeal and rectal mucosa (Edwards and Apicella, 2004; Quillin and Seifert, 2018). Gonorrhea commonly presents itself as urethritis in men and cervicitis in women. It can be asymptomatic and hard to diagnose particularly in women, where asymptomatic cases of gonorrhoea have been reported upwards of 40% (Unemo and Shafer, 2014; Edwards *et al.*, 2016; Martín-Sánchez *et al.*, 2020). Symptoms of *Ng* infection can be non-specific and prone to misdiagnosis(Quillin and Seifert, 2018). Left untreated, gonorrhoea can lead to pelvic inflammatory disease, ectopic pregnancy in women(Quillin and Seifert, 2018; Unemo *et al.*, 2019) as well as infertility in both sexes. Most cases of gonorrhea can be currently treated with widely available antimicrobial drugs namely ceftriaxone, penicillin, sulphonamides and tetracycline. However, resistance to these antimicrobials is rising globally(Alirol *et al.*, 2017; Wi *et al.*, 2017) calling for new therapies as well as the development of an effective vaccine(*Multi-drug resistant gonorrhoea*, 2024).

Similar to *Nm* strains, Type IV pili (T4P) are the major adhesins also in *Ng*, playing an important role in the mobility of the bacterial cell and crucial to colonisation(Punsalang and Sawyer, 1973; Craig, Forest and Maier, 2019). The *Ng* T4P shares a similar architecture with the pilus of *Nm(Craig, Forest and Maier, 2019)*, see **Figure 4.1**. The major pilin subunit in the *Ng* T4P is pilE and its amino acid sequence is highly similar (~70%) to that of the *Nm* pilE, depending on the strain(Wang *et al.*, 2017).

One major difference between *Ng* and *Nm* strains is the absence of the capsular polysaccharide in *Ng.* The presence of the capsular polysaccharide in *Nm* provides an extra layer of protection from the immune cells(Virji, 2009). Four of the *Nm* serogroups (B,C,W,Y) also express polysialic acid as part of their capsular polysaccharide, which we discovered are able to bind effectively to the *Nm* T4P, as discussed in **Chapter 4**. *Ng* strains still express a wide range of surface-exposed carbohydrate structures, many of which are shared with *Nm* strains, including its lipo-oligosaccharides (LOS) and glycan structures incorporated into glycoproteins and glycolipids(Mubaiwa *et al.*, 2017; Semchenko *et al.*, 2019). Although *Ng* is unable to synthesise sialic acids, it can scavenge sialic acid from host cells and add to its own LOS structures(Cardenas *et al.*, 2024). It has also been reported to interact with a wide range of carbohydrate structures expressed on the mucosal surface(Semchenko *et al.*, 2019). In particular, *Ng* can express mannose-binding lectins that can bind effectively with mannosyl groups present on the urogenital epithelial cells(Semchenko *et al.*, 2019).

Ng T4Ps role in cell surface adherence has been previously highlighted, with evidence of binding interactions with cell surface receptors such as the membrane cofactor protein CD46 (Källström *et al.*, 2001), the complement receptor 3 (CR3)(Jennings *et al.*, 2011) and the C4 binding protein (C4BP) (Blom *et al.*, 2001). However, there is no clear evidence about the *Ng* T4P binding capabilities with cell surface glycans and about its role in adherence, colonisation and infection.

In this work, I used molecular dynamics (MD) simulations to explore the molecular architecture and stability of a *Ng* T4P atomistic model built as a polymer of the major pilin subunit, pilE, using a similar protocol and analysis as seen for the *Nm* T4P in **Chapter 4**. This approach allowed me to make a direct comparison between *Nm* and *Ng* T4P structures. Analogously to *Nm* T4P, I investigated the glycan binding specificity of the *Ng* T4P using the same MD-based glycan screening approach, and compared the results with the data from the *Nm* pilin array. Ultimately, my results from the computational screening were validated against surface plasmon resonance (SPR) results by Dr Freda Jen and Dr Chris Day in Prof Micheal Jennings' group at the Institute for Biomedicine and Glycomics at Griffith University in Gold Coast, QLD, Australia.

5.2 Methods

5.2.1 Computational Method

I generated the isolated *Ng* pilE subunit structure using AlphaFold2 (Jumper *et al*., 2021) based on genomic data from a variant of strain MS11 (GenBank accession number EEZ48905). I created the starting structures for the bacterial glycans and all glycan epitopes using the carbohydrate builder tool on GLYCAM-WEB (https://glycam.org/). I then performed MD simulations for each glycan structure and analysed the resulting trajectories, which I uploaded to the GlycoShape Glycan Structure Database (Ives *et al*., 2024). From this analysis, I identified the most populated structures for each glycan.

I ran a minimum of three uncorrelated replicas for each glycan. For glycan structures that theoretically correspond to two energetically accessible conformers, I ran two uncorrelated replicas for each conformer. For structures that theoretically corresponded to three or more energetically accessible conformers, I ran one replica for each conformer. I generated the parameters for phosphoryl-choline (ChoP) and phosphoethanolamine (PE) using the AMBER18 software package (Lee et al., 2018). I linked the *O*-linked bacterial disaccharide, Gal-α1-3-DATDH, ChoP, and PE to Ser 63, Ser 68, and Ser 156 using the *tleap* tool from AMBER18 (Lee et al., 2018).

I arranged the *Ng* pilE subunits to form a 27 pilE system using the cryo-EM structure PDB ID 5KUA (Kolappan et al., 2016) as a template, aligning the backbone of the alpha-helix. Since the 5KUA structure consisted of only 26 subunits, I added the 27th *Ng* pilE subunit by shifting the 5KUA structure down by one subunit. Finally, I analysed the stability and dynamics of this model using MD simulations.

To populate the *Ng* pilin array, I placed the glycan epitopes using two methods. For glycans epitopes not used in the *Nm* pilin array, see **Figure 5.1,** the terminal substituent of each glycan epitope was superimposed onto the terminal sialic acids of both α2,6- and α2,3-linked SialylLacNAc bound to the *Nm* pilE dimer, described in **Chapter 4**.

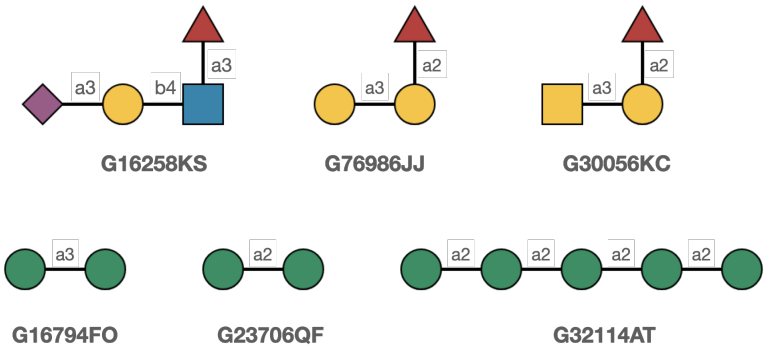


Figure 5.1. Glycan epitopes exclusive to the 27-mer *Ng* pilin array. Each glycan is labelled with its GlyTouCan(Tiemeyer *et al.*, 2017) identifier (https://glytoucan.org/).

I then superimposed this construct onto pilin subunits of the 27-mer *Ng* pilE system using the backbone of the alpha helix as the alignment basis. The *Nm* pilE dimer was then removed, leaving the glycan epitope positioned between two adjacent *Ng* pilin subunits. For glycan epitopes previously used to populate the *Nm* pilin array, their initial placement was achieved by superimposing the pilin subunits which held the glycan epitopes from the *Nm* pilin array onto the *Ng* pilE system. The stability of the bound ligands was determined by MD simulations.

All MDs were run in water with 200 mM NaCl with the AMBER18 simulation package(Lee *et al.*, 2018). Protein atoms and counterions were represented by the AMBER ff14SB(Maier *et al.*, 2015) parameter set, while the carbohydrate atoms were represented using the GLYCAM06j-1version of the GLYCAM06 force field(Kirschner *et al.*, 2008). ChoP and PE were represented using GAFF2(Wang *et al.*, 2004). Water molecules were represented using the TIP3P water model(Jorgensen *et al.*, 1983).

All MD simulations were run following a similar protocol, starting with an energy minimisation stage through 500,000 steps of steepest descent. During minimisation only hydrogen atoms, water molecules, and counterions were left unrestrained, while the position of all other atoms was kept restrained with a force constant of 5 kcal/mol·Å^2. Following minimisation, the system was brought up to standard temperature through a heating phase in the NVT ensemble using a Langevin thermostat. This heating phase was completed in two stages of 500 ps each. During phase one the temperature was raised from 0 to 100 K where the volume was kept constant and from 100 to 300 K during phase two where the pressure was kept constant. The system was then equilibrated for 1 ns in the NPT ensemble to a 1 am pressure

with the Berendsen barostat. During these equilibration stages, we used the same position restraints as for minimisation.

For the free glycan structures, we performed a conformational equilibration for 10 ns with all positional restraints removed. This was followed by a production phase of 500 ns for each replica of the glycan structure with all positional restraints removed.

For the full 27-subunit pilE system both with and without the glycan epitopes, a stepwise equilibration process was employed. Restraints were initially put on the glycan epitopes and pilin subunits 1-7 and 20-27 which were then gradually relaxed to the glycan epitopes and subunits 1-3 and 25-27 and subsequently narrowed down to the the glycan epitopes and alpha helices of subunits 1-3 and 25-27. Production phases followed with all restraints on the heavy atoms removed except for those for the alpha helices of the pilin subunits 1-3 and 25-27.

5.2.2 Surface Plasmon Resonance

As part of this project in February 2024 I visited Prof Micheal Jennings' research group at the Institute of Biomedicine and Glycomics at Griffith University in Gold Coast, QLD, Australia as part of my PhD international placement. During this internship I worked with Dr Freda Jen and Dr Chris Day who ran SPR experiments designed to test the results I obtained from the pilin array screening by MD simulations. The experiments were run using the method described below.

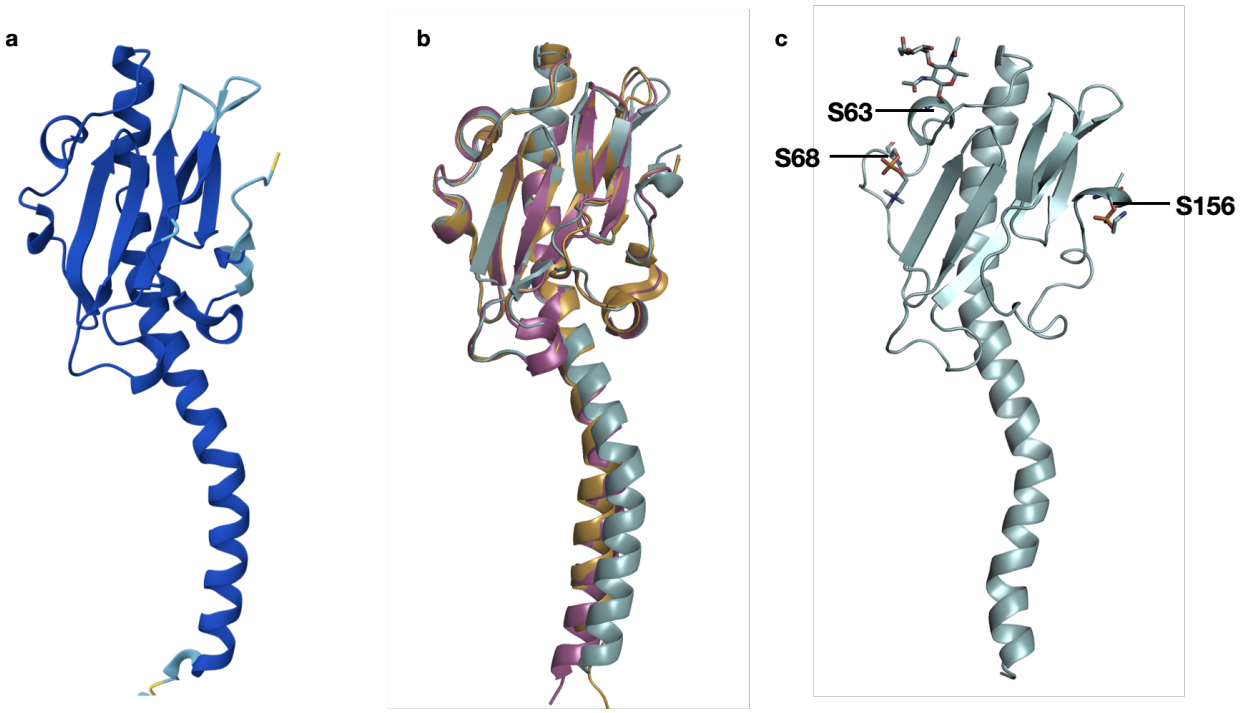
Vortex-released pilin from *Ng* with differing glycosylation modifications were immobilised on a CDH chip (Sartorius) and run using the Pioneer FE SPR system (Sartorius). Immobilisation was carried out at pH 4.5 and the protein flowed at 10 µL per minute for 5 minutes over an EDC/NHS activated surface. The surface was quenched with ethanolamine at 10 µL per minute for 7 minutes. A pilin deletion mutant bacteria vortexed and treated the same as the piliated bacteria was used as the negative surface control for double reference subtraction. Glycans were flowed over immobilised protein at 50 µL per minute at a maximum concentration of 100 µM using the OneStep kinetics with regeneration. Data from each run was exported to QDat analysis software (Sartorius) and analysed using the standard K_A/K_D model. The calculated binding affinity (K_D) are from a minimum of two biological replicates.

5.3 Results

In this section, I present the results of the MD simulation of a 27-mer pilE *Ng* T4P system, which from this point forward I will refer to as the *Ng* pilE system. I will then present results from our MD-based glycan binding screening using the *Ng* pilE system, which I will refer to as the *Ng* pilin array from this point forward. Finally, I will present the results of the SPR experiment of the *Ng* pilin systems tested.

5.3.1 Structure and Dynamics of the *Ng* pilE System

The *Ng* pilE system is based on a variant from strain MS11 which was shown to exhibit high-antimicrobial resistance(Dillard, 2011). The post-translational modifications (PTM) of this strain have been well-characterised (Jennings *et al.*, 2011). Several variants of this strain's pilE are available including some with crystal structures such as the structures corresponding to PDBid 2HIL and 1AY2; however, most of these structures have slight variations in their sequences at variable regions including the semi-variable region (residues 51-114), the hypervariable loop (residues 127-141) and the hypervariable tail (residues 153-159)(Obergfell and Seifert, 2016). To explore how differences in the sequences would affect the protein structure, I decided to use a variant of MS11(GenBank accession number EEZ48905) that was not structurally characterised by crystallography and generated a 3D structure with AlphaFold2 (Jumper *et al.*, 2021). We compared the sequence of the variant generated by AlphaFold (AF) to the sequences of two variants with readily available crystal structures, 2HIL and 1AY2 using Clustal Omega(Madeira *et al.*, 2024). The sequence analysis showed high similarity between the sequence of AF-predicted model and the two variants, with percentage identity values of 94.9% for both, see **Figure 5.2 d-e**. We then compared their 3D structures by structural alignment of the backbone atoms with Pymol (www.pymol.org). The AF predicted structure scored RMSD values of 0.624 Å and 0.708 Å when compared to 2HIL and to 1AY2, respectively.



d

```
EEZ48905.2      FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGKWPEN    60
2HIL            FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGKWPEN    60
1AY2            FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGKWPEN    60
                ************************************************************

EEZ48905.2      NTSAGVASPPTDIKGKYVKEVEVKNGVVTATMLSSGVNNEIKGKKLSLWGRRENGSVKWF    120
2HIL            NTSAGVASSPTDIKGKYVKEVEVKNGVVTATMLSSGVNNEIKGKKLSLWARRENGSVKWF    120
1AY2            NTSAGVASPPSDIKGKYVKEVEVKNGVVTATMLSSGVNNEIKGKKLSLWARRENGSVKWF    120
                ******** *:*************************************.**********

EEZ48905.2      CGQPVTRADDDTVADAKDGKEIDTKHLPSTCRDKASDAK 159
2HIL            CGQPVTRTDDDTVADAKDGKEIDTKHLPSTCRDNFDAK- 158
1AY2            CGQPVTRTDDDTVADAKDGKEIDTKHLPSTCRDNFDAK- 158
                *******:************************: .
```

e

	EEZ48905.2	2HIL	1AY2
EEZ48905.2	100	94.94	94.94
2HIL	94.94	100	98.73
1AY2	94.94	98.73	100

Figure 5.2. Panel a) Structure of *Ng* pilE subunit from strain MS11(GenBank accession number EEZ48905.2) generated using AlphaFold. The structure is coloured based on the per-residue model confidence score(pLDDT) where dark blue represents very high confidence(pLDDT > 90), cyan represents high confidence(90 > pLDDT > 70) and yellow represents low confidence(70 > pLDDT > 50). **Panel b)** Structural alignment of the AlphaFold model(cyan) with PDB structures, 2HIL(orange) and 1AY2(pink) **Panel c)** Single pilE subunit with the positions for the PTMs, S63, S68 and S156, labeled. **Panel d)** Sequence alignment of the pilE AlphaFold sequence (EEZ48905.2), against the sequences of the PDB structures of 2HIL and 1AY2. Sequence alignment was conducted using Clustal Omega(https://www.ebi.ac.uk/jdispatcher/msa/clustalo)(Sievers *et al.*, 2011). **Panel e)** Percentage identity scores for the sequence alignment. Molecular rendering was done with pymol (www.pymol.org).

Using the 'bond' functionality in *tleap*((Lee *et al.*, 2018), we added ChoP and PE to Ser 68 and Ser 156, respectively(Jennings *et al.*, 2011) and the *O*-linked disaccharide Galα1-3DATDH at Ser 63. This latter modification differs slightly from the *O*-linked glycan reported in the *Nm* pilE subunits, which in our model carried the trisaccharide Galβ1-4Galα1-3DATDH(Jen *et al.*, 2013). This difference in glycosylation is due to the fact that the gene responsible for adding the second galactose, pglE, is inactive in most *Ng* strains(Jennings *et al*., 2011; Power *et al*., 2003).

I used the AF pilE predicted structure to build the 27-mer *Ng* pilE system using the 5KUA structure as template, as for the *Nm* pilE system discussed in **Chapter 4**. The system was extended to 27 subunits to prevent the loss of the glycan ligand located between the 25th and 26th pilE subunits, which occurred in the *Nm* pilE system. While this doesn't guarantee that the glycan epitope will remain in the array, this adjustment helps to ensure that the absence of subunits is not the cause of its displacement.

The *Ng* pilE system shares a similar architecture with the 26-mer *Nm* pilE system, where the alpha helices form the core of the system and the globular domains of adjacent pilE subunits face one another. Results from the MD simulations of the 27-mer *Ng* pilE system showed similar trends to the ones observed for the *Nm* pilE system. This includes the greater degree of flexibility exhibited by the globular domains in comparison to the alpha helices as seen from the RMSF plot, see **Figure 5.3c**. The average rise and twist values observed for the *Ng* pilE system were 9.7 Å and 100.6°, respectively, see **Figure 5.3b,** were also very similar to the average rise and tilt values seen for the *Nm* pilE system, which were 9.7 Å and 101.1°.

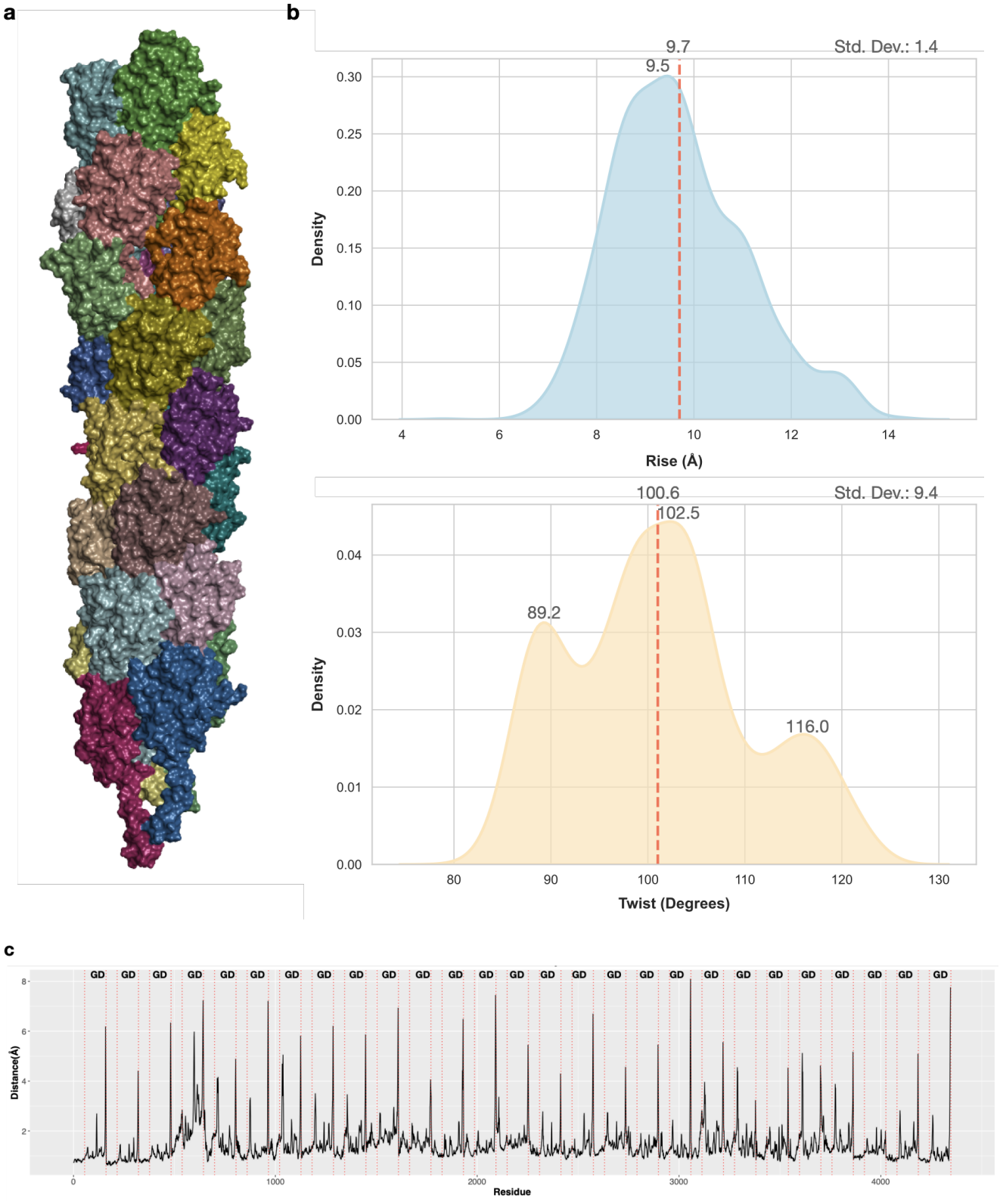


Figure 5.3. Panel a) Graphical representation of the reconstructed *Ng* pilE T4P structure. Each pilE subunit is represented by a different colour. **Panel b)** Plot of the rise and twist parameters used to define the structure of the *Ng* pilE system. The average values for each parameter are labelled and indicated with red dotted lines. The major peaks for each parameter are also labelled. The standard deviation(Std. Dev.) for the rise and twist parameters are listed at the top of the plots. **Panel c)** Line plot of the RMSF values calculated from the MD trajectories of the backbone atoms of the *Ng* pilE system. Red dotted lines and labels

identify the residues corresponding to the globular domains(GD) of the pilE subunits. Molecular rendering was done with pymol (www.pymol.org).

5.3.2 'Pilin Array' Screening by MD Simulation

Previously, I used a set of 13 glycan epitopes to populate the 26-mer *Nm* pilin array in **Chapter 4**. Although there were 26 binding pockets available in the 26-mer construct, I wanted to evaluate whether the *Nm* pilin array could effectively accommodate multiple glycan epitopes simultaneously. To test this, I used half the quantity of binding pockets available. Given that the 26-mer *Nm* pilin array was able to maintain its stability while interacting with multiple glycan epitopes, in our *Ng* pilin array I extended occupancy to all 26 binding pockets. I populated the 27-mer *Ng* with the glycan epitopes listed in **Figure 5.4** with two copies of each glycan. The alignment of the glycan epitopes was performed using a method similar to that described in **Chapter 4** and discussed in detail in the Methods section.

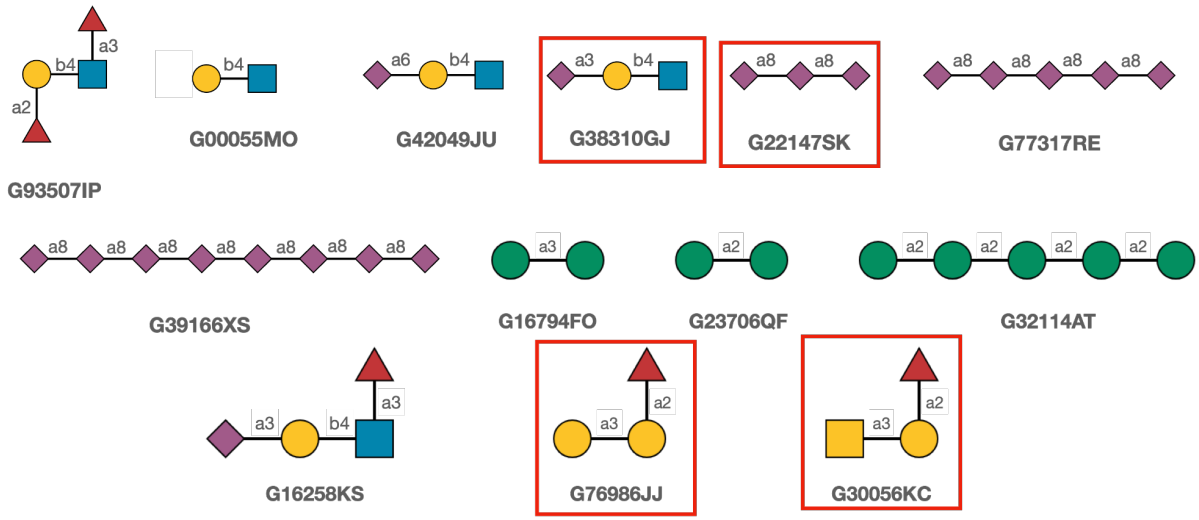


Figure 5.4. Glycan epitopes used to populate the 27-mer *Ng* pilE subunit system. Each glycan is labelled with its GlyTouCan(Tiemeyer *et al.*, 2017) identifier (https://glytoucan.org/). The glycans epitopes analysed from the pilin array are highlighted within a red boundary.

During the MD simulation of our *Ng* pilin array, the pilin subunits at the top of our pilin system drifted away from the core triggering a partial disruption of the top half of our T4P model, see **Figure 5.5a-b**. The displacement of pilin subunits resulted in the deviation from the dynamics observed from the pilE system without the glycan epitopes and characterised by the presence of rise and tilt peak values of 1.3 Å, 4.3 Å, and 15.3 Å, and 78.8°, 135°, respectively, values that are not present in the pilE system without glycans see **Figure 5.4.c.** As a consequence of

this structural collapse, I was unable to assess the binding specificity of the glycan epitopes positioned on the structure that was compromised, which were displaced. However, I am still able to present the results of glycan epitopes unaffected by this shift and located in the bottom half of the model, see **Figure 5.4.**

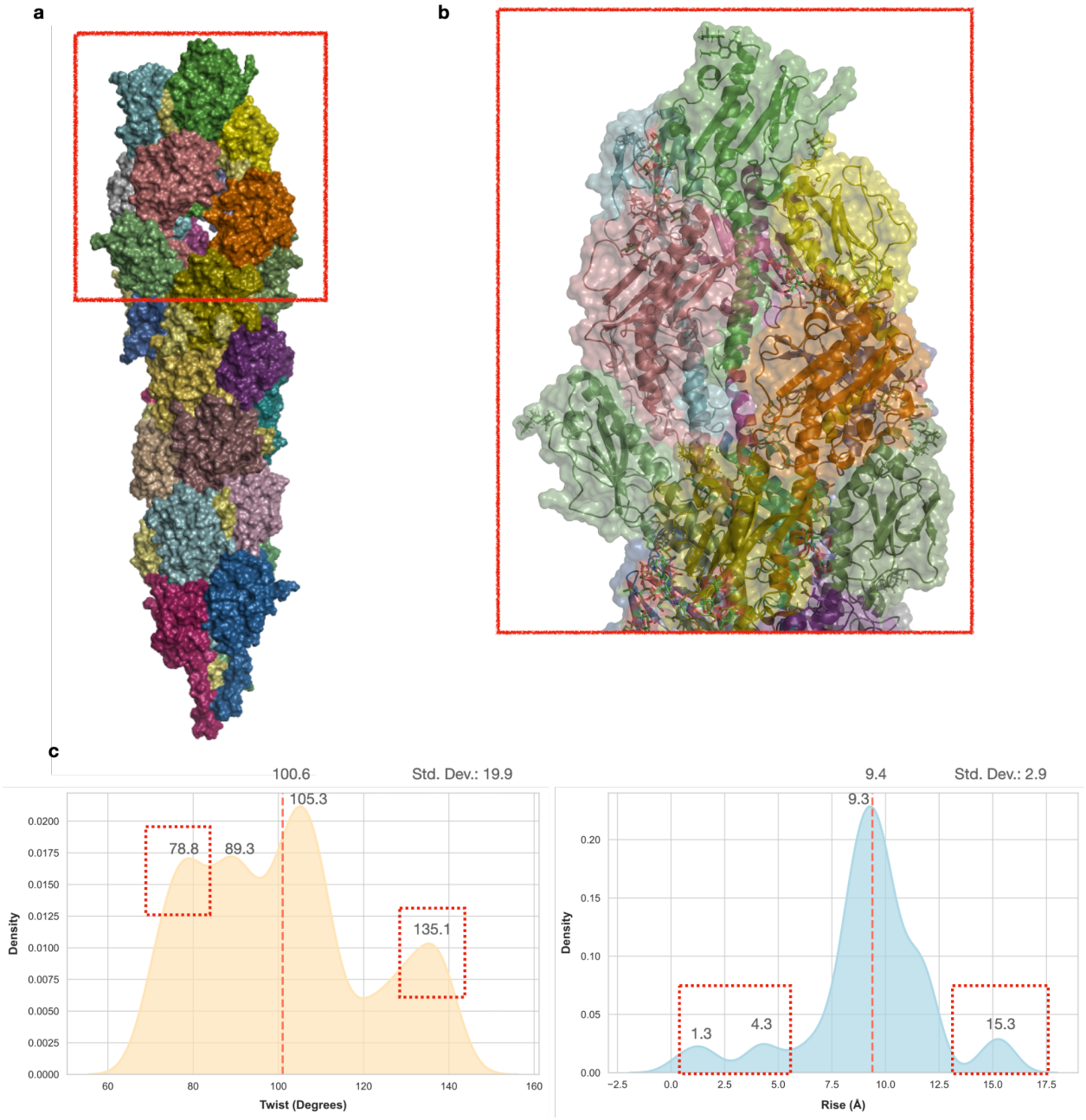


Figure 5.5 Panel a) Graphical representation of the *Ng* pilE array taken from the MD simulation. Highlighted is the gap created by the drift of the *Ng* pilE subunits at the top-half of the system. **Panel b)** Close-up of the top half of the pilE system showcasing the separation of the pilin subunits from the core resulting in the gap in the system. The pilin subunits are rendered as cartoons with transparent surfaces. **Panel c)** Density plot of the rise and twist parameters used to define the structure of the *Ng* pilin array. The average values for each parameter are labelled and indicated with red dotted lines. The peaks for each parameter are also labelled. The peaks corresponding to displaced pilin subunits are highlighted with a red box. The standard

deviation (Std. Dev.) for the rise and twist parameters are listed at the top of the plots. Molecular rendering was done with pymol (www.pymol.org).

5.3.3 Binding of the α2-8 linked polysialic acid

The trisaccharide of α2-8 linked polysialic acid located near the top of the *Ng* pilin array was retained within its binding pocket throughout the 1μs MD simulation of the *Ng* pilin array, forming persistent hydrogen bond interactions with the polar residues lining the binding site, see **Figure 5.6.** As shown for the *Nm* pilin array described in **Chapter 4,** this trisaccharide adopted a stable "hook" conformation, see **Figure 4.4.**

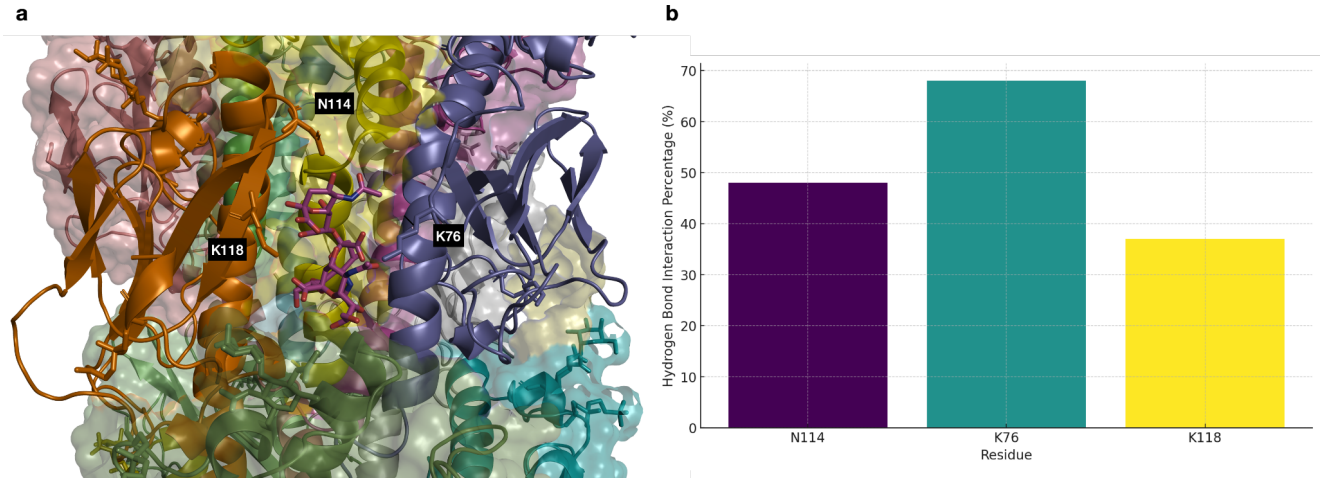


Figure 5.6 Panel a) Bound α2-8 linked polysialic acid trisaccharide(rendered as purple sticks) taken from the MD simulation of the *Ng* pilin array. Key binding residues, namely K118, K76, N114, are labelled. **Panel b)** Barplot indicating the persistence of the hydrogen bonding interactions between the α2-8 linked polysialic acid trisaccharide and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time.

5.3.4 sialyl-α(2-3)-LacNAc

Both copies of the sialyl-α(2-3)-LacNAc were positioned near the bottom half of the *Ng* pilin array and remained in their respective binding pocket throughout the MD trajectory. Both ligands adopted different binding poses likely because of differences in their original placement within the binding sites, see **Figure 5.7.** Both copies were stabilised by interactions with polar residues lining the binding pocket, however, while one copy interacted with residues from both pilE subunits, see **Figure 5.7c-d**, the other only engaged with only one of the pilin subunits, see **Figure 5.7a-b.**

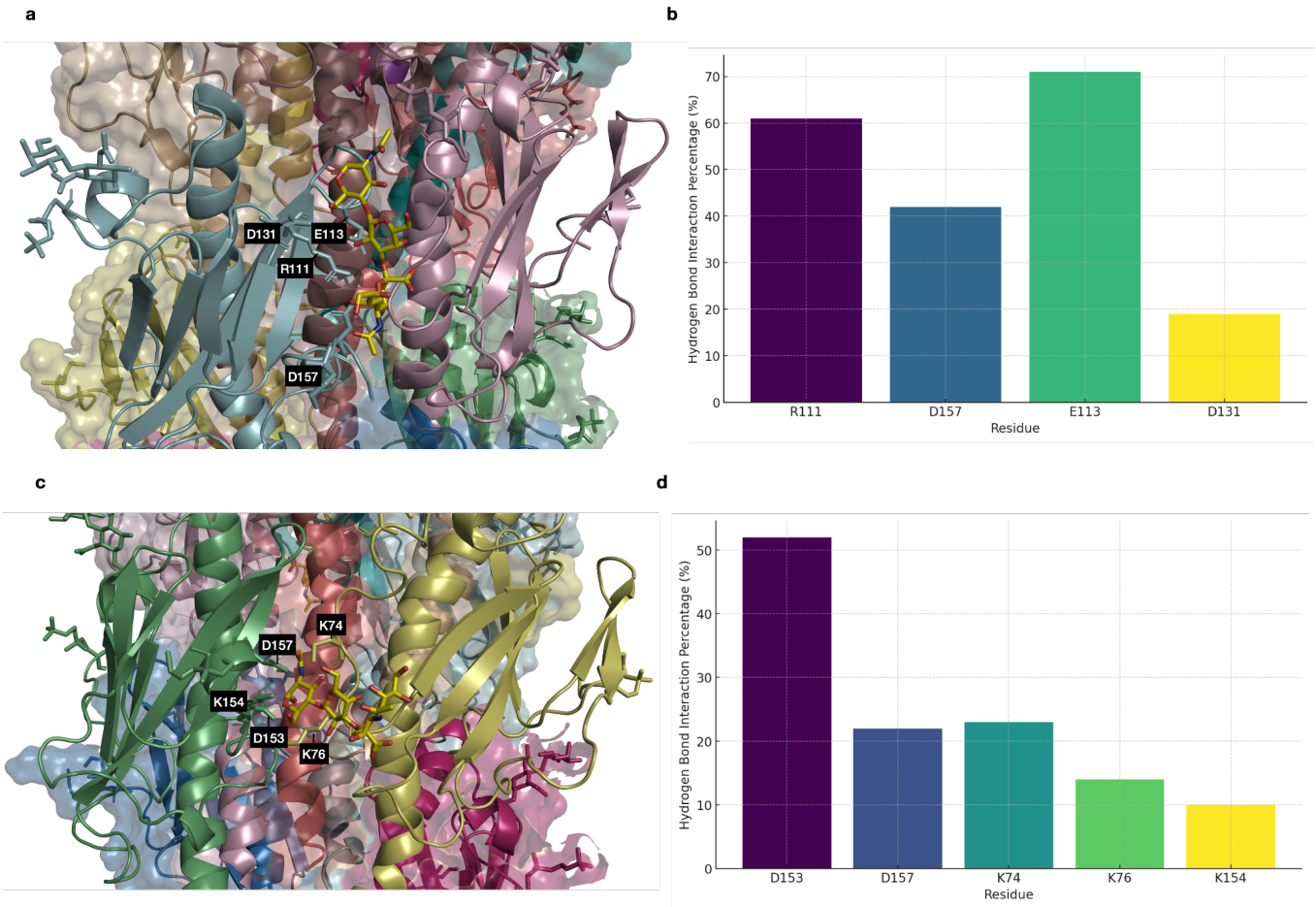


Figure 5.7.Panel a) Bound copy of sialyl-α(2-3)-LacNAc(rendered as yellow sticks), originally aligned using the binding motif of α2,6-linked SialylLacNAc bound to the *Nm* pilE dimer, taken from the MD simulation of the *Ng* pilin array. Key binding residues, namely R111, E113, D131 and D157, are labelled. **Panel b)** Barplot indicating the persistence of the hydrogen bonding interactions between the copy of sialyl-α(2-3)-LacNAc from panel a and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time. **Panel c)** Bound copy of sialyl-α(2-3)-LacNAc(rendered as yellow sticks), originally aligned using the binding motif of α2,3-linked SialylLacNAc bound to the *Nm* pilE dimer, taken from the MD simulation of the *Ng* pilin array. Key binding residues, namely K74, K76, D153, K154, D153, are leablled. **Panel d)** Barplot indicating the persistence of the hydrogen bonding interactions between the sialyl-α(2-3)-LacNAc from panel c and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time.

5.3.5 Blood Groups A and B

Blood group antigens are expressed as part of glycoconjugates in the epithelium of various tissue cells(Marionneau *et al.*, 2001), including the respiratory, urinary and rectal epithelia, regions that are colonised by the *Ng* bacterium. To investigate their potential interactions with the pilin system, I analysed copies of both blood group antigen A type I (BGA), located at the bottom half of the pilin subunits, see **Figure 5.8a-b**, and blood group antigen B type I (BGB), located at the start of the pilin array, see **Figure 5.8c-d**. Both antigens were retained and stable

within their respective binding sites. BGA primarily interacted with residues from the globular domain of a single pilin subunit, whereas BGB formed persistent interactions with residues from the globular domain of both pilin subunits comprising the binding pocket.

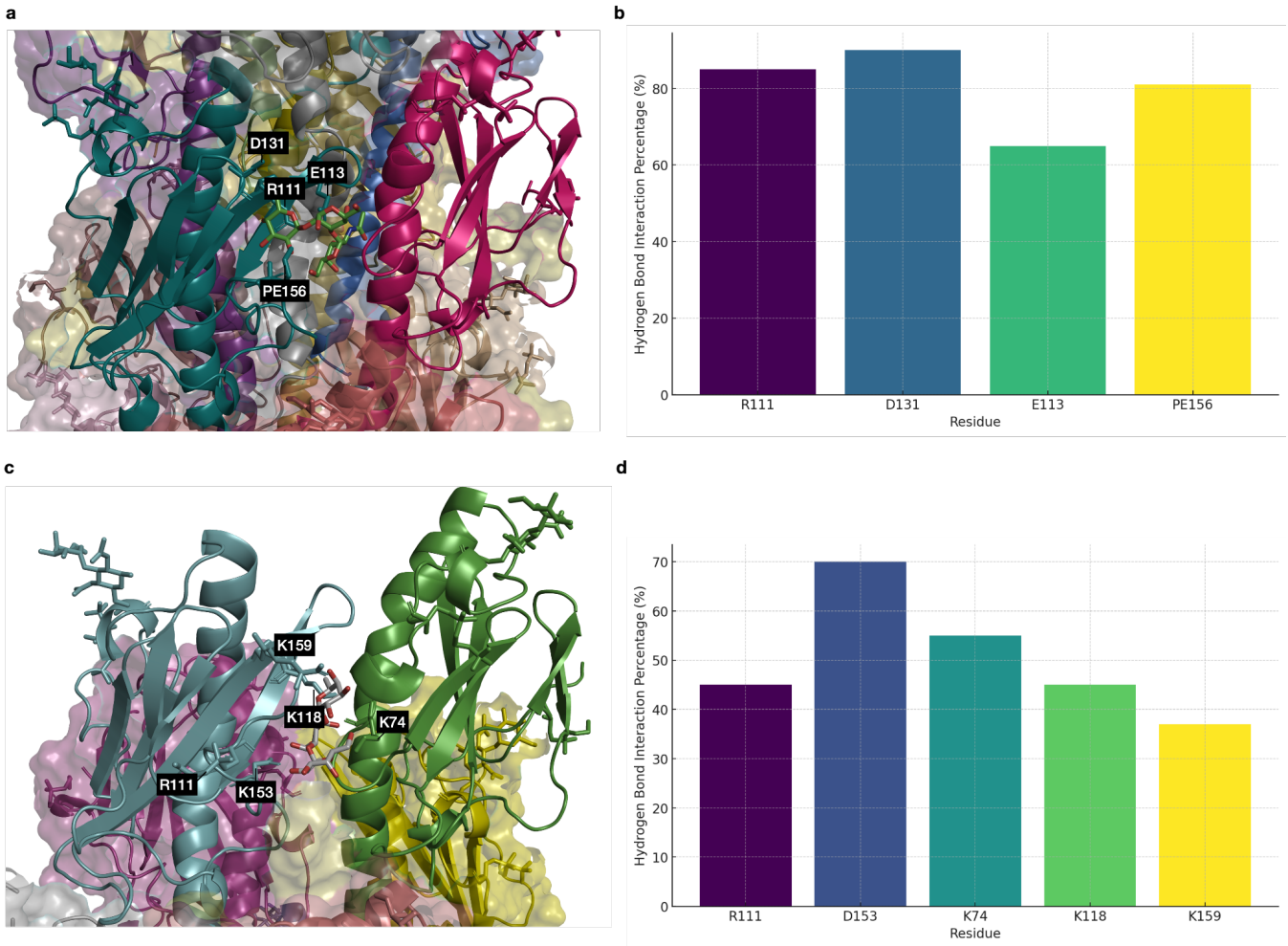


Figure 5.8.a) Bound copy of BGA taken from the MD simulation of the *Ng* pilin array. Key binding residues, namely R111, E113, D131 and PE156, are labelled. **Panel b)** Barplot indicating the persistence of the hydrogen bonding interactions between the BGA and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time. **c)** Bound copy of BGB from the MD simulation of the *Ng* pilin array. Key binding residues, namely K74, R111, K118, K153 and K159, are labelled. **Panel d)** Barplot indicating the persistence of the hydrogen bonding interactions between the BGB and specific binding site residues calculated during the MD trajectory. The values represent the percentage of time

5.3.6 Experimental Binding Assays

To further investigate the binding specificity of the T4P, I performed SPR assays in collaboration with Dr Freda Jen and Dr Chris Day from Prof Micheal Jennings'group at the Institute for Biomedicine and Glycomics at Griffith University in Gold Coast, QLD. SPR experiments were run to screen the binding of the *Ng* bacterial pilin to a wide variety sialylated and non-sialylated glycans. To assess the impact of bacterial glycosylation on the binding

affinity, we tested a ΔPglL mutant that expresses pilin without the bacterial glycan and a ΔPglA mutant that expresses pilin with an *O*-linked monosaccharide, DATDH. Results indicated that no binding of non-sialylated glycans, including blood group antigens, Lewis antigens, and mannose-containing glycans, was observed at the maximum concentration of 100 µM. The calculated binding affinity obtained for the different variants of the *Ng* pilin for sialylated glycan epitopes are shown in **Figure 5.9**.

Table 5.1. Calculated binding affinity (K_D in µM) from a minimum of two biological replicates for WT *Ng* pilin, ΔPglL *Ng* pilin and ΔPglA *Ng* pilin where α2-3SL = Neu5Acα2-3Galβ1-4Glc; α2-8α2-3SL=Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc; α2-8α2-8α2-3SL= Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;
(α2-8Neu5Ac)n~15 = Neu5Aα2-8Neu5Ac with an average of 15 repeats (colominic acid)

Glycan	WT Ng pilin	ΔPglL Ng pilin	ΔPglA Ng pilin
α2-3SL	0.91µM±0.12	1.2µM±0.07	1.9µM±0.09
α2-8α2-3SL	1.4µM±0.2	6.3µM±0.8	3.4µM±0.7
α2-8α2-8α2-3SL	1.9µM±0.4	14.0µM±2.8	4.0µM±0.6
(α2-8Neu5Ac)$_{n\sim15}$	11.0µM±1.0	41.4µM±4.9	23.1µM±4.4

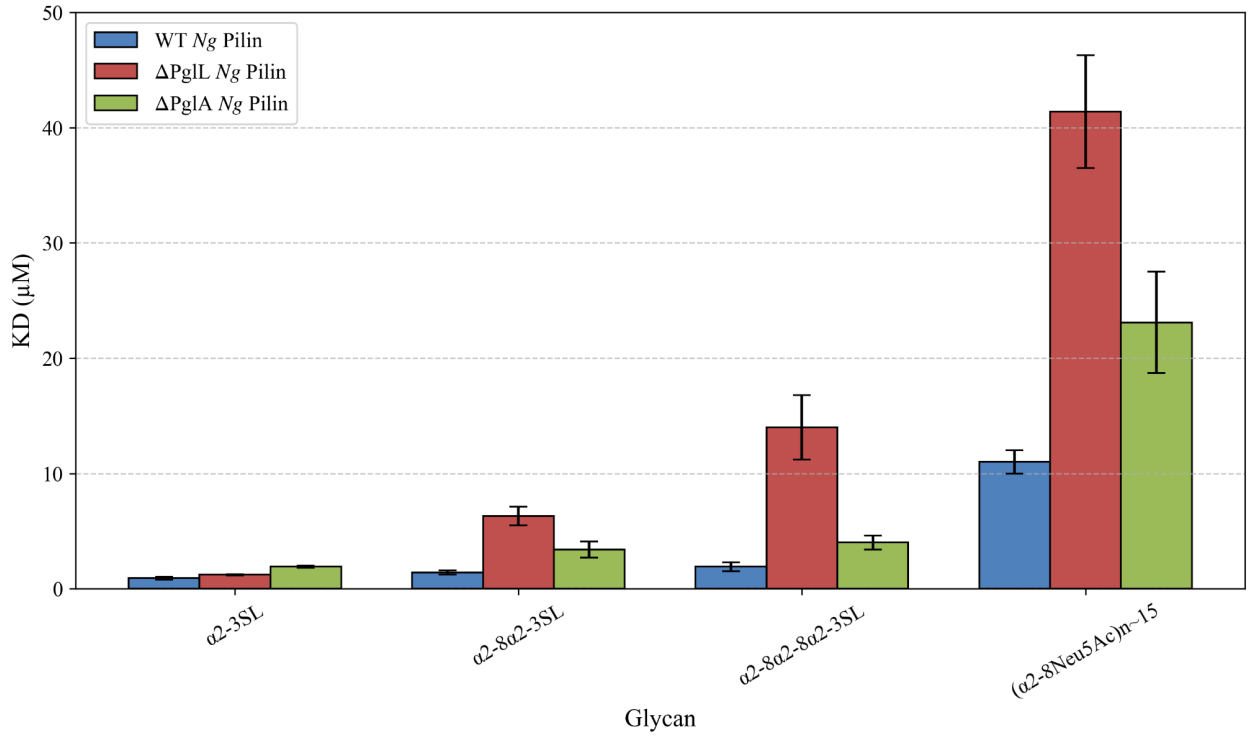


Figure 5.9. Barplot of the calculated binding affinity (K_D in µM) from a minimum of two biological replicates for WT *Ng* pilin, ΔPglL *Ng* pilin and ΔPglA *Ng* pilin where α2-3SL = Neu5Acα2-3Galβ1-4Glc; α2-8α2-3SL=Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;
α2-8α2-8α2-3SL= Neu5Acα2-8Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc;

(α2-8Neu5Ac)n~15 = Neu5Aα2-8Neu5Ac with an average of 15 repeats (colominic acid)

5.4 Discussion

I used MD simulations to study the structure, stability and dynamics of a 27-mer *Ng* T4P model composed of its major pilin subunit, pilE. The results I obtained for the *Ng* pilE system are directly comparable and highly similar to those of the *Nm* pilE system described in **Chapter 4.** As seen in the *Nm* pilE system, the globular domain of the pilE subunits in the *Ng* system showed a high degree of flexibility while the alpha helices remained structurally restricted stabilising the core of the pseudo helical T4P architecture. The rise and twist parameters values I used to assess the dynamics and structural stability of the system, were also comparable to those observed for the *Nm* pilE system and aligned well with experimental results(Craig *et al.*, 2006).

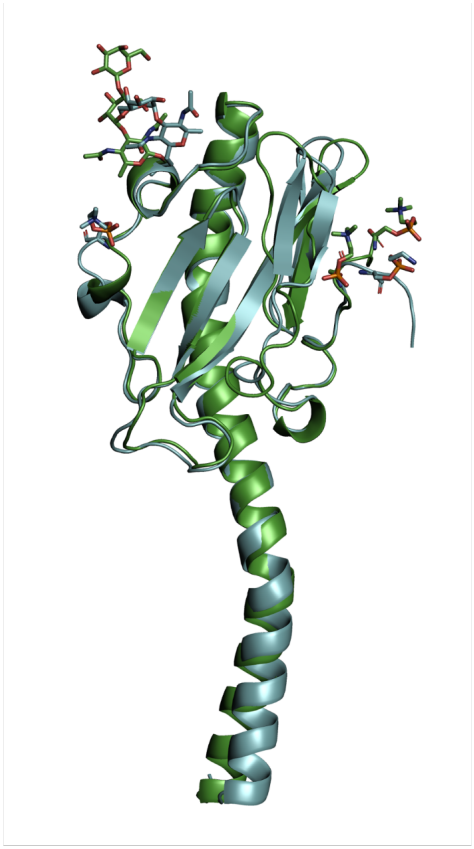
Unlike *Nm* pilin array which was loaded with glycan ligand at half capacity (13 glycans), the MD simulation of the fully loaded *Ng* pilin array (26 glycans) was unable to maintain its structural integrity, with pilin subunits at the top of the system separating away from the core. The increased number of ligands, which now occupied much of the space between pilin subunits, may have restricted the flexibility and space required for each subunit to maintain the structural integrity of the system resulting in the drift of pilin subunits and ultimately in the collapse of the structure. This suggests that the T4P architecture is not built to be fully occupied, but it requires empty clefts to compensate for the structural rigidity imposed by the presence of some glycans between adjacent pilE subunits. Due to the complexity of the system and potential variability of the effects on the T4P structure determined by different ligands, I was not able to assess optimal occupancy. Yet it is structurally unlikely for any T4P to reach a full or nearly full load of glycan ligands even within a crowded environment, such as the extracellular matrix and/or a bacterial colony.

I analysed the binding specificity of the *Ng* T4P for a subset of the glycan epitopes unaffected by the breakdown of the top-half of the array. The *Ng* T4P shows a similar binding preference as the *Nm* pilin, where the glycan epitopes engage through hydrogen bonding interactions with polar residues lining the binding site. More specifically both *Ng* and *Nm* pilE subunits form binding pockets with positively charged residues at neutral pH such as K74, K118 and K76

aswell as other polar residues including N114 and Q/E113. While all of these residues can form hydrogen bond contacts with appropriately oriented hydroxyl groups, the relative abundance of positive charge indicates a preference for negatively charged glycans, such as sialylated epitopes. The 3D structure of the glycans and their inherent flexibility greatly modulate binding specificity. Previously, we have seen that branch fucosylated N-glycans, such as sLex terminating triantennary species, cannot be accommodated in the *Nm* T4P binding sites, in agreement with previous work(Le Guennec *et al.*, 2020). However, smaller fucosylated structures such as the blood group and Lewis antigens,as seen in both arrays, were able to fit within the binding pocket, suggesting that glycan size and branching also influence recognition. Meanwhile, I observed that both *Nm* and *Ng* T4P bind the relatively rigid pseudo-helical structure α2-8 linked polysialic acid engaging it in interactions with similar residues, with a trisaccharide unit as a minimum motif. These results were experimentally confirmed by SPR.

Comparison of the *Nm* and *Ng* pilE subunits shows a high level of similarity, both in terms of sequence and structure, see **Figure 5.10.** In particular, the key polar residues involved in the binding activity for both the *Nm* and *Ng* pilin arrays are conserved. I also noted the presence of polar residues near the end of the *Ng* pilE subunits specifically K159, K157 and D153 that were involved in the overall binding profile of the *Ng* pilin array. These residues are exclusive to the *Ng* pilE subunit; see **Figure 5.10b,** which suggests that their presence enhances the binding capabilities of the *Ng* T4P system compared to the *Nm* T4P system. This was corroborated by the results of the SPR analysis conducted by our collaborators, where the calculated binding affinities of the *Ng* pilin systems for the sialylated glycans epitopes were found to be overall larger than those calculated for the *Nm* pilin system for the same epitopes.

a



b

```
QXZ29465.1      FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGEWPGN    60
EEZ48905.2      FTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGKWPEN    60
                ************************************************:** *

QXZ29465.1      NTSAGVASS-STIKGKYVKEVTVANGVITATMLSSGVNKEIQGKKLSLWAKRQNGSVKWF    119
EEZ48905.2      NTSAGVASPPTDIKGKYVKEVEVKNGVVTATMLSSGVNNEIKGKKLSLWGRRENGSVKWF    120
                ********  : ***** ***** * ***:**********:**:********.:*.:******

QXZ29465.1      CGQPVTRNDTDDTVAAVAADNTGNINTKHLPSTCRDASDAS-      160
EEZ48905.2      CGQPVTRADDDTVAD---AKDGKEIDTKHLPSTCRDKASDAK      159
                ******* * * ..    *.:  :*:********** :. :
```

c

	EEZ48905.2	QXZ29465.1
EEZ48905.2	100.00	80.89
QXZ29465.1	80.89	100

Figure 5.10: Panel a) Structural alignment of the *Ng*(cyan) and *Nm*(green) pilE structures. The backbone alignment of the two protein structures produced a RMSD value of 0.631 Å **Panel b)** Sequence alignment of the *Ng* (EEZ48905.2) and *Nm* (QXZ29465.1) pilE subunits. Highlighted in red are the key residues involved in the binding profile shared by both *Nm* pilin array and *Ng* pilin array. Highlighted in green are the key residues seen more exclusively in the binding profile of the *Ng* pilin array. Sequence alignment conducted using Clustal Omega(https://www.ebi.ac.uk/jdispatcher/msa/clustalo)**Panel c)** Percentage identity scores for the sequence alignment. Molecular rendering done with pymol (www.pymol.org).

The results from the SPR experiments on *Ng* T4P showed that altering the pilE *O*-linked glycosylation has a significant effect on the glycan binding affinity. More specifically, the

ΔPglA mutant that expresses pilEs with reduced *O*-linked glycosylation, retaining only a monosaccharide, showed a higher affinity for the glycan epitopes than the WT. Meanwhile, the ΔPglL mutant that expressed the pilE with no *O*-linked glycosylation showed the highest binding affinity. Similar results were obtained for the *Nm* T4P as discussed in **Chapter 4.** This suggests that the presence of a larger *O*-linked glycosylation as seen in the *Nm* pilin array, while it may be able to interact with the glycan epitopes, may hinder the overall binding ability of the pilin by limiting accessibility to the binding site.

It is also worth noting that while in both the pilin array and SPR experiments, the glycan epitopes we tested were free, in their natural environment these glycans are normally found as part of glycoconjugates expressed on the surface of cells. As such, it is important to keep in mind that we are observing a simplified version of the binding profile between our pilin and the glycan epitopes where we don't consider several factors including the steric limitations imposed by the linkage to their conjugate counterpart. In this respect, both results from the simulations and SPR experiments could be coupled with other experimental techniques including X-ray crystallography or Cryo-Electron Microscopy which could provide an atomistic insight to the binding interaction and build a better picture of the overall binding profile of these interactions (Angulo et al., 2022). NMR methods can also help to characterise these interactions. In particular, ligand-based approaches such as Saturation Transfer Difference NMR can confirm binding events, identify the glycan protons most directly involved in the interaction, and provide affinity estimates, while exchange-transferred Nuclear Overhauser Effect Spectroscopy(Tr-NOESY) experiments can reveal the conformation adopted by the glycan in the bound state (Angulo et al., 2022, 2024).

5.5 Conclusion

In this work, I analysed the structure and dynamic of an *Ng* T4P system composed of the major pilE subunits using MD simulations. The results of this analysis allowed me to make comparisons to the *Nm* pilE system previously described in **Chapter 5.** My analysis of the *Ng* system shows similar structure, stability and dynamics to the *Nm* system. Analogously to the *Nm* T4P study, I explored the binding specificity of the *Ng* pilE system for glycan epitopes using an MD-based pilin array and tested our predictions by SPR. The results of the pilin array

and SPR analysis showed that the *Ng* T4P shows similar binding specificities to *Nm* T4P, e.g. a distinct preference for a(2-8) linked polysialic acids, yet with a slightly higher binding affinity in comparison to the *Nm* pilin system.

References:

1. Alirol, E. *et al*., 2017. Multidrug-resistant gonorrhea: A research and development roadmap to discover new medicines. *PLoS medicine*, 14(7), p.e1002366.

2. Angulo, J. *et al*. 2022 'Structural biology of glycan recognition', in Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 403–418.

3. Angulo, J. *et al*., 2024 'NMR investigations of glycan conformation, dynamics, and interactions', *Progress in nuclear magnetic resonance spectroscopy*, 144-145, pp. 97–152.
4. Blom, A.M. *et al*., 2001. A novel interaction between type IV pili of Neisseria gonorrhoeae and the human complement regulator C4B-binding protein. *The journal of immunology*, 166(11), pp.6764–6770.
5. Cardenas, A.J. *et al*., 2024. Neisseria gonorrhoeae scavenges host sialic acid for Siglec-mediated, complement-independent suppression of neutrophil activation. *mBio*, 15(5), p.e0011924.
6. Craig, L. *et al*., 2006. Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Molecular cell*, 23(5), pp.651–662.
7. Craig, L., Forest, K.T. & Maier, B., 2019. Type IV pili: dynamics, biophysics and functional consequences. *Nature reviews. Microbiology*, 17(7), pp.429–440.
8. Dillard, J.P., 2011. Genetic manipulation of Neisseria gonorrhoeae. *Current protocols in microbiology*, Chapter 4(1), p.Unit4A.2.
9. Edwards, J.L. *et al*., 2016. Is gonococcal disease preventable? The importance of understanding immunity and pathogenesis in vaccine development. *Critical reviews in microbiology*, 42(6), pp.928–941.

10. Edwards, J.L. & Apicella, M.A., 2004. The molecular mechanisms used by Neisseria gonorrhoeae to initiate infection differ between men and women. *Clinical microbiology reviews*, 17(4), pp.965–81.
11. Jen, F.E.-C. *et al*., 2013. Dual pili post-translational modifications synergize to mediate meningococcal adherence to platelet activating factor receptor on human airway cells. *PLoS pathogens*, 9(5), p.e1003377.
12. Jennings, M.P. *et al*., 2011. Neisseria gonorrhoeae pilin glycan contributes to CR3 activation during challenge of primary cervical epithelial cells. *Cellular microbiology*, 13(6), pp.885–896.
13. Jorgensen, W.L. *et al*., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), pp.926–935.
14. Jumper, J. *et al*., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp.583–589.
15. Källström, H. *et al*., 2001. Attachment of Neisseria gonorrhoeae to the cellular pilus receptor CD46: identification of domains important for bacterial adherence. *Cellular microbiology*, 3(3), pp.133–143.
16. Kirschner, K.N. *et al*., 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of computational chemistry*, 29(4), pp.622–655.
17. Kolappan, S. *et al*., 2016. Structure of the Neisseria meningitidis Type IV pilus. *Nature communications*, 7, p.13015.
18. Lee, T.-S. *et al*., 2018. GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. *Journal of chemical information and modeling*, 58(10), pp.2043–2050.

19. Le Guennec, L. *et al*., 2020. Receptor recognition by meningococcal type IV pili relies on a specific complex *N*-glycan. *Proceedings of the National Academy of Sciences of the United States of America*, 117(5), pp.2606–2612.
20. Madeira, F. *et al*., 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic acids research*, 52(W1), pp.W521–W525.
21. Maier, J.A. *et al*., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation*, 11(8), pp.3696–3713.
22. Marionneau, S. *et al*., 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie*, 83(7), pp.565–573.
23. Martín-Sánchez, M. *et al*., 2020. Clinical presentation of asymptomatic and symptomatic women who tested positive for genital gonorrhoea at a sexual health service in Melbourne, Australia. *Epidemiology and infection*, 148(e240), p.e240.
24. Mubaiwa, T.D. *et al*., 2017. The sweet side of the pathogenic Neisseria: the role of glycan interactions in colonisation and disease. *Pathogens and disease*, 75(5). Available at: http://dx.doi.org/10.1093/femspd/ftx063.
25. Obergfell, K.P. & Seifert, H.S., 2016. The pilin N-terminal domain maintains Neisseria gonorrhoeae transformation competence during pilus phase variation. *PLoS genetics*, 12(5), p.e1006069.
26. Punsalang, A.P., Jr & Sawyer, W.D., 1973. Role of pili in the virulence of Neisseria gonorrhoeae. *Infection and immunity*, 8(2), pp.255–263.
27. Quillin, S.J. & Seifert, H.S., 2018. Neisseria gonorrhoeae host adaptation and pathogenesis. *Nature reviews. Microbiology*, 16(4), pp.226–240.

28. Semchenko, E.A. *et al*., 2019. Glycointeractome of Neisseria gonorrhoeae: Identification of host glycans targeted by the gonococcus to facilitate adherence to cervical and urethral epithelial cells. *mBio*, 10(4). Available at: http://dx.doi.org/10.1128/mBio.01339-19.

29. Sievers, F. *et al*., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.

30. Tiemeyer, M. *et al*., 2017. GlyTouCan: an accessible glycan structure repository. *Glycobiology*, 27(10), pp.915–919.

31. Unemo, M. *et al*., 2019. Gonorrhoea. *Nature reviews. Disease primers*, 5(1), p.79.

32. Unemo, M. & Shafer, W.M., 2014. Antimicrobial Resistance in Neisseria gonorrhoeae in the 21st Century: Past, Evolution, and Future. *Clinical microbiology reviews*, 27(3), pp.587–613.

33. Virji, M., 2009. Pathogenic neisseriae: surface modulation, pathogenesis and infection control. *Nature reviews. Microbiology*, 7(4), pp.274–286.

34. Wang, F. *et al*., 2017. Cryoelectron Microscopy Reconstructions of the Pseudomonas aeruginosa and Neisseria gonorrhoeae Type IV Pili at Sub-nanometer Resolution. *Structure (London, England: 1993)*, 25(9), pp.1423–1435.e4.

35. Wang, J. *et al*., 2004. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9), pp.1157–1174.

36. Wi, T. *et al*., 2017. Antimicrobial resistance in Neisseria gonorrhoeae: Global surveillance and a call for international collaborative action. *PLoS medicine*, 14(7), p.e1002344.

Conclusion

My PhD research focused on the analysis of the 3D structure and dynamics of glycans using Molecular Dynamics simulations. Using that knowledge, I explored how their 3D signatures and conformational diversity affects molecular recognition, with a particular focus on proteins involved in the glycan biosynthesis pathway and in bacterial adhesion.

My analysis of the free glycan structures used throughout my PhD highlighted their distinct conformational behaviour, demonstrating how their monosaccharide makeup, branching patterns and linkage types influence glycan flexibility and structural preferences. The MD data collected from this not only contributed to the GlycoShape Glycan 3D Structure Database, but also provided valuable insights into the conformational landscape of these glycans, helping to rationalise their behaviour in the molecular recognition events I studied.

I investigated how the structural modification of bisection can alter the conformational equilibrium of *N*-glycans, making comparisons to triantennary and biantennary *N*-glycan systems. I showed how this modification can hinder interactions with enzymes in the glycan biosynthesis pathway, namely FUT8 and b4GalT1, highlighting its potential role as a "stop-codon" that prevents further functionalisation.

I also explored the role of glycan structure in bacterial adhesion, focusing on the Type IV pili (T4P) of *Neisseria meningitidis* and *Neisseria gonorrhoeae.* My findings showed how the size and arrangement of the T4P system would make it an effective glycan binding scaffold facilitating interactions with epitopes through binding pockets formed between adjacent pilin subunits. Using results from MD simulations that were later validated through surface plasmon resonance (SPR) experiments, I demonstrated that the T4P had a strong preference for polysialylated glycans, particularly α(2-8)-linked polysialic acid, which closely resembles the capsular polysaccharide of *Neisseria meningitidis* Serogroup B. This finding suggests that interactions between T4P and the bacterial capsule may contribute to the colonisation of the bacteria. My results also provided insights into how the post-translational modifications influence glycan binding, specifically how the bacterial *O*-linked glycosylation may hinder binding by limiting access to the binding site.

Overall, my PhD work has provided valuable insights into how glycan 3D structure influences molecular recognition in both glycan biosynthesis and bacterial adhesion. Given that bisected *N*-glycan structures are often implicated in disease states of cells, including cancer progression, the findings from my study of bisected glycans offer structural insights to how bisection hinders functionalisation. These insights could be used to inform diagnostic approaches such as selection of glycan biomarkers for early disease detection. The results from my analysis of the interactions between Type IV pili (T4P) and glycans in their environment, especially the strong preference for α(2-8)-linked polysialic acids, provide new perspectives on the colonisation mechanism of the *Neisseria* bacteria. These insights can be applied to strategizing anti-adhesive therapeutics that can prevent bacteria-cell adhesion and destabilise colonisation.