University of Cape Town

Doctoral Thesis

Modelling and Validation of Bacterial O-antigen Conformations: Ring Puckering in *Shigella flexneri* 7a and 7b O-antigens as a Case Study

---

Author:
Zaheer Timol

Supervisors:
Prof. Michelle Kuttel
Assoc. Prof. Neil Ravenscroft

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the
Faculty of Science

December 2023

Declaration of Authorship

I, Zaheer Timol, declare that this thesis titled, "MODELLING AND VALIDATION OF BACTERIAL O-ANTIGEN CONFORMATIONS: RING PUCKERING IN *Shigella flexneri* 7A AND 7B O-ANTIGENS AS A CASE STUDY" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Abstract

This work forms part of a larger project investigating the conformation and dynamics of *Shigella flexneri* O-antigens. *S. flexneri* is a leading cause of diarrhoeal related diseases, especially in Sub-Saharan Africa and Southeast Asia. There are over 30 *S. flexneri* serotypes and, with the exception of serotype 6, all share a common O-antigen backbone (serotype Y), with variations in glucosylation, phosphorylation, and O-acetylation. O-antigens are carbohydrate polymers on the outer membrane of gram-negative bacteria. In *S. flexneri*, O-antigens are a primary antigenic component and are a target for conjugate vaccines currently in development.

Analysis of O-antigen conformation for different *S. flexneri* serotypes may inform the vaccine development process. However, determining molecular conformation experimentally is challenging. Systematic molecular modelling protocols have proven useful in elucidating conformations of polysaccharide antigens, especially when experimental methods, such as nuclear magnetic resonance, are used for verification of modelling results.

Here we use a combination of molecular modelling and ^{1}H nuclear magnetic resonance spectroscopy experiments to probe the O-antigen conformations of *S. flexneri* 7a and 7b. Simulations of six repeating units of both O-antigens show that they are highly flexible, similar to the *S. flexneri* Y O-antigen. However, we found frequent puckering of the β-D-Glc*p*NAc ring away from the canonical $^4C_1$ conformer, which has not been previously observed in studies of *Shigella* or other bacterial O-antigens. To provide further insight, molecular dynamics and metadynamics simulations of a range of 3,4-disubstituted β-D-GlcNAc trisaccharides with two carbohydrate force fields (CHARMM36 and GLYCAM06) were performed. The simulations reveal that 3,4-α-linked β-D-Glc*p*NAc puckers to a similar extent as the *S. flexneri* 7a and 7b O-antigens. Moreover, for both force fields the range of β-D-GlcNAc puckering is dependant on the anomeric configuration of both the 3- and 4- linkage, with non-$^4C_1$ conformations dominant in 3,4-α-linked β-D-Glc*p*NAc trisaccharides.

^{1}H nuclear Overhauser effect spectroscopy experiments were used to calculate ^{1}H-^{1}H distances in *S. flexneri* 7a and 7b O-antigens. These experimentally derived distances match those calculated from the repeating unit simulations when a mix of $^4C_1$ and boat/skew states of β-D-Glc*p*NAc are considered at a ratio of 85:15. The results also suggest that puckering of β-D-Glc*p*NAc does occur in *S. flexneri* 7a and 7b O-antigens but that it may be over represented in the simulations.

Two general observations can be drawn from this study: (1) the 3,4-disubstitution of β-D-Glc*p*NAc with two bulky substituents leads to the ring puckering out of the $^4C_1$ conformer; and (2) the conformation and dynamic behaviour of β-D-Glc*p*NAc puckering is not accurately modelled by current methods, identifying the need for further enhancements to existing carbohydrate force fields.

Contents

List of Figures

List of Tables

List of Abbreviations

BSE	Block Standard Error
CSD	Cambridge Structural Database
CASPER	Computer-Assisted Spectrum Evaluation of Regular Polysaccharides
CV	Collective Variable
DFT	Density Functional Theory
EABF	Extended System Adaptive Biasing Force
GAG	Glycosaminoglycan
GBS	Group B Streptococcus
GEMS	Global Enteric Multicenter Study
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single Quantum Coherence
IUPAC	International Union of Pure and Applied Chemistry
LD	Langevin Dynamics
LPS	Lipopolysaccharide
MD	Molecular Dynamics
NAMD	Nano Scale Molecular Dynamics
NMR	Nuclear Magnetic Resonance
nOe	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NPT	Constant Number of Particles, Pressure, and Temperature
NUS	Non-Uniform Sampling
NVT	Constant Number of Particles, Volume, and Temperature

O-Ag	O-Antigen
PANIC	Peak Amplitude Normalization for Improved Cross-Relaxation
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PLUMED	Plugin for Molecular Dynamics
PMF	Potential of Mean Force
QM	Quantam Mechanics
RU	Repeating Unit
$\mathbf{R_g}$	Radius of Gyration
RMS	Root Mean Square
RMSD	Root Mean Square Deviation
ROE	Rotating Frame Overhauser Enhancement
TSP	3-Trimethylsilyl-(2,2,3,3-2h4)-Propanoate
***Sf* 7a**	*Shigella flexneri* 7a O-antigen
***Sf* 7b**	*Shigella flexneri* 7b O-antigen
US	Uniform Sampling
VMD	Visual Molecular Dynamics

Chapter 1. Introduction

Carbohydrates are among the most abundant and diverse set of naturally occurring molecules. They are responsible for a vast array of vital biological functions, including defence, cell adhesion, storage, structure, and signaling.[1,2] Monosaccharides are the most basic building block of carbohydrates and can occur in linear or cyclic form. Monosaccharides link to each other to form more complex molecules in the form of di-, oligo-, or polysaccharides.[3]

In bacteria, polysaccharides play an important role as structural components on the outermost layer of the bacterial cell. One example is lipopolysaccharides (LPS), which are uniquely found in the outer-membrane of gram-negative bacteria. A single LPS molecule comprises three covalently linked but distinct domains: a lipid A tail embedded in the membrane bilayer, an oligosaccharide core directly above, and an O-antigen (O-Ag) as the outermost component (Figure 1.1). O-Ags mediate cellular interactions and often play an essential role in bacterial virulence.[4–6] O-Ags are carbohydrate polymers formed by repeated fragments of glycosidically linked monosaccharides.

The primary chemical structure of carbohydrates is the sequence of monosaccharide residues in the glycan chain and their linkage positions. The molecular conformation is the three-dimensional shape adopted by the polysaccharide, and it may have one or more distinct conformations depending on the flexibility of the molecule. The chemical composition and primary structure of O-Ags is strain-specific.[*][2,7] The primary chemical structure of O-Ags are generally well defined for a wide-range of bacterial types.[7] In comparison, molecular conformations of O-Ags have not been as widely studied.[8]

Determining the conformation of polysaccharides can be challenging as their inherent flexibility means they can adopt multiple conformations. Molecular conformation can be investigated using either experimental or computational techniques. However, using a combination

---

[*] A bacterial strain is a variant of a species that has unique genetic, biochemical, or phenotypical characteristics that distinguish it from other bacteria of the same species.

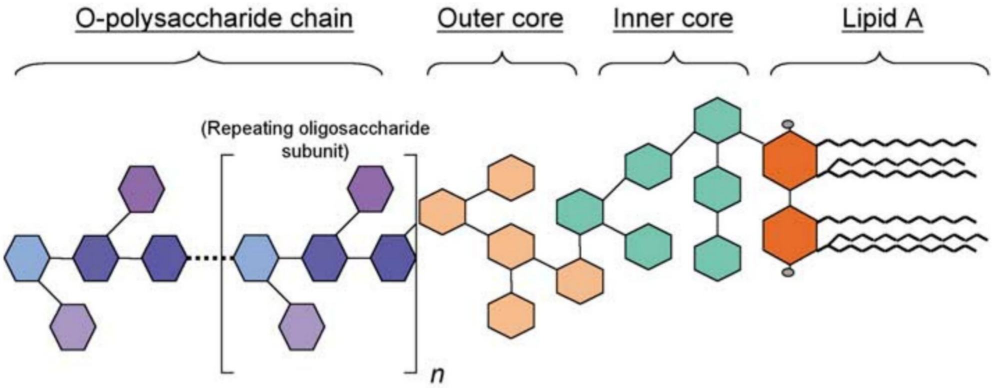


Figure 1.1: Structural illustration of a general lipopolysaccharide molecule illustrating its three component parts. Reproduced with permission from [9] © Elsevier.

of both approaches can provide a more reliable measure of the molecule's conformational behavior.[8,10,11]

The direct conformational elucidation of polysaccharides conformation is commonly performed by two approaches: Nuclear Magnetic Resonance (NMR) spectroscopy (typically in solution) and X-ray crystallography (for a crystalline solid). While these methods are capable of providing detailed atomic level information they have several limitations, such as those associated with sample preparation, data resolution, and reproducibility.[12–14]

Molecular modelling using quantum mechanics (QM) or molecular dynamics (MD) methods can provide useful theoretical insights into the chemical, conformational, and dynamic properties of molecules that are not easily measured experimentally. Although QM methods are more accurate in energy and structure calculations, classical MD simulations (without an explicit quantum effect) are the preferred method for modelling large molecules, such as polysaccharides, due to its lower computational cost and ability to simulate the dynamics of a molecule over time.[8,10]

A central component in all MD simulations is the force field — a mathematical functional form and its corresponding parameter sets that define the potential energy of a system of atoms based on their positions and orientations. Simulations with different force fields can yield different molecular conformations and dynamics and these differences may at time be significant. Therefore, modelling using multiple force fields together with experimental measurements can enhance the reliability of conformational analysis.[8,10]

Conformational studies of polysaccharide antigens have aided in understanding the immunogenicity[†] of numerous bacterial pathogens[8,15–27] and may be useful in informing the production of novel carbohydrate-based vaccines.[8,17,19] Such carbohydrate-based vaccines have proven to be effective against a number of infectious pathogens including *Neisseria meningitides*, *Haemophilus influenzae*, and *Streptococcus pneumoniae* among others.[5]

Novel vaccines are developed to have low complexity and cost, broad serogroup/serotype coverage (through cross-protection), and maximum immune response. Both coverage and immune response to a carbohydrate-based vaccines is dependant on the primary chemical structure as well as the molecular conformations of the polysaccharide antigen.[4–6,8] The conformations of polysaccharide antigens can be studied by MD simulations to provide insights into the immune response of vaccines and their potential ability to provide cross-protection.

There are two primary types of carbohydrate-based vaccines: polysaccharide vaccines and glycoconjugate vaccines.[5] The cell-surface polysaccharide of the bacteria serves as the primary antigenic component in polysaccharide vaccines. In contrast, glycoconjugate vaccines comprise a polysaccharide or oligosaccharide component linked to a protein carrier that is suitably immunogenic. These vaccines have been observed to be more immunogenic, effective in young children, and are now favoured over polysaccharide vaccines.[28–30]

Carbohydrate-based vaccines are currently in development against several pathogens including those associated with diarrhoeal or dysentery diseases,[5,31] such as *Escherichia coli*, *Salmonella*, *Campylobacter*, and *Shigella*,[32] which are among the leading causes of mortality in the developing world.[32–35] Over ninety-percent of all infections and deaths occur in impoverished countries, with the highest rates reported in south Asia and sub-Saharan Africa (Figure 1.2).[32–37]

This work focuses on comparing the conformations of two structurally similar O-Ags from *Shigella flexneri* to better understand the relationship between bacterial antigenic structure and activity. *S. flexneri* is one of the leading causes of shigellosis; a gastrointestinal disease caused by the *Shigella* family of bacteria.[38] Shigellosis is among the foremost causes of death worldwide,[38–40] particularly among infants in developing countries.[32–35] Treatment of shigellosis infection has suboptimal outcomes in regions without universal provision of potable water and sanitation. Moreover, increasingly widespread resistance to antibiotic therapies means that preventing infection is even more important. Vaccines provide a safe and highly effective means of disease prevention. However, there are currently no licensed vaccines against *Shigella*.[41]

[†] Immunogenicity is the ability of a foreign substance, such as an antigen, to elicit an immune response in an organism.

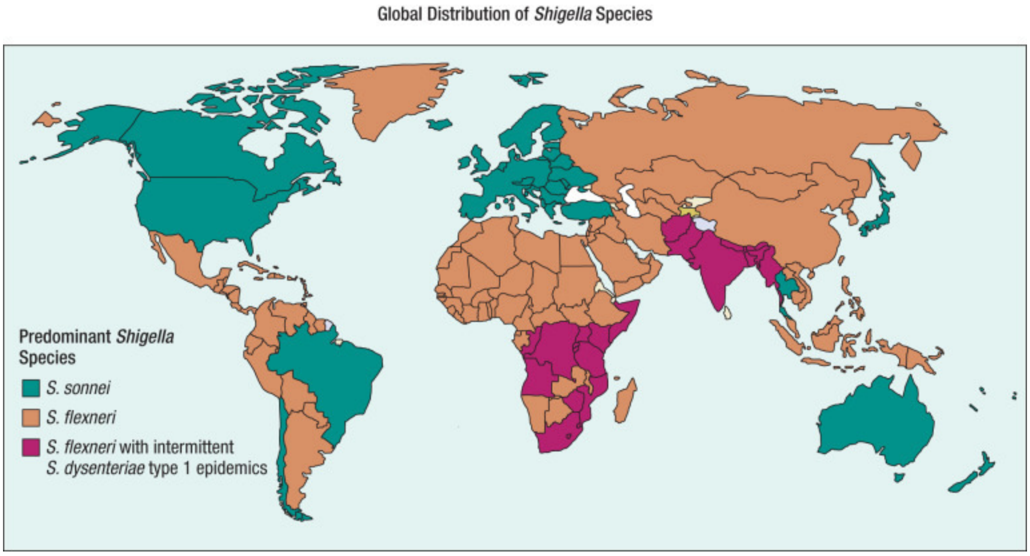


Figure 1.2: Global distribution of predominant *Shigella* species. Reproduced with permission from [35] © Elsevier

Prevention of shigellosis should ideally be based on well established personal and food hygiene practices. However, this is challenging in resource-limited areas (particularly in the developing world) where access to clean water and sanitation is often scarce. A low-cost and efficacious vaccine would complement existing efforts and accelerate disease reduction.[42] There are numerous *Shigella* vaccines under development including live attenuated,[43,44] formalin-killed whole-cell, subunit, and novel antigen vaccines.[42] All carbohydrate vaccines in development against *Shigella* are glycoconjugate based and multivalent to provide cross-protection against multiple serotypes.[41,45–47]

Vaccine cross-protection refers to a phenomenon in which a vaccine developed against one type of bacterial strain provides a certain degree of protection against related, yet distinct, strains. This occurs when shared antigens or structural similarities between different pathogens trigger an immune response that offers defense beyond the target pathogen. Cross-protection can be particularly beneficial when it is difficult to create vaccines for every variant of a pathogen, as in the case of *S. flexneri* which has multiple serotypes. By generating immune responses that recognize and combat multiple strains or species, cross-protective vaccines enhance the effectiveness of immunization strategies and contribute to broader disease control efforts.

The Global Enteric Multicenter Study (GEMS) suggests that a vaccine including *S. sonnei* and three *S. flexneri* serotypes (2a, 3a, and 6) could offer direct protection against 64% of *Shigella* strains responsible for causing moderate-to-severe diarrhea in children residing in low-income and middle-income settings. If cross-protection between various *S. flexneri* serotypes can be demonstrated a four-component vaccine encompassing *S. sonnei* and *S. flexneri* 2a, 3a, and 6 O-Ags could provide a comprehensive coverage (both direct and cross-reactive) for as much as 88% of Shigella strains. The O-Ag of *S. flexneri* is used in glycoconjugate vaccines and is a key antigenic and immunogenic component. The O-factor is the antigenic structure of the O-Ag and is responsible for eliciting a protective immune response.[32,48,49]

All but one *S. flexneri* serotype (serotype 6) share a common O-Ag repeating unit backbone chemical structure: a tetrasaccharide containing three L-rhamnose residues (Rhap^{I} - Rhap^{III}) and a single 2-acetamido-2-deoxy-D-glucose residue.[50] Differences in serotype and subtypes arise due to O-acetylation, glucosylation, and phosphorylation at varying positions on the repeating unit.[51] While the primary structures of *S. flexneri* O-Ags are similar, their unique substitution patterns may give rise to distinct molecular conformations.

The O-Ag conformations of *S. flexneri* serotypes Y, X, 1, 2, 3, 5 and 6 have been explored using both MD simulations and NMR spectroscopy.[15,16,22–25] The serotype Y O-Ag is highly flexible and adopts both extended and compact conformations, including right-handed helices. Backbone substitution with glucosyl and O-acetyl groups in *Sf* 2, 3, and 5 generally restricts O-Ag flexibility. Moreover, the specific substitution position can also either reduce chain extension or induce helical conformations in the backbone in O-Ags of the three serotypes.[22–25] Similarly, the serotype 6 O-Ag is also less flexible than the serotype Y O-Ag and adopts predominantly extended conformations, while O-acetylation does not significantly affect backbone structure.[24] Additionally, the O-Ag conformations of serotypes 1 and 4 have been studied using molecular modelling but in less detail.[15] *S. flexneri* 7 O-Ag conformations have not been previously investigated.

The *S. flexneri* 7a and 7b O-Ags have an α-D-Glcp^{II}-(1→2)-α-D-Glcp^{I} side-chain connected to the 4-position of β-D-Glc*p*NAc.[52,53] In addition, the *S. flexneri* 7b O-Ag repeating unit has an O-acetyl group at the 2-position of α-L-Rhap^{I} (Figure 1.3). The common O-factor in *Sf* 7a and 7b: α-L-Rhap^{I}-(1→3)-[α-D-Glcp^{I}-(1→4)]-β-D-Glc*p*NAc, has an unusual 3,4-disubstitution. The conformational behavior of this vicinal disubstitution motif of β-D-Glc*p* derivatives has only been studied in trisaccharides, which revealed unusual orientations of the 1→4 linkage.[54] As such, elucidating the conformation and dynamics of the *S. flexneri* 7a and 7b O-Ags and understanding the behaviour of the 3,4-disubstituted β-D-Glc*p*NAc residue in the common O-factor is the focus of this study, to provide further insight into *S. flexneri* O-Ag conformations and potentially aid in future vaccine development processes.

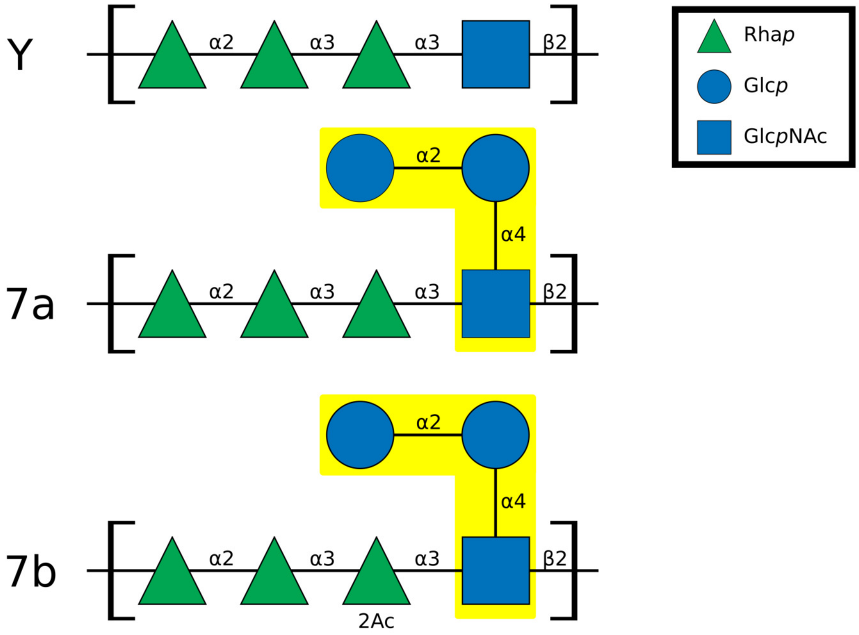


Figure 1.3: SNFG schematic representations of the repeating unit of *S. flexneri* Y, 7a, and 7b. The O-factor common to serotypes 7a and 7b is highlighted in yellow.[51]

1.1 Problem statement

Modelling *S. flexneri* O-Ag conformation can inform future vaccine development by providing insight into immune responses as well as potential cross-protection between serotypes. The conformations of several *S. flexneri* O-Ags have been elucidated, however, the effect of the α-D-Glc*p*$^{\text{II}}$-(1→2)-α-D-Glc*p*$^{\text{I}}$ side-chain on backbone conformation and dynamics for *Sf* 7a and 7b is unknown; as is the influence of α-L-Rha*p*$^{\text{I}}$ 2-O-acetylation on *Sf* 7b. Molecular modelling, combined with NMR spectroscopy, is proven to be a useful method for investigating polysaccharide conformation. Conformations predicted by molecular modelling are dependant on a number of factors including the type of force field employed and results between force fields sometimes differ significantly. Molecular modelling of *S. flexneri* O-Ags with different force fields and its effect on conformation is yet to be investigated.

1.2 Aims

We aim to study the conformation and dynamics of *S. flexneri* 7a and 7b O-Ags by molecular modelling and NMR spectroscopy, and elucidate the conformational characteristics of the β-D-Glc*p*NAc residue, which is central to the common O-factor in the O-Ags. We will perform the simulations with two widely used carbohydrate force fields (CHARMM36[55–60] and

GLYCAM06[61,62]) and compare interproton distances from the simulations with measurements from NMR experiments of the two O-Ags. We will also compare our observations with the previously modelled *S. flexneri* Y O-Ag conformations. Through this work we aim to further our understanding of the relationship between *S. flexneri* O-Ag conformations, and the immune response and cross-protection of potential vaccines.

1.3 Research questions

This work addresses the following research questions:

1. **Does the α-D-Glc*p*$^{\text{II}}$-(1→2)-α-D-Glc*p*$^{\text{I}}$ side-chain on *Sf* 7a and 7b affect the O-Ag conformation and dynamics as compared to *Sf* Y?** Does 2-OAc of α-L-Rha*p*$^{\text{I}}$ in *Sf* 7b change the O-Ag conformation and dynamics compared to *Sf* 7a? Are the conformations of *S. flexneri* 7a and 7b similar to other *S. flexneri* O-Ags studied?
2. **Do the conformations of *S. flexneri* 7a and 7b O-Ags differ across force fields?** Are there conformational and dynamic differences in the backbone and residues of the O-factors when modelled with CHARMM36 compared to GLYCAM06?
3. **Do NMR spectroscopy measurements corroborate the conformations of *Sf* Y, 7a, and 7b identified from simulation?** Can conformationally diagnostic inter-residue ^{1}H-^{1}H distances be measured by NMR spectroscopy? If so, do these distances closely align with those observed in the MD simulations?

1.4 Approach

We employ a systematic approach to the modelling and validation of *S. flexneri* 7a and 7b O-Ag conformations. For each O-Ag we build a six repeating unit (6 RU) polysaccharide using low energy torsion angles for the glycosidic linkages, as calculated from potential of mean force (PMF) simulations of representative disaccharides. This 6 RU polysaccharide is used as the starting conformation for the all-atom solution MD simulations with the CHARMM36 force field. CHARMM36 is employed as it has been used repeatedly in the research group to effectively model *Shigella* and other O-antigens/capsular polysaccharides. We evaluate key conformational metrics, such as the polysaccharide end-to-end distance, glycosidic torsion angles, and ring pucker geometries. MD simulations of a *Sf* Y 6 RU model was used the primary point of comparison with contrasts also drawn against O-Ag modelling of other *S. flexneri* serotypes.[22,24]

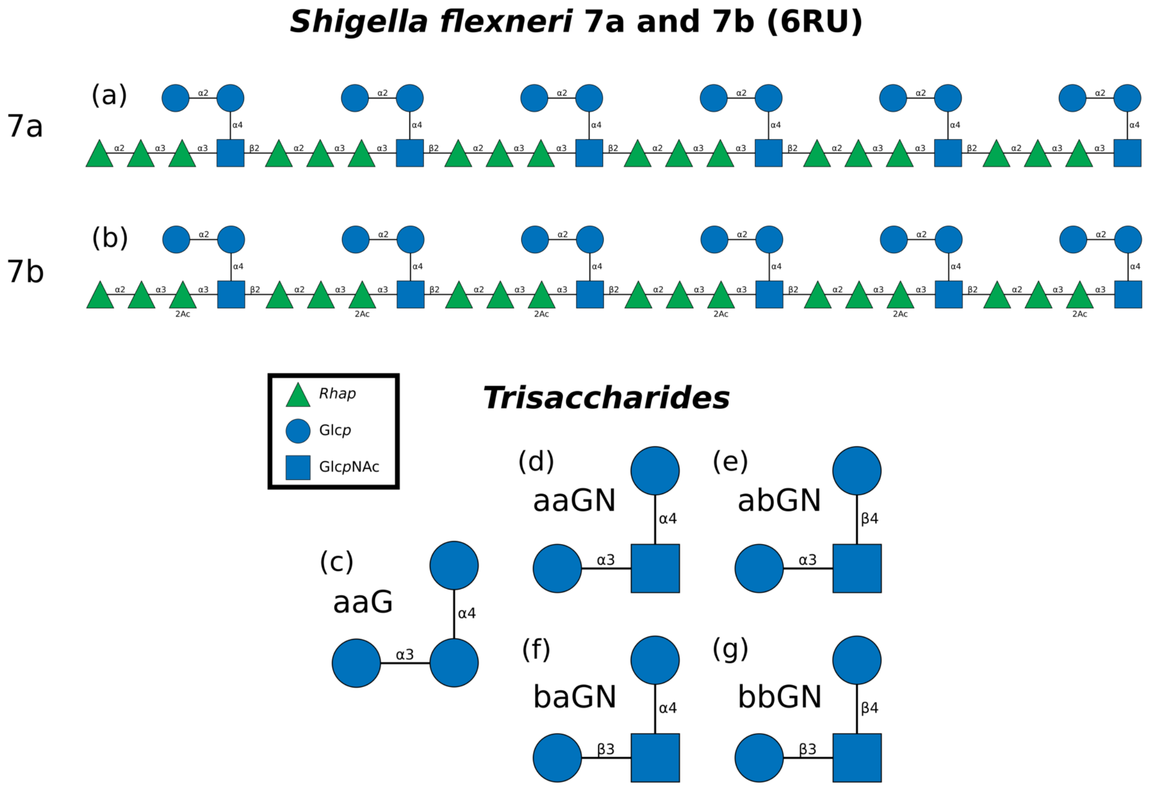


Figure 1.4: SNFG schematic representations of the saccharides modelled within this work. (a) and (b) RU of the *S. flexneri* 7a and 7b O-Ag. (c)-(g) 3,4-disubstituted *β*-D-Glc*p* and *β*-D-Glc*p*NAc trisaccharides.

Six-repeating unit simulations of *Sf* 7a and 7b (Figure 1.4: a, b) using the CHARMM36 force field showed puckering of *β*-D-Glc*p*NAc residues. These observations were also observed in simulations of shorter 1 RU and 3 RU chains of *Sf* 7a using the same force field and a number of different starting orientations for the glycosidic linkage torsion angles.

We performed further modelling to study *β*-D-Glc*p*NAc ring conformation to determine if the results from the *S. flexneri* 7a and 7b simulations may be an artefact of CHARMM36. A series of MD simulations were carried out with 3,4-disubstituted trisaccharides (Figure 1.4: c-g), using the CHARMM36 and GLYCAM06 force fields. While a different force field could be chosen for comparison with CHARMM36, we selected GLYCAM06 as it has been used to extensively study *β*-D-Glc*p*NAc ring pucker.[63] We modelled the central glucose residue with and without a 2-NAc substitution as well as all possible *α* and *β* anomeric configurations. This was done to measure and compare the conformation, dynamics, and energies of the 3,4-disubstituted trisaccharides across three metrics: (1) different force fields; (2) varying anomeric configurations; and (3) with and without 2-N-acetylation on *β*-D-Glc*p*NAc. We also compare the different effects of CHARMM36 and GLYCAM06 on *β*-D-Glc*p*NAc conformation in 1 RU of *Sf* 7a.

We conducted NMR spectroscopy experiments to investigate *S. flexneri* 7a and 7b O-Ag conformations and to study the β-D-Glc*p*NAc ring empirically. This included ^{1} NMR spectroscopy temperature measurements, homonuclear ^{3}J coupling analysis, and ^{1}H, ^{1}H-NOESY studies. Finally, we combine the results from the MD simulations and NMR spectroscopy experiments to predict O-Ag conformation.

1.5 Contribution

We establish that the α-D-Glc*p*$^{\mathrm{II}}$-(1→2)-α-D-Glc*p*$^{\mathrm{I}}$ side-chain within the O-Ag repeating unit of *Sf* 7a and 7b significantly influences backbone conformation. We also reveal unusual puckering of the β-D-Glc*p*NAc ring in the molecular modelling of both *Sf* 7a and 7b and show that ^{1}H NMR experiments of the O-Ags indicate a less pronounced puckering of the β-D-Glc*p*NAc ring in comparison. Furthermore, we identify similar ring pucker conformations in 3,4-disubstituted β-D-Glc*p*NAc trisaccharides and demonstrate correlation between substituent anomeric configuration and the duration of β-D-Glc*p*NAc ring pucker. These results further our understanding of *S. flexneri* O-Ag conformation and provide valuable insights on the contrasting effects of glycoslation and acetylation on the O-Ag backbone, which may be useful in the development of polysaccharide-based vaccines against *Shigella*.

1.6 Thesis Overview

Chapter 1 (this chapter) provides background as well as the aim and approach.

Chapter 2 includes an overview of carbohydrate structure and conformation and details the different *Shigella* O-Ags.

Chapter 3 describes the NMR spectroscopy of polysaccharides.

Chapter 4 outlines the molecular modelling of polysaccharides.

Chapter 5 provides a review of bacterial polysaccharide antigen modelling.

Chapter 6 discusses the molecular modelling of *S. flexneri* 7a and 7b O-Ags. The structure and flexbility of the modelled repeating units are analysed and compared to previously modelled of *S. flexneri* O-Ags.

Chapter 7 details the molecular modelling of 3,4-disubstituted β-D-Glc*p*NAc trisaccharides with two force fields: CHARMM36 and GLYCAM06.

Chapter 8 contains conformational studies of *S. flexneri* O-Ags 7a and 7b O-Ags by NMR.

Chapter 9 gives the conclusions.

Chapter 2. Bacterial Carbohydrates

Carbohydrates are a diverse class of organic molecules that are essential for life processes and play an important role in bacterial cell communication and defence. Comprising monomeric units called monosaccharides, carbohydrates range from simple sugars to complex, branched structures, such as disaccharides and polysaccharides. These molecules serve as vital energy sources, structural elements, and signaling molecules in organisms. Carbohydrates also play key roles in cell recognition, immune responses, and cellular communication. The three-dimensional configurations, or conformations, of carbohydrates can influence their biological function, making their study important in fields such as biochemistry, molecular biology, and vaccine development. This chapter provides an overview of the carbohydrate structure, highlights the role of lipopolysaccharides in gram-negative bacteria, and discusses the different *S. flexneri* carbohydrate O-Ags.

2.1 Carbohydrate Structure

Carbohydrates molecules have a wide range of structures. Monosaccharide units, such as glucose and fructose, have various sugar types (Figure 2.1) and form the fundamental building blocks of more complex carbohydrates. These range from simple disaccharides, where two monosaccharides are linked, to oligosaccharides, comprising short chains of monosaccharides. The structural variety further extends to polysaccharides, which are composed of extended chains or networks of monosaccharide units. Variations in the type, arrangement, and linkage of monosaccharides contribute to the vast structural diversity of carbohydrates.

2.1.1 Monosaccharides

Monosaccharides are the basic building blocks of carbohydrates and have the molecular formula $C_nH_{2n}O_n$, where $n \geq 3$. A typical unsubstituted monosaccharide has a single aldehyde or ketone group and a single hydroxyl on each of the remaining carbon atoms. They commonly exist in both an open-chain form and a closed ring form, where the aldehyde/ketone

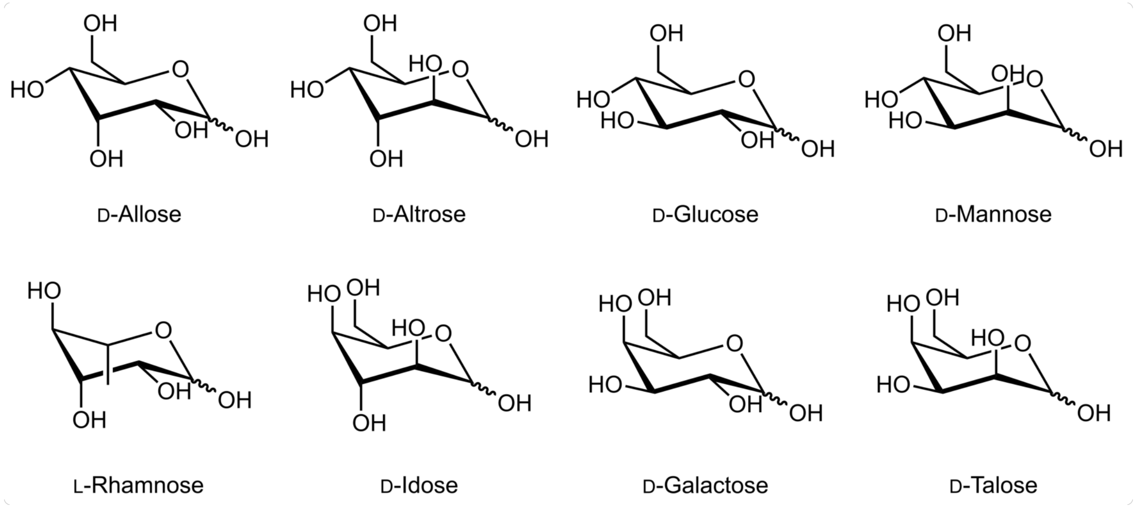


Figure 2.1: A few common unsubsituted hexose sugars in pyranose form.

carbonyl group carbon atom and hydroxyl group combine to form a hemiacetal with a new C — O — C bond. Each open-chain monosaccharide yields two cyclic diasteriomers, denoted by the prefixes $\alpha-$ and $\beta-$ (Figure 2.2). The molecule can change between these two forms by a process called mutarotation. Monosaccharides can also occur as two enantiomers designated with a D– prefix or an L– prefix.

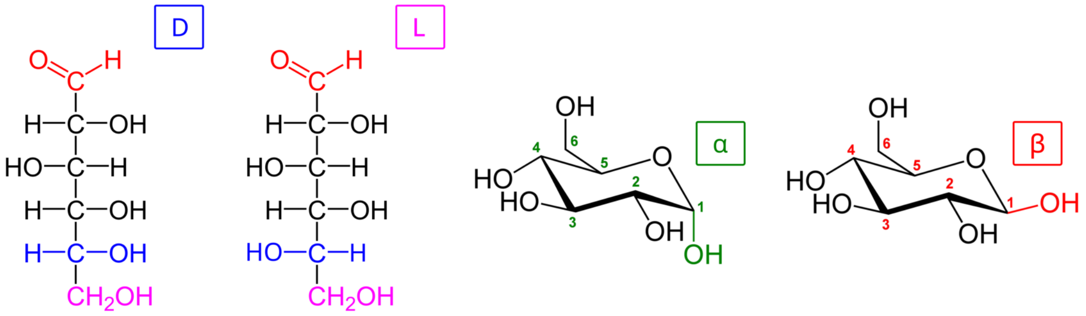


Figure 2.2: (Left) Fischer projection of D-glucose and L-glucose. The red atoms highlight the aldehyde group and the blue atoms highlight the asymmetric centre furthest from the aldehyde. In D-glucose this -OH is on the right of the Fischer projection and in L-glucose it is on the left. (right) The α and β anomers of glucopyranose. Note the position of the hydroxyl group (green or red) on the anomeric carbon relative to the CH_2OH group bound to carbon 5: they either have identical absolute configurations (R,R or S,S) and are given an α assignment, or opposite absolute configurations (R,S or S,R) and are given a β assignment.

Different functional groups can also be substituted at various positions on a monosaccharide residue. These substitutions can change the physical and chemical properties of a monosaccharide, altering their roles in various biological processes. The wide range of possible substituents and the position at which they bond contribute to the diversity and complexity of monosaccharides. Typical monosccharide substituents include methyl, O-acetyl, and N-acetyl. In bacteria, such as *Shigella*, common monosaccharides are glucose with and without 2-N-acetylation and Rha*p*, which is methylated at the 5-position. Substitution can alter monosaccharide conformation by interactions with other functional groups or bending (puckering) of the molecular ring.[63,64]

The ring atoms in cyclic monosaccharides are not planar. Puckering of the ring atoms away from a planar state minimises the ring torsional strain. Typically, pyranose (cyclic hexose monosaccharides) sugars have their ring conformations in the low-energy chair conformation.[65] The pyranose form of aldohexose sugars is predominant in aqueous solution and $^4C_1$ is the prevalent state for all but D-idose.[65–67]

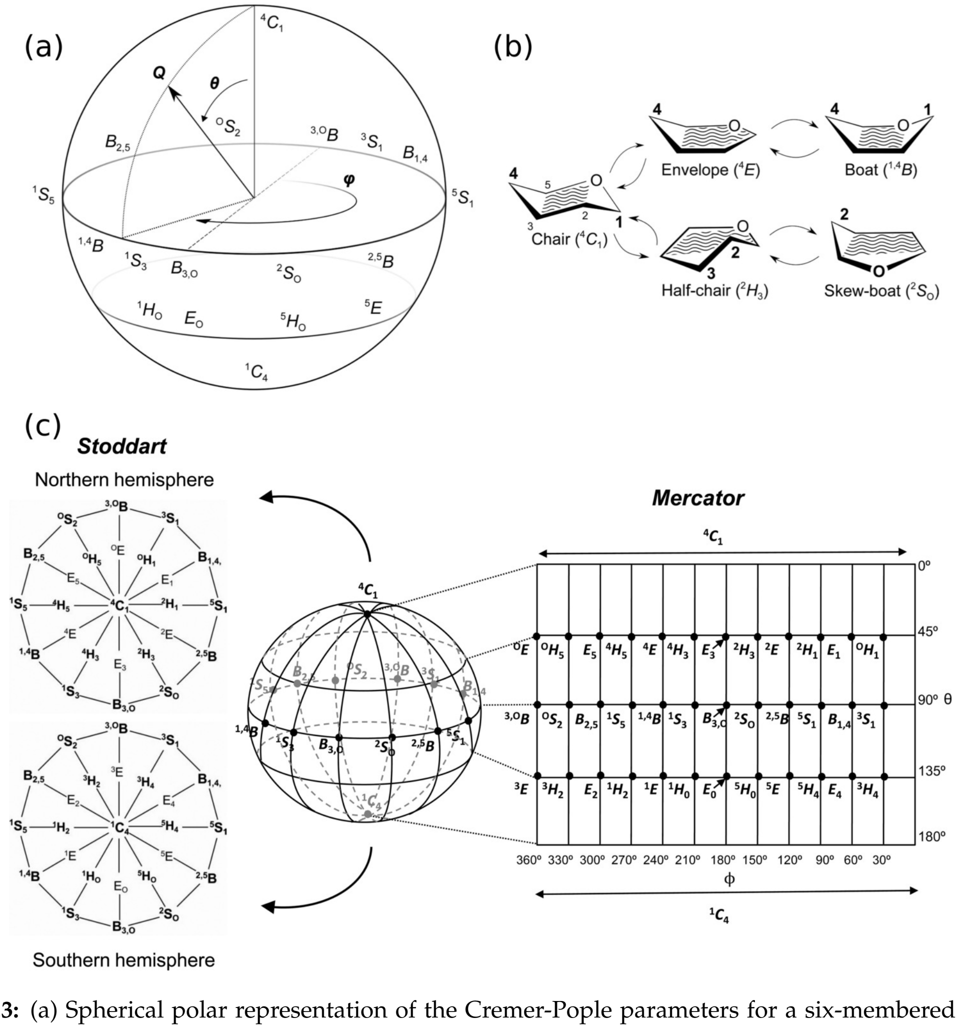


Figure 2.3: (a) Spherical polar representation of the Cremer-Pople parameters for a six-membered ring, depicting the meridian angle (ϕ), azimuth angle (θ), and pucker radius (Q);[68] (b) illustrations of the boat, skew (skew-boat), envelope, and half-chair ring pucker conformers;[68] (c) Stoddart and Mercator representations of the Cremer-Pople sphere. Reproduced with permission from [69] © American Chemical Society.

The set of possible ring pucker conformations are defined by the International Union of Pure and Applied Chemistry (IUPAC) on nomenclature originally described by Shwarz.[70] Cremer and Pople defined a set of spherical coordinates to uniquely describe different ring pucker conformers for rings of any size.[71] With the Cremer-Pople method ring coordinates are initially mapped from Cartesian space to spherical space where three parameters (ϕ, θ, and Q) fully describe the conformation of the ring (Figure 2.3). The first two parameters provide the precise canonical ring pucker, while Q gives the pucker amplitude. The two idealised chair conformers are at the poles of the Cremer-Pople sphere with values of 0° or 180°. Boat (B) and skew (S) conformers are at the equator of the sphere with $\theta = 90°$, while half-chair (H) and envelope (E) puckers are at the tropics ($\theta = 45°$ and $\theta = 135°$). The specific atom numbering associated with the different ring conformations are defined by their IUPAC designation.[72]

2.1.2 Disaccharides

Disaccharides are formed by the linkage of two monosaccharides through a glycosidic bond. The glycosidic bond forms from the anomeric carbon of one monosaccharide to any position on a second monosaccharide. Two torsion angles, ϕ and ψ* (and a third, ω, in the case of 1→6 linkages) define the conformation of a glycosidic linkage. There are a large range of possible spatial configurations that two glycosidically linked monosaccharides can adopt as each residue has the potential to rotate about the bond. The anomeric configuration, either α to β, can also increase or decrease the rotational freedom of a glycosidic linkage.

Glycosidic linkage orientation is the largest contributing factor in determining disaccharide conformation.[8,13,65,73,74] The rotation around glycosidic linkages in disaccharides have been extensively studied using both NMR spectroscopy and molecular modelling.[17,75–78] Syn-ψ orientations are energetically favourable and more frequently adopted in disaccharides, while anti-ψ orientations are preferred over anti-ϕ orientations (Figure 2.4).[13,16,17,75–78] As a glycosidic bond can form between any two monosaccharides with many different points of linkage, the range of different glycosidic bonds adds significant complexity to the study of disaccharides and more complex carbohydrates, such as oligo- and polysaccharides.

---

[*] defined by IUPAC nomenclature: $\phi = H_1 - C_1 - O_1 - C'_x$ and $\psi = C_1 - O_1 - C'_x - H'_x$

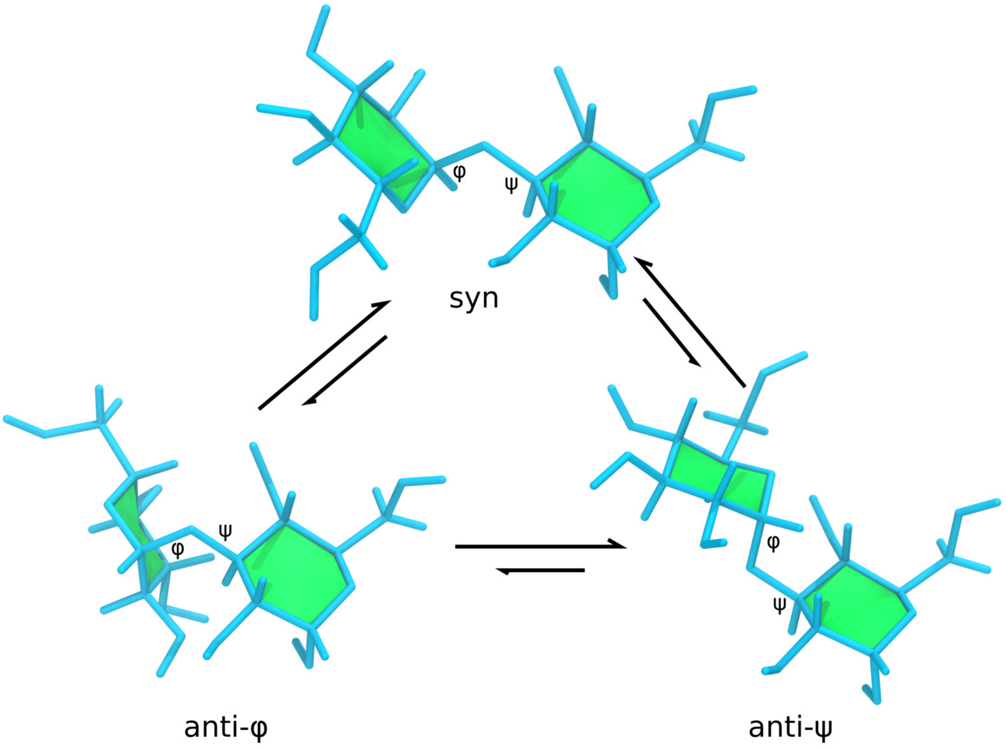


Figure 2.4: α-D-Glcp^{II}-(1→2)-α-D-Glcp^{II} in syn, an anti-ϕ, and an anti-ψ orientation. Glycosidic linkages in disaccharides often prefer the syn orientations anti-ψ and anti-ϕ orientations. The anti-ψ orientation is also favoured over the anti-ϕ orientation due to lower steric interactions between the residues.

2.1.3 Oligosaccharides and polysaccharides

Oligo- and polysaccharides are carbohydrate molecules composed of multiple monosaccharide units, linked by glycosidic bonds. Oligosaccharides consist of a small number of monosaccharides (typically 3 to 10), while polysaccharides are larger, with up to thousands of monosaccharides. Both oligo- and polysaccharides can be either be unbranched, where each monosaccharide residue only contains a single glycosidic linkage, or branched, where at least one monosaccharide residue contains two or more glycosidic linkages.

Oligo- and polysaccharide conformation is strongly dependent on the primary chemical structure.[8,17–24,26,73,74] Significant differences in molecular conformation can be caused by a seemingly small alteration in primary structure, either by the addition or change in configuration of a substituent or a change in linkage position.[8] For a particular glycan, conformation is determined principally by the orientations of the glycosidic linkages connecting the constituent

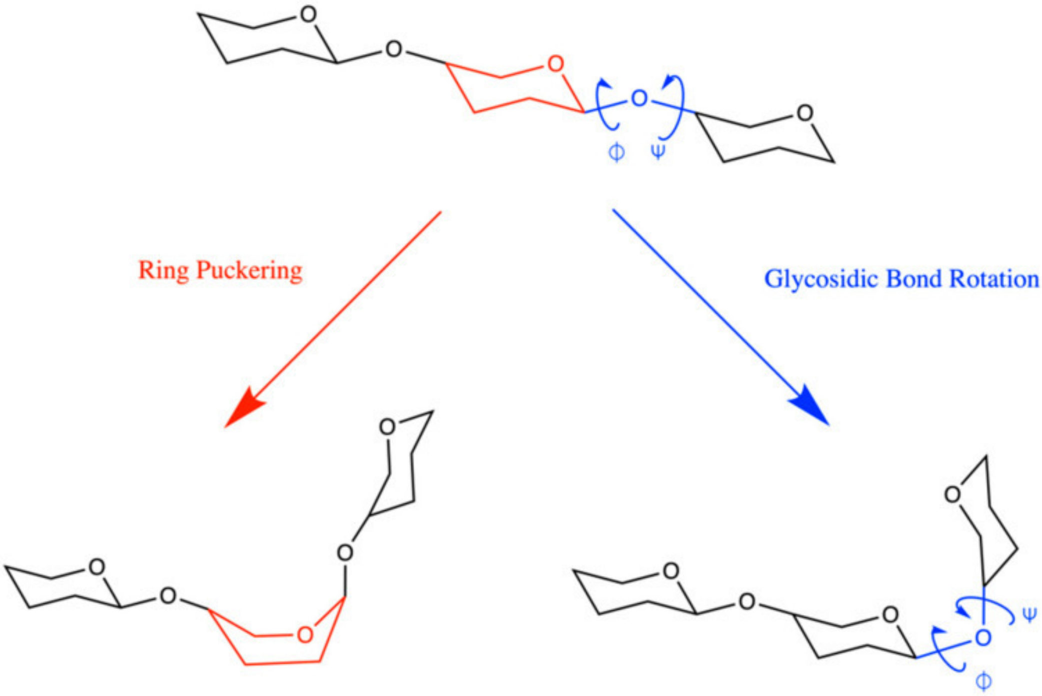


Figure 2.5: Pyranose ring puckering (red) and glycosidic bond rotation (blue) are the major sources of conformational flexibility in polysaccharide antigens. Reproduced with permission from [67] © MDPI.

monosaccharides and the flexibility of the individual monosaccharide rings (Figure 2.5). Syn-orientations of glycosidic linkages often occur in short oligosaccharides leading to helical spirals.[77] However, glycosidic linkage orientations of short chained carbohydrates cannot always be directly extrapolated to long polysaccharide chains, where glycosidic linkage orientations can be more complex.[18] Residues within an oligo- or polysaccharide rarely undergo large changes in ring conformation,[8,10,11] although, subtle variations in pucker angles can affect overall oligo- and polysaccharide conformation.[73,79]

2.2 Lipopolysaccharides

Lipopolysaccharides (LPS) are important components of the outer membrane in gram-negative bacteria, such as *Shigella*. They have a wide range of functions including cellular adhesion, defence, signalling, and structural integrity. They also often mediate cellular interactions and play a key role in bacterial virulence. A single LPS molecule comprises three covalently linked but distinct domains: a Lipid A tail, an oligosaccharide core, and an O-Ag polysaccharide (Figure 2.6).

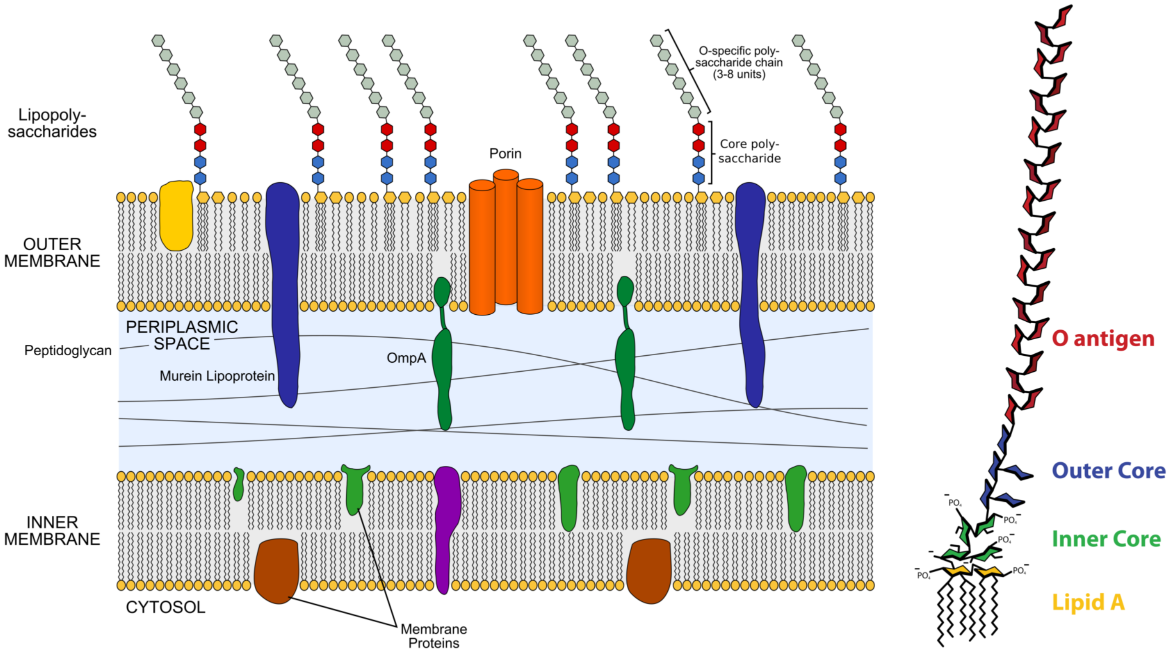


Figure 2.6: (a) Cross-section of a gram-negative cell envelope depicting LPS molecules on the extracellular surface of the outer membrane.[82] (b) Structural illustration of a lipopolysaccharide molecule illustrating its three component parts.[83]

Lipid A is the innermost portion of LPS and anchors the molecule to the bacterial outer membrane. It is a potent endotoxin and can trigger a strong immune response. The core oligosaccharide connects Lipid A to the O-Ag and varies in structure among different bacterial species. It provides structural stability to LPS and contributes to bacterial cell integrity. The O-Ag is the outermost component of LPS and is a polysaccharide built from repeating unit fragments. It is highly variable among bacterial strains and plays a crucial role in determining bacterial serotypes and antigenic variation. The O-Ag can act as a virulence factor and is often the target of the immune response against bacteria.[80,81]

LPS molecules are versatile and adaptable due to its structural heterogeneity, which is primarily found in the O-Ag.[84] O-Ags are commonly around 150-200 residues in length, range in weight from 10 to 20 kDa and extend approximately 100-400 nm from the cell surface.[2,7,85–87] The O-Ag primary chemical structure has a high degree of diversity. A repeating unit can be two to several monosaccharide residues in length and can include pentoses, hexoses, hexosamines, their 6-deoxy derivatives, and hexuronic acids among others.[7] Further, the glycosidic linkage between any two monosaccharides within a repeating unit can attach to the at any of the ring hydroxyl groups, have either α- or β- anomeric configurations, and can contain substituents, such as O-Ac, N-Ac, glycosyl, and phosphate groups.

2.3 *Shigella flexneri* O-antigens

Shigella is a genus of four gram-negative bacterial species that causes shigellosis. There are four serogroups of *Shigella* and approximately fifty serotypes based on O-Ag composition.[51,88] The disease prevalence of each subgroup is different: *Shigella flexneri* (66%), *Shigella sonnei* (24%), *Shigella dysenteriae*, (5%), and *Shigella boydii* (5%).[89] *S. flexneri* is predominant in Africa, Asia, and South America while *S. sonnei* occurs more frequently in Australasia, North America, and Europe. Approximately 90% of all infection from *S. flexneri* is caused by five serotypes: 1b, 2a, 2b, 3a, and 6; and these are therefore of particular interest as vaccine targets.[42,89]

There are nine serotypes of *S. flexneri* and over thirty subtypes as classified by O-Ag specificity (Figure 2.7).[51,88] *S. flexneri* can be classified into serogroups based on the type O-factor groups (I; II; III; IV; V; VI; VII). Further classification into subtypes is based on group O-factors (3, 4; 6; 7, 8; 9; 10; IV-1) as determined by phage-mediated glucosylation, phosphorylation, and O-acetylation of the O-Ag.[51,90]

Investigation into the structures of *S. flexneri* O-Ags began in the 1960s. By 1988, the fundamental carbohydrate primary structures of all known *S. flexneri* O-Ags had been established with a single O-acetylation site at position 2 of α-L-Rha$^{\mathrm{I}}$ identified in types 1b, 3a, 3b, 3c, and 4b. More recent surveys of the *S. flexneri* O-Ag structure have listed O-Ag variants which include glucosylation, additional sites of O-acetylation, and phosphorylation with phosphoethanolamine.

O-acetylation of the O-Ag may be full or partial.[51] The level and position of O-acetylation measured can be affected by the bacterial strain, the method of isolation and purification, as well as storage conditions.[51] The degree of glucosylation at each position is close to stoichiometric.[51] However, the first O-unit of the O-Ag chain linked to the LPS core lacks any glucosyl residue.[51,91] The common backbone across the majority of O-Ags, as well as the shared substituent groups, suggest *S. flexneri* serotypes may be susceptible to a significant level of vaccine cross-protection.[42,89,92,93]

α-L-Rha*p*$^{\mathrm{I}}$ is stoichiometrically O-acetylated at position 2 in serotypes 3a, 3b, 3c, and 4b,[90] whereas it is 80% O-acetylated in serotypes 1b and 7b.[53,90] O-Ac also occurs non-stoichiometrically at positions 3 and 4 of α-L-Rha*p*$^{\mathrm{III}}$ in serotypes 1a, 1b, 2a, and 5a, and at approximately 60% frequency at position 6 of β-D-Glc*p*NAc in type 2a. Similar O-Ac patterns as α-L-Rha*p*$^{\mathrm{III}}$ have been demonstrated in types Y, 6, and 6a, as well as for β-D-Glc*p*NAc in types 3a and Y. The extent of O-Ac of α-L-Rha*p*$^{\mathrm{III}}$ at positions 3 and 4 varies between strains of the same type in the range of 30-70% and 15-30%, respectively, in types 1a, 1b, 2a, 3a, 5a, and Y. Additionally, the same range of variation is observed between types 6 and 6a.[90]

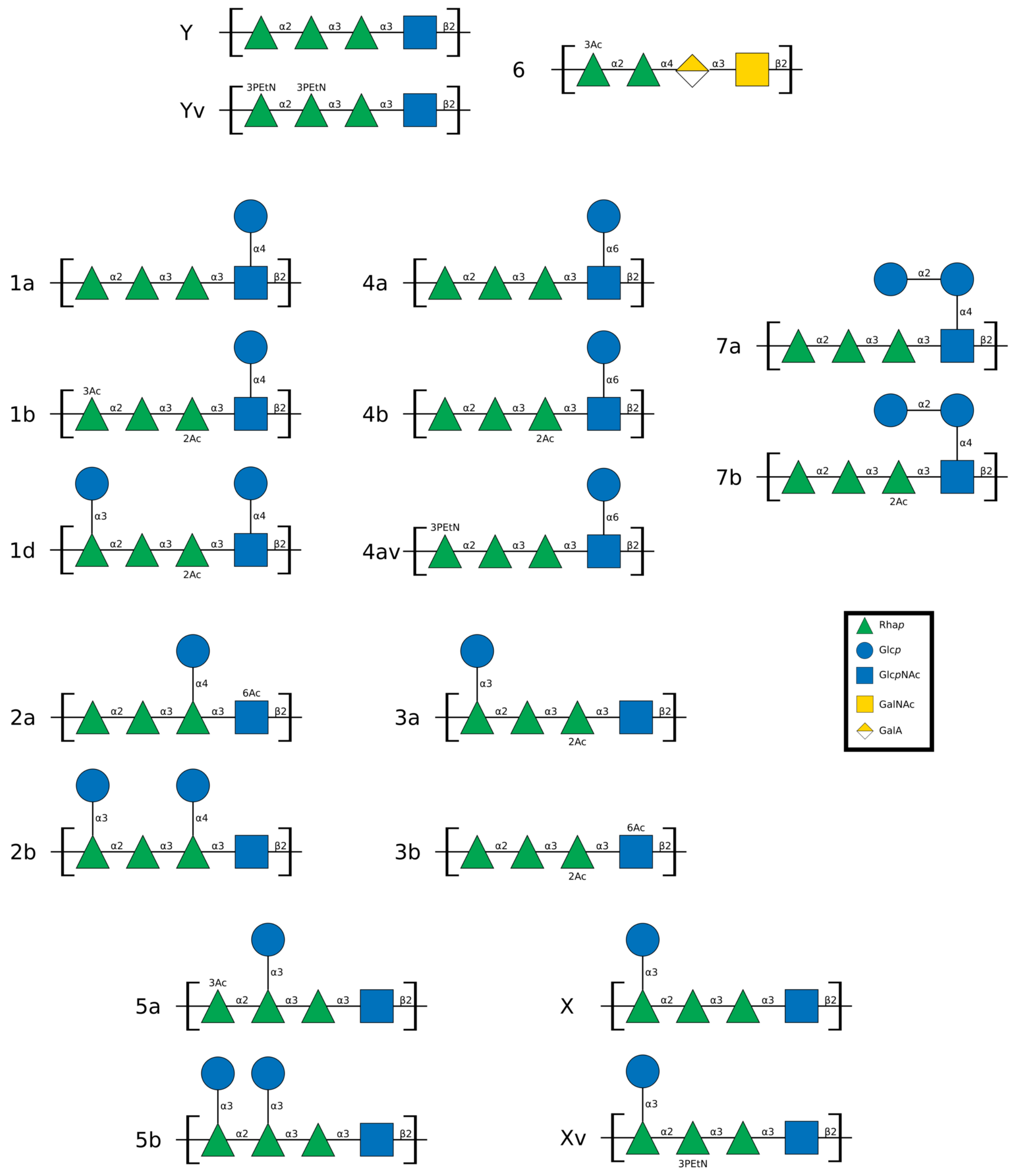


Figure 2.7: SNFG representations of common *Shigella flexneri* O-Ag repeating units. All O-Ag repeating units, with the exception of serotype 6, share the common *S. flexneri* Y O-Ag backbone.

The function of O-Ac in *S. flexneri* infection is not well established. However, glucosylation of the O-Ag is known to have significant impact on the virulence of *S. flexneri* by changing the polysaccharide conformation from a more filamentous to a more compact structure that facilitates bacterial invasion of gut epithelium.[94]

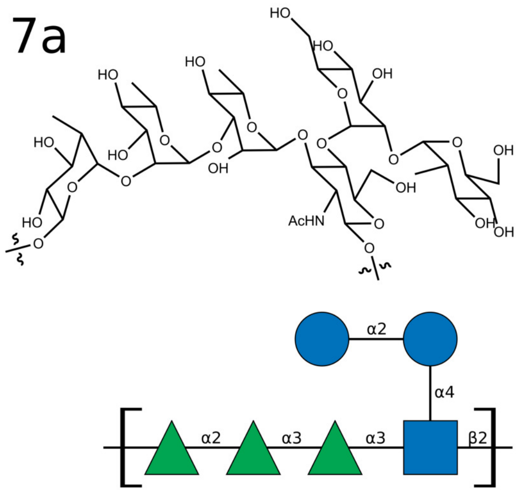
The role of O-factors in antibody binding and immune response is important, although direct correlation between O-factor binding and vaccine cross-reactivity cannot always be made.[8] Shared antigenic primary structures may also not always lead to cross-reactivity among closely related antigens due to subtle differences in carbohydrate structure that can profoundly influence chain conformation and molecular flexibility. As a result, investigating vaccine cross-reactivity between antigens can be informed by conformational analysis.

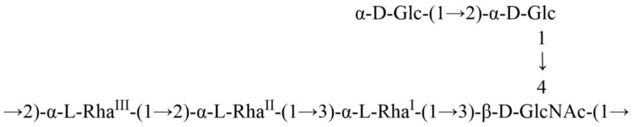
It is expected that a multivalent vaccine that includes *S. flexneri* serotypes 2a and 3a, which are far more widespread, would provide shared group antigens and potentially elicit cross-protection against subtypes 2b, 3b and serotype 7b - the latter via group O-factor 6.[32,42,49] However, serotype 7a lacks any shared group antigens as it only contains group O-factor IC.[51] It has been suggested that excluding *S. flexneri* 7a from a multivalent vaccine would have little impact on the global breadth of coverage.[49] Nonetheless serotype 7 remains of interest as a potential emerging strain as well as part of a comprehensive study of all *S. flexneri* O-Ag conformations.

Serotype 7a was discovered in clinical isolates from Bangladesh and provisionally named serotype 1c.[52] However, as wild-type strains of this serotype do not react with antibodies against O-factor I it was subsequently renamed provisionally to serotype Y394. In 1999, Trevejo et al. discovered an untypeable *S. flexneri* strain from isolates of an outbreak in California.[95]. This strain was provisionally named serotype 88-893. Foster et al. later performed discovered structural similarities between provisional serotypes Y394 and 88-893.[53] Consequently, a new serotype, named serotype 7, was created and provisional serotypes Y394 and 88-893 were renamed to serotypes 7a and 7b respectively.[53]

The Global Enteric Multicenter Study (GEMS) suggests that serotype 7 is accountable for a very small fraction of *Shigella* disease prevalence.[32,42,49] Isolates collected from afflicted patients in Bangladesh, Egypt, Indonesia, Pakistan and Vietnam over the period of study indicated that serotype 7 was accountable for less than 5% of all *S. flexneri* detected in these regions.[89,96–103] However, serotype 7, as with all serotypes of low disease prevalence, remains of interest – particularly to know if they are affected by existing treatments or if they may become potential emerging strains in the future.

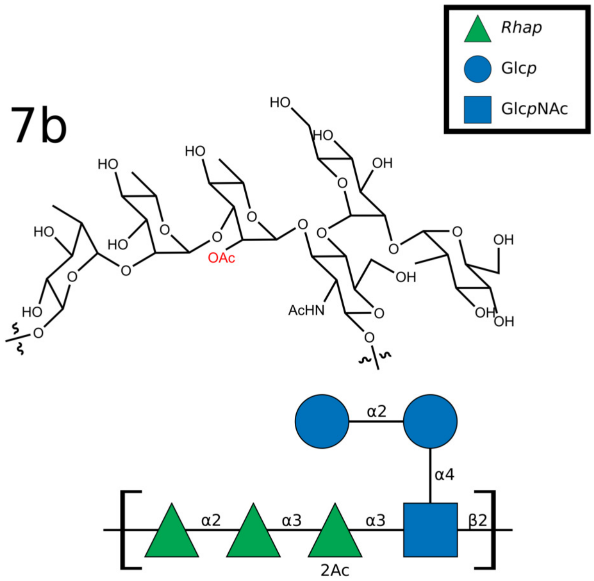
7a

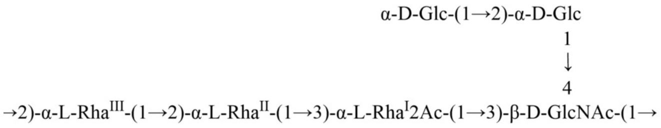




$$\begin{array}{r} \alpha\text{-D-Glc-}(1\rightarrow2)\text{-}\alpha\text{-D-Glc} \\ 1 \\ \downarrow \\ 4 \\ \rightarrow2)\text{-}\alpha\text{-L-Rha}^{III}\text{-}(1\rightarrow2)\text{-}\alpha\text{-L-Rha}^{II}\text{-}(1\rightarrow3)\text{-}\alpha\text{-L-Rha}^{I}\text{-}(1\rightarrow3)\text{-}\beta\text{-D-GlcNAc-}(1\rightarrow \end{array}$$

7b





$$\begin{array}{r} \alpha\text{-D-Glc-}(1\rightarrow2)\text{-}\alpha\text{-D-Glc} \\ 1 \\ \downarrow \\ 4 \\ \rightarrow2)\text{-}\alpha\text{-L-Rha}^{III}\text{-}(1\rightarrow2)\text{-}\alpha\text{-L-Rha}^{II}\text{-}(1\rightarrow3)\text{-}\alpha\text{-L-Rha}^{I}\text{2Ac-}(1\rightarrow3)\text{-}\beta\text{-D-GlcNAc-}(1\rightarrow \end{array}$$

Figure 2.8: Structures of the biological repeating units of *S. flexneri* 7a and 7b O-Ags. Each structure is represented in schematic chemical representation (top), SNFG notation (middle) and standard nomenclature (bottom).

The O-Ag backbone for both serotype 7a and 7b is the same as *S. flexneri* Y. However, the two serotypes also include an α-D-Glcp^{II}-(1→2)-α-D-Glcp^{I} side-chain connected to the 4-position of β-D-Glc$p$NAc, which forms part of O-factor IC. The *S. flexneri* 7b O-Ag repeating unit also has an O-acetyl group at position 2 of α-L-Rhap^{I} that comprises O-factor 6. (Figure 2.8). *Sf* 7a therefore contains a single immunodeterminant O-factor while *Sf* 7b has two O-factors (Figure A.1). These O-factors are found in serotypes 1, 3, and 4, while they also share structural similarities with other *S. flexneri* O-factors. Together with the O-Ag backbone these immunodeterminants play an important role in bacterial infection. Understanding the molecular conformations of *Sf* 7a and 7b backbone and O-factors is of interest and methods such as NMR spectroscopy and molecular modelling can be useful in this regard.

Chapter 3. NMR Spectroscopy of Polysaccharides

NMR spectroscopy is the principle method for determining the primary structures and conformation of polysaccharides. Carbohydrates do not crystallize readily, making conformational analyses by widely used X-ray crystallography techniques challenging.[14] In contrast, NMR spectroscopy allows for the study of molecules in both solid and solution states, enabling analysis of primary structure and conformation.[13] NMR provides a convenient and non-destructive method to probe the structure of biomolecules in general and carbohydrates in particular.[65,104] However, as with all experimental methods, NMR has a number of limitations. For example NMR cannot resolve atomic resolution as X-ray crystallography[105] and also certain molecular processes are too fast or too slow on the NMR time scale, and cannot be measured using NMR spectroscopy techniques.[13] Nevertheless, NMR remains a valuable tool for studying polysaccharides.

3.1 NMR spectroscopy

We provide here an overview of NMR spectroscopy, outlining the concepts of chemical shifts and their temperature dependence, coupling constants and several NMR spectroscopy methods as used in this study, such as diffusion-filtered NMR, zero quantum suppression, and nOe spectroscopy.

In an NMR spectroscopy experiment a molecular sample is placed within a strong, constant magnetic field, which causes spin active nuclei (such as those of ^{1}H and ^{13}C) to align in the direction of the magnetic field. Pulses of radio waves at a range of frequencies are applied to the sample and, depending on the frequency and the electronic environment of the nuclei, the nuclei are perturbed away from their aligned position.[106] The protons relax and spin align with the magnetic field after the radio pulse. In doing so they emit specific frequencies that are characteristic of their electronic environment. These frequencies can be measured to unambiguously determine the relative position of atoms within a molecule and therefore provide information on the structure, dynamics, and interactions of the molecule itself. NMR exploits the magnetic properties of diamagnetic atoms and therefore spectroscopy that measure the frequencies of hydrogen (^{1}H NMR) and carbon-13 (^{13}C NMR) are the most common, although, other atoms such as ^{15}N, ^{19}F and ^{31}P can also be detected by NMR spectroscopy.

3.1.1 ^{1}H NMR chemical shifts

The ^{1}H NMR chemical shift is a fundamental aspect of NMR spectroscopy that provides crucial information about the chemical environment of hydrogen atoms in a molecule. The value of the chemical shift is the difference between the resonant (Lamour) frequency of the spinning protons and the signal of the reference molecule usually tetramethylsilane (TMS). The shift is influenced by the electronic environment surrounding a proton, including factors such as electronegativity, neighboring atoms, and molecular structure. Distinctive chemical shifts for different proton environments enable the identification of functional groups and aid in determining the overall molecular structure. The higher the chemical shift value, the more shielded the proton, while lower chemical shift values indicate a less shielded or more electronegative environment. ^{1}H NMR chemical shift analysis is a powerful tool in elucidating molecular structures and is widely utilized in fields such as organic and medicinal chemistry.

3.1.2 Coupling constants in ^{1}H NMR spectroscopy

Coupling constants in proton NMR spectroscopy provide valuable insights into the connectivity and arrangement of atoms in a molecule. Scalar coupling constants (J-couplings), both homonuclear ($J_{H,H}$ and $J_{C,C}$) and heteronuclear ($J_{C,H}$), are mediated through bonds. The magnetic interaction between coupled protons can be measured and these constants quantify the splitting patterns observed in NMR spectra. The number of peaks in a multiplet, determined by the coupling constants, reveals the neighboring protons influencing each an atoms resonance frequencies. Analysis of coupling constants aids in identifying molecular fragments, determining stereochemistry, and structure of organic compounds. The magnitude of a coupling constant reflects the extent of spin-spin coupling, with larger values indicative of stronger coupling. The magnitude of the coupling constant is also related to geometry by the Karplus type relationship:[107]

$$J(\theta) = Acos^{2}\theta - Bcos\theta + C \tag{3.1}$$

where A, B and C are constants and θ is the torsion angle. A single value of the coupling constant corresponds to several torsion angles (Figure 3.1), which means that the torsion angle cannot be unambiguously determined by measuring a single coupling constant. Analysis of

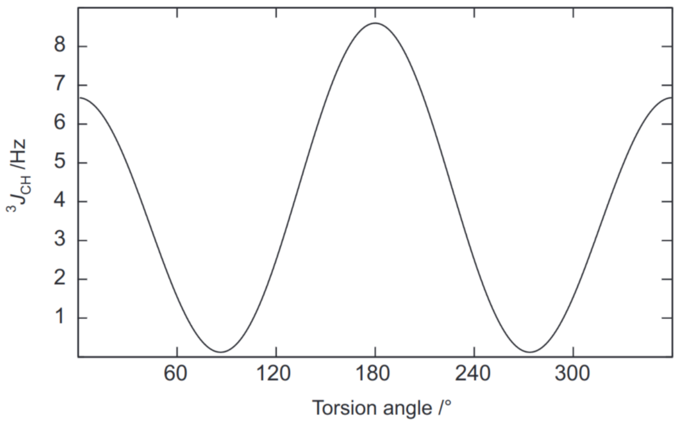


Figure 3.1: Coupling constant as a function of the torsion angle /° for $^3J_{COCH}$.[108]

coupling constants aids in identifying molecular fragments, determining stereochemistry and structure of organic compounds.

3.1.3 Temperature dependence of ^{1}H NMR chemical shifts

^{1}H NMR chemical shifts are temperature dependant. Typically, changes in temperature will cause a molecule to tumble at a faster (for increased temperature) or slower rate (for reduced temperature) but will not significantly affect the position of the measured chemical shifts. However, in certain circumstances the chemical shifts of certain protons are temperature dependent.[109]

Higher temperature may cause conformational changes by overcoming bond rotation barriers and change an atoms electronic environment. Hydrogen bonds may also be weakened, lessening the electron withdrawing effect of the hydrogen bond acceptor on the proton. As a result the proton becomes more shielded and its chemical shift decreases. The temperature dependence on glycosidic proton chemical shift, and consequently, the molecular conformational changes with temperature, has been observed in human milk oligosaccharides[110], type III Group B Streptococcus CPS[111], and *Streptococcus pneumoniae* type 4.[112]

3.1.4 Diffusion-filtered ^{1}H NMR

Diffusion filtered ^{1}H NMR experiments can be used to remove signals of solvents or other impurities with lower molecular weights than the compound of interest. With respect to solvent suppression, this method allows for the selective removal of the solvent peak without affecting resonances of high molecular weight compounds that overlap it, as is the case with pre-saturation techniques.

This diffusion filtered technique uses the same principle as that of diffusion-ordered spectroscopy. Signals of different compounds are separated by their translational diffusion coefficients and all resonances belonging to faster moving molecules are removed.

3.1.5 Zero quantum suppression in NMR spectra

In two-dimensional experiments the inability to separate z-magnetization and zero quantum coherence can cause numerous difficulties. For example, in nuclear Overhauser effect spectroscopy (NOESY) cross-peaks evolve due to z-magnetization during the mixing time* giving in-phase peaks on the spectrum. However, in coupled two-spin systems anti-phase terms will also develop during t_1 and some of these will be turned into zero-quantum coherence by the 90° pulse at the end of t_1.

By applying a systematically calibrated pulse gradient, known as the z-filter,[113] post t_1, the peaks that arise from zero-quantum coherence can be fully eliminated from an NMR spectra. Figure 3.2 illustrates the effect of zero-quantum coherence and its ability to swamp an in-phase resonance of a NOESY cross-peak.

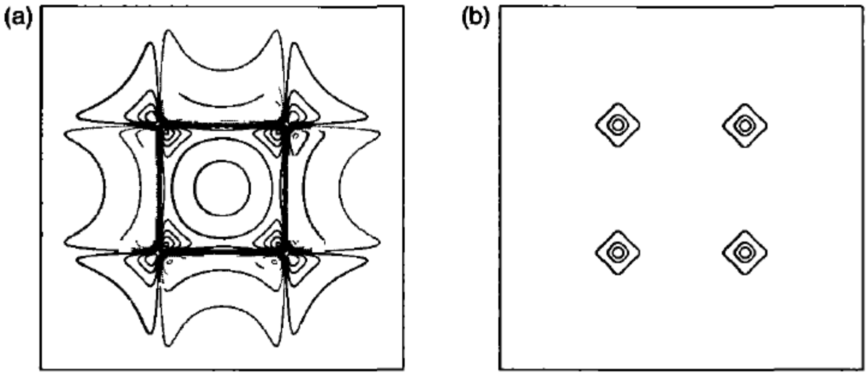


Figure 3.2: Example of NOESY cross-peaks of dipolar coupled spins (a) without zero-quantum suppression showing the effect of zero-quantum coherence and its ability to swamp an in-phase resonance of a NOESY cross-peak; and (b) with zero quantum suppression showing well defined NOESY cross-peaks. Reproduced with permission from [106] © John Wiley & Sons, Inc.

3.1.6 The nuclear Overhauser effect

The nuclear Overhauser effect (nOe) arises from the mutual dipolar relaxation of two or more protons. In large molecules, such as proteins and polysaccharides, there are typically many protons that are dipolar coupled to at least one other proton. In a NOESY experiment this coupling will give rise to a peak in the spectrum if the protons are within approximately 5 Å.[114,115]

---

*the duration during which magnetization transfer occurs between different nuclear spins.

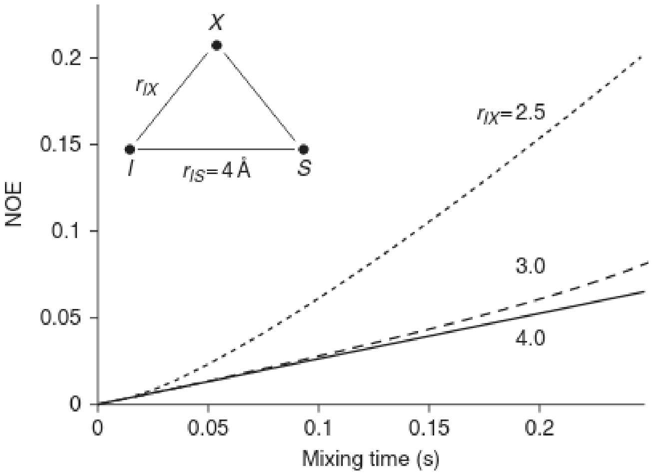


Figure 3.3: The effect of spin diffusion on nOe derived distances. Protons I and S are 4 Å apart and there is a third spin X equidistant between I and S. The I-S NOESY cross-peak build-up is shown for different distances of I-X. Spin diffusion has a marked effect on the distance measured between I-S when a third spin is present between the proton pair. The nOe build-up between I-S is faster and the initial rate approximation is valid for shorter mixing times with shorter distances of I-X. Reproduced with permission from [114] © Elsevier.

A NOESY cross-peak is not quantitative with respect to interproton distances. At best it can be described as being semi-quantitative.[115] A cross-peak suggests that the protons may be within 5 Å. However, under certain assumptions, such as the Isolated Spin Pair Approximation (ISPA) and the initial rate approximation, the cross-relaxation rate between a pair of protons is inversely proportional to the sixth power of the distance between them.[115]

Arguably the largest contributor to breakdown of the Isolated Spin Pair Approximation is the effect of spin diffusion (Figure 3.3).[114] In multi-spin systems where several or more proton are within 5 Å of each another there will be mutual dipolar relaxation between the spins. This can lead to spin diffusion contributing to the nOe enhancement and as a result one would measure a distance between proton pairs shorter than they are in reality.

The sign and magnitude of the nOe depends on the rotational correlation time (Figure 3.4), which is a function of the size of the molecule, viscosity and temperature among other factors.[114] Typically, the ^{1}H-^{1}H nOe is small and positive for small molecules and large and negative for large molecules. Selecting appropriate experimental parameters plays a large role in determining the quality of the spectra recorded and the measurements obtained. For example, a temperature at higher experiment may provide better peak resolution, but it also decreases the correlation time of the molecule, leading to a faster decay in the nOe transfer reducing the intensity of the peaks for a given mixing time.

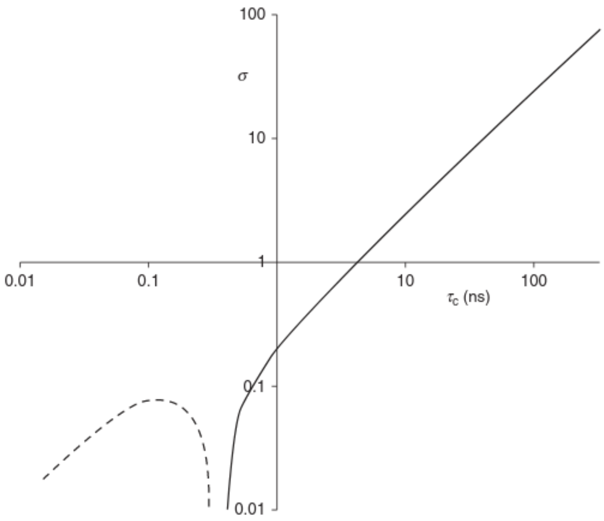


Figure 3.4: Variation of cross-relaxation rate with correlation time. O-Ags have high molecular weights and, therefore, longer correlation times, faster cross-relaxation, and faster nOe build-up. The nOe can thus be measured with shorter mixing times, which is better as it minimizes the effect of spin diffusion. Reproduced with permission from [114] © Elsevier.

Conformational analysis using nOe build-up curves

The cross-relaxation rate, σ, for a dipolar coupled proton pair (IS) can be determined from the following relation:

$$\eta_{IS}(\tau_m) = 2\sigma\tau_m \tag{3.2}$$

where η is the nOe intensity at mixing time, τ_m. There are three general approaches for calculating the cross-relaxation rate base on Equation 3.2: the Classical method, the PANIC (Peak Amplitude Normalization For Improved Cross-Relaxation)[116,117] method and the Dixon[118] method.

The Classical method has two different forms. The first and most basic form of the Classical method takes the form of Equation 3.2; the second involves normalisation of the cross-peak intensity by the auto-peak intensities (η^{ap}) at zero mixing time as described in Equation 3.3.

$$\frac{\eta_{IS}(\tau_m)}{\frac{1}{2}\eta_I^{ap}(0) + \frac{1}{2}\eta_S^{ap}(0)} = 2\sigma\tau_m \tag{3.3}$$

The normalised classical method provides an improvement over the Classical method. Normalising by the auto-peak intensity and zero mixing time reduces the experimental differences, such as shimming, between the NOESY spectra run at different mixing times. A further improvement over the Normalised Classical method is the PANIC method (Equation 3.4) whereby the cross-peak intensity is normalised with respect to the auto-peak at that mixing time.

$$\frac{\eta_{IS}(\tau_m)}{\frac{1}{2}\eta_I^{ap}(\tau_m) + \frac{1}{2}\eta_S^{ap}(\tau_m)} = 2\sigma\tau_m \tag{3.4}$$

The PANIC method largely restores the linear dependence of the nOe enhancement on τ_m typically allowing for longer mixing times for which the initial rate approximation remains valid. The Dixon method (Equation 3.5) is simply an extension of the PANIC method whereby the cross-relaxation rate is calculated as a function of mixing time.

$$\frac{\eta_{IS}(\tau_m)}{\eta^{ap}(\tau_m)\tau_m} = 2\sigma\tau_m \tag{3.5}$$

Given the calculated cross-relaxation rate for a cross-peak for proton pair IS, σ_{IS}, and reference peak, σ_{ref}, with a known interproton distance, r_{ref}, we can calculate the distance between IS assuming the ISPA is valid (Equation 3.6).

$$r_{IS} = r_{ref}(\sigma_{ref}/\sigma_{IS})^{1/6} \tag{3.6}$$

Equation 3.6 is typically correct for short and middle internuclear distances. For longer distances, spin diffusion may occur whereby magnetization is transferred between two nuclei via a third.

3.2 Polysaccharide structure analysis by NMR spectroscopy

Structure elucidation and conformational analysis of polysaccharide antigens by NMR spectroscopy typically employ a multi-step process. Firstly, the polysaccharide antigen are extracted from the bacterial cell and then purified. Subsequently, NMR spectroscopic analysis is carried out to determine the monosaccharide residues present in the carbohydrate antigen, and their linkage positions are identified. Finally, based on heteronuclear $^3J_{C,H}$ coupling and ^1H-^1H distance measurements observed, representative three-dimensional structures can be constructed.

3.2.1 Primary structure determination

NMR spectroscopy is a leading method for elucidating the primary structure of polysaccharides.[13] Solution-state NMR spectroscopy requires homogeneous and not overly viscous material.[13] Achieving this requires the use of several physicochemical separation techniques.[13,104] In the case of LPS, it has been customary to analyse the polysaccharide portion by removing the lipid component through a mild acid treatment. However, this technique may result in undesired alterations and degradation, necessitating alternative approaches to address these issues.[12,119] Nonetheless, in some instances, it is feasible to analyse the complete LPS molecule, including its polysaccharide component as has been done with *Escherichia coli* O124 and *S. dysenteriae* type 3 LPS.[120]

NMR spectroscopy of polysaccharide usually begins with recording 1D ^{1}H and ^{13}C NMR spectra or a 2D ^{1}H, ^{13}C HSQC NMR spectrum.[13,121,122] The number of constituent monosaccharides can be obtained from anomeric resonances, typically between 4.4–5.5 ppm for ^{1}H NMR and 90–110 ppm for ^{13}C NMR spectra. Each spin system and the resonance of ring proton and carbon atoms can be assigned from 2D techniques such as ^{1}H, ^{1}H COSY; ^{1}H, ^{1}H TOCSY; and ^{1}H, ^{13}C HSQC experiments. Nuclear Overhauser effect spectroscopy (NOESY), for through space interactions, and ^{1}H, ^{13}C HMBC experiments, for long-range chemical shift correlation, can be used when coupling constants are small and magnetization is not transferred around the whole spin system.

After assigning all the ^{1}H and ^{13}C resonances within the ring, the linkage positions can be investigated. Glycosylation shifts can be obtained by comparing the chemical shifts of the polysaccharide to those of corresponding monosaccharides and compensating for the effects caused by glycosylation.[13] Inter-residue ^{1}H, ^{1}H NOE and ^{1}H, ^{13}C HMBC correlations can be employed to find the sequence of sugar residue in the chain.

Computerised methods can assist in polysaccharide structure determination. CASPER (Computer-Assisted SPectrum Evaluation of Regular polysaccharides) calculates chemical shifts based on constituent monosaccharides within oligo- and polysaccharides as well as glycosylation shifts from substitutions.[123–125] CASPER determines ^{1}H and ^{13}C chemical shifts from known structures allowing for comparison and validation of experimental values. For unknown structures, CASPER can determine the carbohydrate sequence by finding structures with the smallest difference between experimental and calculated chemical shifts.

The primary structures of polysaccharide antigens for a wide range of bacteria have been resolved by NMR spectroscopy. These include, but are not limited to, approximately 79 *Klebsiella* capsular polysaccharides (CPS), 176 *Escherichia coli* O-Ags[126], 98 *Streptococcus pneumoniae*

CPSs,[127] and 30 *Shigella flexneri* O-Ags.[7] These studies have been able to identify small differences in linkage positions, O-Ac patterns, chemical substitutions, and the presence of unusual residues with methods that have been greatly simplified by NMR spectroscopy.

3.2.2 Conformational analysis

NMR can also be used to provide insight into molecular conformation. Such analysis usually measures transglycosidic heteronuclear coupling constants, residual dipolar coupling constants and nOe ^{1}H-^{1}H distances.[13,121]

Carbohydrate conformation can be investigated through experimental measurements, such as ^{1}H-^{1}H distances derived from nOe spectroscopy, or ^{3}J couplings. However, such comparisons are not always directly possible or, if so, may be ambiguous[8,13,128] For example, glycosidic linkage torsion angles can be inferred indirectly through heteronuclear $^3J_{C,H}$ coupling measurements but these only provide a weighted average of all torsion angles populated.[13,129,130] Further, ^{1}H-^{1}H distances can be used to determine the orientations of glycosidic linkages[13,131] and measurement of inter-residue ^{1}H-^{1}H distances can help identify possible molecular conformations.[8] Measurements from NOESY spectra can also help reveal the presence of chair or non-chair pucker states.[128] However, conformations of non-chair pucker states in pyranose and furanose rings in solution cannot be easily established with current experimental approaches.[128]

The most common method carbohydrate conformational analysis by NMR employs the measurements of ^{1}H-^{1}H distances using NOESY experiments.[13] However, these are restricted to interproton distances that are $\sim$5 Å or less, as beyond this distance nOe correlations are too weak to measure accurately.[114] Additionally, polysaccharide spectra are generally not well defined, even when compared to other biomolecules such as nucleic acids and proteins, and may have many peaks overlapping, making it challenging to take accurate measurements with high degree of confidence.[132,133]

Although NMR spectroscopy has been useful in determining the primary structures of many polysaccharide antigens, the number of antigens with resolved conformations is relatively small. Early studies involved the use of through space interactions to measure atoms which non-bonded atoms may be near to one another allowing for conformations to be inferred. For example, nOe and Rotating Frame Overhauser Enhancement (ROE) spectroscopy has been used to measure ^{1}H-^{1}H distances in *Moraxella catarrhalis* and *Escherichia coli* O-Ags and static conformational models have been built based on these measurements.[130,134,135] More recently, heteronuclear $^3J_{C,H}$ coupling and ^{1}H-^{1}H distances derived from nOe spectroscopy has often been combined with other techniques, such as MD simulations, to provide detailed information on both molecular conformation and dynamics, as has been demonstrated with studies of *Escherichia coli* and *Shigella flexneri*.[15,135–139]

Chapter 4. Molecular Dynamics Simulations of Polysaccharides

MD simulations employ the principles of classical mechanics to model the motion of atoms over time as a function of the interatomic forces in a molecular system. These forces are described by a series of functions and parameters that are specific to a particular group of atoms or class of molecules.[140,141] For example, there are different force fields for proteins,[142–146] lipids,[147–149] carbohydrates,[55–58,60,62] and nucleic acids[150–153] as the behavior and interactions of molecules can vary widely based on their chemical composition, environment, and structure.

Computational methods are capable of supplementing experimental measurements as well as providing theoretical physical, chemical and biochemical insights that may otherwise be very challenging or impossible to obtain.[8,10,154–156] The computational modelling of carbohydrates is typically performed with two common methods: QM and/or MD simulations.[8,10,140,154,157–159] QM methods, in comparison to MD, are capable of highly accurate energy, electronic, and structure determination. However, these are typically restricted to relatively small chemical systems.[160,161] MD studies are generally limited to structure calculations and high-level chemical analysis but can be performed on much larger systems, such as those involving polysaccharides.

4.1 The model

A force field is a set of functions that describes the interactions and behavior of atoms in a molecular system. The parameters of a force field are determined in a process known a force field parameterisation, through QM calculations, experimental methods or, ideally, a combination of both.[162] A classical mechanics function and set of force constants together define the potential energy of a molecular system. The potential energy, V_{pot}, of a molecular system with respect to its individual components can be defined by the general function:

$$V_{pot} = \sum_{Bonds} v_{Stretch} + \sum_{Angles} v_{Bend} + \sum_{Dihedrals} v_{Torsion} + \sum_{Urey-Bradley} v_{1,3-interaction} + \sum_{Pairs} v_{Non-Bonded} \quad (4.1)$$

The individual energy components of bonding and external non-bonding interactions between each atom pair i, j are as follows:

Bonding terms

Bonds:

$$v_{Stretch} = k_{stretch}(r - r_0)^2 \quad (4.2)$$

where r represents the distance between the atom pair, r_0 denotes the equilibrium distance for the atom pair, and $k_{stretch}$ refers to the bond force constant.

Angles:

$$v_{Bend} = k_{bend}(\theta - \theta_0)^2 \quad (4.3)$$

where θ represents the angle formed by the three-atom group, θ_0 corresponds to the equilibrium angle of the three-atom group, and k_{bend} denotes the angle force constant.

Dihedrals:

$$v_{Torsion} = k_{proper}[1 + cos(n\phi - \delta)] + k_{improper}(\omega - \omega_0)^2 \quad (4.4)$$

ϕ represents the torsion angle, δ corresponds to the phase shift, k_{proper} denotes the force constant for proper dihedral interactions, and $k_{improper}$ represents the force constant for improper dihedral interactions.

1,3 interactions:

$$v_{1,3-interaction} = k_{UB}(u - u_0)^2 \quad (4.5)$$

where u represents the measured 1,3 atom distance, u_0 denotes the equilibrium 1,3 atom distance within the harmonic potential, and k_{UB} represents the Urey-Bradley force constant.

Non-Bonding term

Nonbonded :

$$v_{Non-Bonded} = \epsilon_0 \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{r_{ij}} \quad (4.6)$$

The non-bonded term in the equation consists of two components: the van der Waals energy, described by the 12-6 Leonard-Jones potential, and the standard Coulombic potential. R represents the distance at which the 12-6 Leonard-Jones potential becomes zero.

4.2 The integrator

In MD simulations, the integrator algorithm calculates the change in atomic position over time based on the forces acting on the atoms. To initiate an MD run, the initial position, connectivity, and velocity of all atoms at the start time are required. In each subsequent time step, the forces exerted on each atom are computed, taking into account bonded and non-bonded interactions, as well as any restraints if applicable. The Velocity-Verlet algorithm,[163] calculates atom positions (r), velocities (v), and accelerations (a) at time $t + \Delta t$ based on their values at time t by:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}t + \Delta t^2 \tag{4.7}$$

$$\mathbf{v}(t + \Delta t/2) = \mathbf{r}(t) + \frac{1}{2}\mathbf{a}t \tag{4.8}$$

$$\mathbf{a}(t + \Delta t) = -\frac{1}{m}\Delta V(r(t + \Delta t)) \tag{4.9}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t) + \frac{1}{2}\Delta t\mathbf{a}(t + \Delta t) \tag{4.10}$$

4.3 Metadynamics

Metadynamics simulations are a type of potential of mean force (PMF) calculation and a subset of MD simulations. While MD and metadynamics employ similar methods, they differ in how they explore system states.

MD is a deterministic approach that models the forces acting on atoms and their subsequent motion over time using classical mechanics. As a result, systems simulated with MD naturally tend to transition towards lower energy states determined by the potential energy functions of the system.[162] However, the range of accessible system states within a given period of time depends on the inherent variations of forces in the system. If a particular state is energetically distant from lower energy states, longer simulation times may be necessary, or lower energy states may be poorly sampled or missed altogether.[162]

In contrast, metadynamics simulations use external forces to accelerate the sampling of system space.[164] These external forces act on specific sets of atoms within the system known as collective variables (CVs), which can include inter-atomic distances, bond angles, and torsion angles.[165] During the simulation, the potential of the system, V_{meta}, is measured as a function of one or more CVs. Positive Gaussian energy potentials are then added, reducing the likelihood of the system returning to its previous state. By running the simulation for a sufficient duration, a statistically representative average of all CV configurations is obtained.[164,165] The free energy obtained through this process, being dependent on one or more CVs, is referred to as the potential of mean force.

In addition to metadynamics other approaches also be employed to enhance conformational sampling. These include, increasing the simulation time-step, replica exchange (temperature or Hamiltonian), and hydrogen mass repartitioning among others. However, it should be noted that these methods also have their limitations: the loss of time-domain information (replica exchange) and loss of the ability to analyse certain chemical and physical properties (increased time-steps, hydrogen mass repartitioning).

4.4 Periodic boundary conditions

MD simulations typical model a small segment of a much larger chemical system. The model system has a significantly greater surface area to volume ratio and the simulated molecules spend a greater proportion of time closer to the outer physical boundary of the system. Consequently, the interactions at the edge parts of the system have a disproportionally large effect on the overall simulation. This results in erroneous behaviour and is termed the ‘edge effect’.

Periodic boundary conditions (PBC) are employed to minimise the impact of this ‘edge effect’. A system is modelled as a unit cell of a much larger theoretical environment (Figure 4.1). During a simulation, the motion of all particles in adjacent unit cells replicate the same movement of the simulated unit cell. When a particle moves out of the simulated unit cell during its motion, a corresponding particle of the same type enters the unit cell at the exact opposite position. This process effectively eliminates the walls within the simulated unit cell, greatly reducing errors that could arise in the abscence of periodic boundaries.

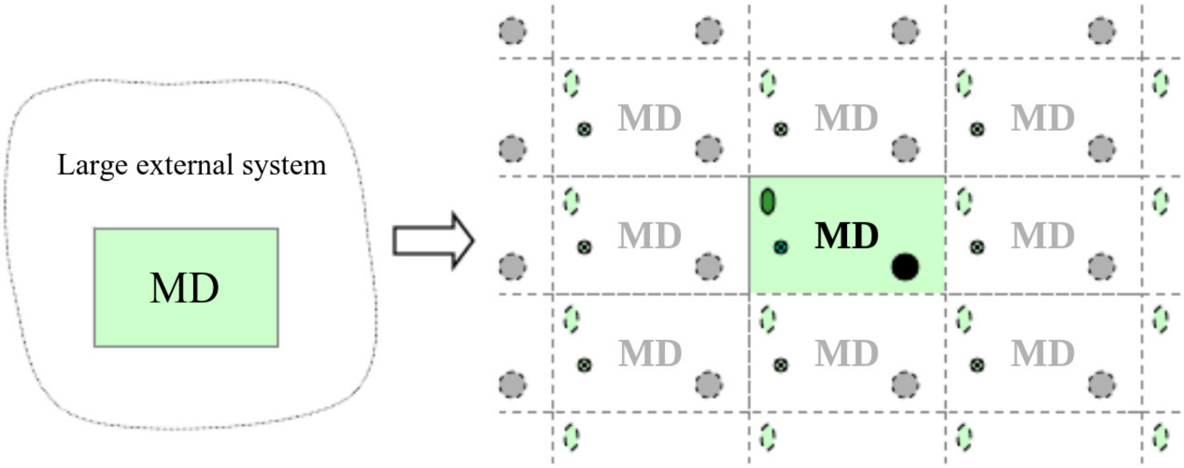


Figure 4.1: Illustration of an MD simulation with (right) and without (left) periodic boundary conditions.[166]

4.5 Water models

Water models for the simulation of water molecules in MD simulations range from a simple two site model[167], that models a water molecule as two single point charges, to more complex six site models,[168–170] which employs a six point charge system. Most carbohydrate force fields were parameterised with TIP3P,[171] a three site model force field. Water models are generally developed to model the properties of pure water and rarely optimized for interactions with solutes, such as proteins, carbohydrates or lipids.[10] Rather, it is the force field of the solute that is usually modified to perform optimally with a given water model.[10] It is not advisable to use a water model different to that employed in the carbohydrate force field parameterisation process.

4.6 Carbohydrate force fields

The development of a force field for carbohydrates is particularly challenging due to the molecules structural complexity and diversity.[172] Moreover, their relative flexibility means that they generally do not adopt a single conformation in solution.[173] Due to this, suitable experimental information to derive potential energy functions remain sparse and many carbohydrate force field parameters are derived from non-carbohydrate molecules that have structurally similar components.[141,172,174] Nonetheless, carbohydrate force fields continue to improve and have advanced to the level where they have been shown to outperform more sophisticated semi-empirical QM methods in certain instances.[175]

There are currently four main all-atom carbohydrate-specific force fields used in molecular modelling: CHARMM36[55–58,60], GLYCAM06[62], GROMOS[176–179], and OPLS-AA[180–182]. The former two force fields are more widely used and can model a wide range of carbohydrates. The latter two are parameterised for a narrower selection of monosaccharides and glycosidic linkages and are more suitable for the simulation of unlinked glycans.[156] These force fields are primarily tailored for biologically and chemically relevant furanoses and pyranoses monomers and their polysaccharides — no larger ring carbohydrates are defined.[141,154] We provide a brief overview of CHARMM36 and GLYCAM06.

4.6.1 CHARMM36

The CHARMM all-atom biomlecular force field[144,183] is among the most widely used general force fields. The CHARMM force field for carbohydrates[55–60] (CHARMM36) is parameterised for the simulation of common furanose and pyranose molecules — such as fucose, glucose, galactose, xylose, N-acetylglucosamine, N-acetylgalactosamine, glucuronic acid, iduronic acid, and N-acetylneuraminic acid among others — allowing for the simulation of molecules from single monomeric units to extended polymer chains. CHARMM36 specifies unique atom types for different anomers which is not a feature of other force fields[154] and reproduces structure, dynamics and thermodynamic properties of numerous important monomers as crystal structures and in aqueous solution.[154,156]

Recent developments of a polarizable empirical force field for hexopyranose based on the CHARMM Drude polarisable force field[184] has increased the modelling accuracy of furanose and pyranose monosaccharides[185–187] and their glycosidic linkages,[184,188,189] by supporting electronic polarization instead of the standard fixed charge approximation.[190] Recent studies on α-linked mannopyranose disaccharides demonstrates the improved utility of Drude versus the additive CHARMM36 force field in investigating complex carbohydrates and demonstrate that Drude polarizability improves "the underlying physical forces dictating the wider conformational sampling dipole moments and radial distribution functions with water of the hydroxyl groups".[191]

4.6.2 GLYCAM06

The GLYCAM06 force field provides a set of complete parameters for the simulation of a wide range of carbohydrates and glycoconjugates.[61,62] Similarly to CHARMM36, it can be used to model numerous standard and non-standard furanose and pyranose monosaccharides as well as glycan chains. The parameter set can also be extended to incorporate glycoproteins, glycolipids,[192,193] lipopolysaccharides,[194] lipids,[195] proteins, and nucleic acids.[196]

A strength of GLYCAM06, compared to other carbohydrate optimised force fields, comes in large part from its very careful treatment of electrostatic interactions. In GLYCAM06, each atom within a sugar unit is assigned a particular partial atomic charge as opposed to being assigned partial charges based on a restricted set of atom types, as is done in many other formulations.[141,154] This feature was achieved by fitting torsion terms to rotational data from small achiral molecular fragments, rather than to tetrahydropyran-based models of intact α- and β-pyranosides as used with CHARMM36. An intrinsic benefit of this feature is that it naturally facilitates an equilibrium between conformers with axial and equatorial substituents at the anomeric centre through the simulation of ring-flipping.

GLYCAM06 does not employ any generic torsion terms (e.g. of the type X–C–C–X). All torsion terms are derived in a hierarchical manner using fitting valence parameters to a series of more than 100 model compounds encompassing several molecular classes. These include hydrocarbons, alcohols, ethers, amides, esters, carboxylates and mixed functional groups, such as ether alcohols, ether amides, alcohol amides and ether carboxylates.[196] Another feature unique to GLYCAM06 is the manner in which atomic partial charges are calculated. MD simulations are used generate an ensemble of conformations employed in the charge averaging.[197,198]

4.6.3 Carbohydrate force field limitations

The strength of current additive all-atom carbohydrate force fields has been demonstrated by many studies across numerous applications. For example, CHARMM36[55–58,60] has been shown to provide molecular ensembles of polysaccharides from MD simulations that are in very good agreement with ^{1}H-^{1}H distances derived from NOESY experiments of the same polysaccharides;[138,139] while GLYCAM06 has been able to accurately model ring pucker in heparonome[79] and amylose polymers[73] Nonetheless, limitations in the force fields become apparent as more experimental data becomes available.[8,61,140,141,158] Comparative modelling between force fields have also revealed significant differences in several measurable parameters and resultant conformational ensembles.[8,154,196,199]

There are a number of limitations with current carbohydrate force fields. The force fields only support common carbohydrate molecules, which are a small subset of all known sugars. However, as force fields are continuously improved, further molecules will be parameterised, increasing the number molecule types that can be simulated by molecular dynamics. Nonetheless, carbohydrates that contain sugar residues or functional groups that are not parameterised cannot be modelled without such updates to existing force field parameter sets. While parameters for undefined molecules and functional groups can, and often are, estimated (without going through a full force field parameterisation process) this can lead to increased errors and decreased accuracy of a simulation.

For disaccharides and longer chains, torsion angles adopted by glycosidic linkages have been shown to be force field dependant. For example, anti-ψ orientations of the glycosidic linkages are less favoured in CHARMM36[78,200] and OPLS-AA[18] compared to GLYCAM06.

Similarly, the accurate parameterisation of carbohydrate ring pucker conformations has proven challenging. This is as the degree to which a force field is able to correctly reproduce the conformation and dynamics, of a sugar ring as well as the interactions between rings, depends on a delicate balance between inter- and intramolecular forces.[201]

In both CHARMM36 and GLYCAM06, ring puckering thermodynamics in solution is not used as target data during the force field parameterisation process. Instead, puckering information obtained from experiment and quantum mechanical gas phase puckering energetics for tetrahydropyran are used during validation, post the parameterisation protocol.[202,203] However, validation can also be challenging. Pyranose ring puckering occurs at the multi-microsecond and beyond timescale,[66,79,175] which is near the upper limit of typical present-day all-atom explicit-solvent MD simulations. Thus, limitations in force field accuracy may not be readily apparent simply based on analysis of such simulation test results.[67]

Recent studies on the ring dynamics of unsubstituted aldohexoses using the CHARMM Drude polarizable force field have demonstrated that, while the inclusion of polarizability increases ring flexibility, the sampling does not agree with experimental data.[204] Development of the CHARMM Drude polarizable force field is on-going and further improvements of the polarizable Drude parameters are required to improve ring pucker dynamics.[191,204]

Development of a robust carbohydrate force field capable of modelling oligosaccharides and polysaccharides is a particularly challenging task. Ideally, during the parameter development and validation processes one needs to account for and consider the influence of the inherent flexibility of glycans. None of the carbohydrate force fields explicitly optimise for oligosaccharides and polysaccharide parameter sets as it is simply not experimentally and computationally feasible to do so. Instead, they are parameterised for mono- or disaccharide units and later validated for longer chain residues by explicit modelling. The accuracy and validity of the modelling is subject to the limitations of molecular modelling techniques and it is also highly dependant on the force field used as well as the methodology employed.

However, studies that specifically compare carbohydrate force fields are rare. Small molecules show comparable results[205,206] while simulations of larger or more complex systems of glycans reveal significant differences in free-energies and molecular conformations and dynamics between the force fields studied.[78,199,207–212] For example, several studies report conformation collapse of polysaccharides that are abundant with rhamnose residues. MD simulations of *S. flexneri* serotype Y O-Ag repeating units by Kang et al. demonstrate that stable hairpin bend conformations are preferred in when modelled with the GLYCAM06 force field, but not when CHARMM36 is used.[16] Galochkina et al. observed similar irreversible conformational collapse when modelling a *Salmonella typhimurium* O-Ag. They report the formation of a hairpin "globule" formation with GLYCAM06, but not with the OPLS-AA force field.[18] Similar results are shown for the *Escherichia coli* serotype O25b O-Ag.[213] Lazar et al. confirm this conformational collapse with the GLYCAM06 force field and demonstrate that it occurs in saccharide chains rich with α-L-rhamnose residues after relatively long simulation intervals.[199] They show that this phenomenon arises due to specific orientation of α-L-Rha$\rightarrow\alpha$-L-Rha glycosidic linkages that have anomalously low energy, which are subsequently sustained by intramolecular interactions in the saccharide chain. Moreover, the lack of partial aliphatic hydrogen charges in the GLYCAM06 are identified as the source of this anomaly.

Chapter 5. Modelling Carbohydrates

Carbohydrates of a wide range of forms, from monosaccharides to polysaccharides, have been studied extensively by molecular modelling.[8,10,155,214] In this chapter we discuss molecular conformations and dynamics from the modelling of small pyranose carbohydrate molecules (with a specific focus on ring pucker) and bacterial polysaccharides.

5.1 Modelling small carbohydrate molecules

Mono-, di-, and trisaccharides have been studied extensively by computational simulations.[154,155,214] Their relatively small physical size and mass means that modelling of many chemical and structural properties do not require extensive resources, while their broad presence in chemical and biochemical systems lend themselves to particular scientific interest. Nonetheless, the high energy interconversion barriers of ring conformers in water for each of the various carbohydrate ring forms as well as the relatively slow transitions, often in the microsecond timescale, means that that long simulation times are required to study the different ring conformations.[214]

α-D-glucopyranose and β-D-glucopyranose are amongst the most intensively explored of all carbohydrate rings, in both vacuum and solution. There are 14 canonical pucker states that minimize the angle strain within the β-D-glucopyranose ring: 2 chair forms ($^4C_1$ and $^1C_4$), 6 possible distinct boat forms ($^{1,4}B$, $^{2,5}B$, $^{3,O}B$, $B_{1,4}$, $B_{2,5}$ and $B_{3,O}$), and 6 skew forms ($^1S_3$, $^1S_5$, $^3S_5$, $^3S_1$, $^5S_1$, $^5S_3$). All free energy studies of β-D-glucopyranose to date[64,215–218] identify the $^4C_1$ chair as the global free energy minimum in aqueous solution, in agreement with empirical evidence.[13,155,219,220]

Mayes et al. used DFT methods to compile the first complete library of low-energy local minima and puckering interconversion transition states for five biologically relevant pyranose sugars: β-xylose, β-mannose, α-glucose, β-glucose, and β-N-acetylglucosamine; and demonstrate that energy require ring pucker transitions increases with greater ring substitution.[64] Lyu et al. mapped the complete conformational energy surface and the corresponding pseudorotation

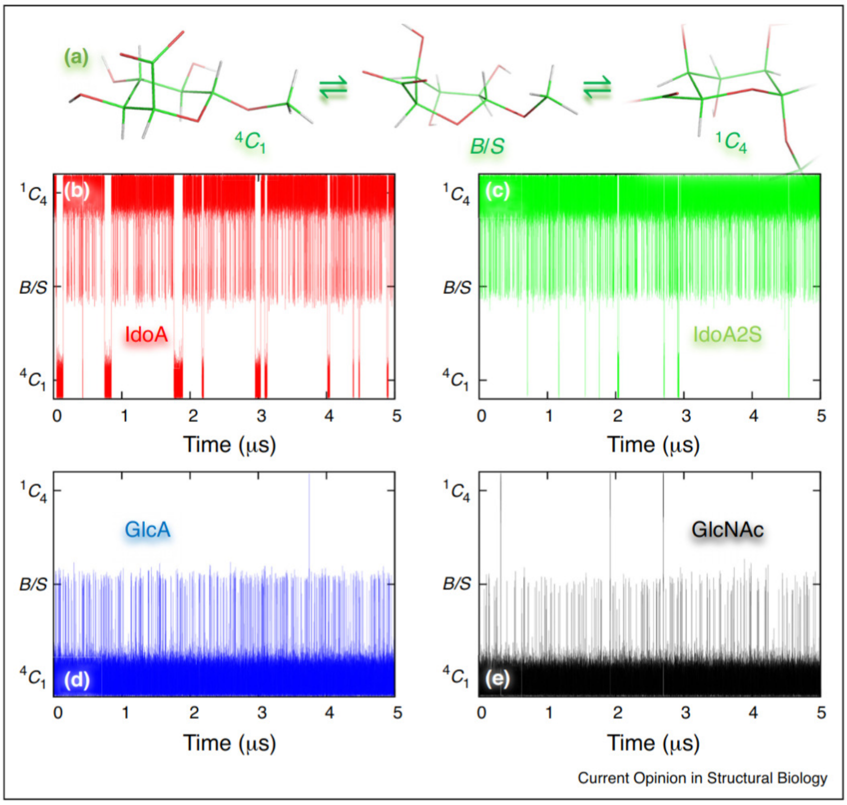


Figure 5.1: Puckering dynamics for key monosaccharides (O-methyl) found in glycosamino glycans from 5 ms MD simulations using GLYCAM06 in water, for (b) iduronicacid(IdoA), (c) iduronic acid-2-O-sulfate (IdoA2S), (d) glucuronic acid (GlcA) and (e) N-acetylglucosamine.[63,175] The dynamics are visualized using the Cremer-Pople azimuthal angle (labelled according to canonical pucker) against time and (a) shows the typical 3D-conformers found in the puckering equilibrium. Reproduced with permission from [155] © NCBI

path of the interplay between deoxyribose ring puckering and intramolecular H-bonding in 2′-deoxycytidine, 2′-deoxyadenosine, 2′-deoxythymidine, and 2′-deoxyguanosine.[221]

A metadynamics study evaluating three carbohydrate force fields (GLYCAM06, GROMOS 45a4 and OPLS) to model different ring conformations of β-D-glucopyranose was performed by Spiwok et al.[217] All three force fields show the $^4C_1$ conformation as the most stable in vacuum and water, with a high barrier for interconversion to other conformations, while GLYCAM06 was found to be consistent with previous metadynamic studies.

Guvench et al. characterised the ring puckering thermodynamics of ten common biological pyranose monosaccharides.[67] They demonstrate that the CHARMM36 force field reliably models ring puckering across the molecules studied while also capturing the subtle variations of the $^4C_1$ and $^1C_4$ states.

Sattelle and Almond[63] investigated ring pucker kinetics of on N-acetylglucosamine monosaccharide and derivatives by multi-microsecond length MD simulations in solution using the GLYCAM06 force field. They showed that, while the $^4C_1$ chair is overwhelmingly the dominant state, the ring conformation is dynamic and puckering does occur to a notable extent (Figure 5.1). For methyl α- and β-D-idopyranosides the $^1C_4$ chair is preferred.

Later MD simulations demonstrated ring flexibility for some monosaccharides over microsecond timescales[66,73,155,222] suggesting that while saccharide ring puckering can and does occur, pyranose rings predominantly remain in low-energy $^4C_1$ (or $^1C_4$ for idopyranose) conformations, with bulky substituents generally maintained in the equatorial orientation. Alibay et al. employed enhanced sampling MD simulations to probe the ring conformations of Glycosaminoglycan (GAG) related monosaccharides, as well as associated acetylated and sulfated GAG compounds. They suggest that in some cases, due to the high energy barriers between ring pucker states, an excess of 15 μs is required for adequate sampling.

For aldohexopyranose idose, the unique presence of three axial ring hydroxyl groups causes considerable conformational flexibility and can be used as a model for rationalising the relationship between anomeric configuration and pyranose ring pucker.[66] Sattelle et al. studied the ring pucker in methyl α- and β-L-idopyranosides with multi-microsecond MD simulations using the GLYCAM06 force field in three explicit water models.[66] They found that the calculated puckering exchange rates and free energies were independent of the water model. Both the α- and β-anomers favoured the $^1C_4$ state but this was less so for the α configuration (85% of the simulation time) compared to the β configuration (>99% of the simulation time) and with $^1C_4 \leftrightarrow^4 C_4$ exchange at rates of 20 μs^{-1} and 1 μs^{-1} respectively.

Moreover, they show that computed α-anomer $^1C_4 \leftrightarrow^4 C_4$ puckering rates depend on the exocyclic C6 substituent, when comparing hydroxymethyl and carboxyl groups. The slower ring pucker kinetics and restricted pseudorotational profile of the β-anomer "is caused by water occupying a cavity bounded by the anomeric 1-O-methyl and the C6 hydroxymethyl groups".[66] NMR measurements of pyranose ring vicinal ^{1}H-^{1}H spin-couplings ($^3J_{H,H}$) trend with computed values from the MD simulations in all instances. These simulation results highlight the effect of both anomeric configuration as well as substitution on ring pucker dynamics and help rationalise the different methyl α- and β-L-idopyranoside $^3J_{H,H}$ values.

Chythra et al. analysed the influence of the polarizable CHARMM Drude force field "on the ring dynamics of five major types of unsubstituted aldohexoses: glucose, mannose, galactose, altrose, and idose — and their anomers".[204] Their study revealed that the CHARMM36 additive force field favours $^4C_1$ conformations while the inclusion of polarizability enhances ring flexibility. Nonetheless, comparisons with experimental data revealed inconsistencies in the

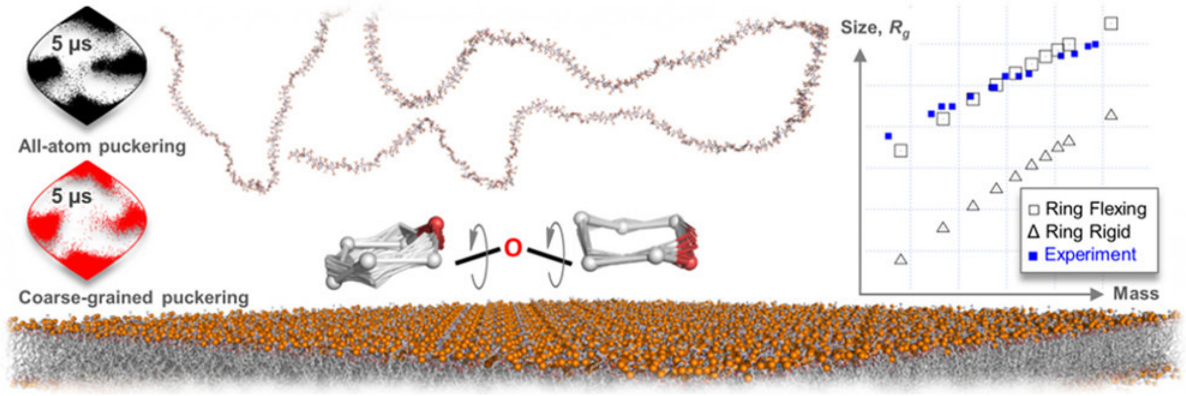


Figure 5.2: Conformations of the modelled heparanome glycans and the effect of ring flexing and ring rigid behaviour in all-atom and coarse-grained simulations. Reproduced with permission from [79] © American Chemical Society.

observed simulation ring pucker sampling and which will require updates to the polarizable Drude parameters.

The conformational behavior of linear oligo-β-(1$\rightarrow$3)-D-glucoside biopolymers were studied using NMR experiments and molecular modelling.[223] Interestingly a significant amount of anti-orientations for the β-1$\rightarrow$3-linkage, although the simulations were short and the observations were not confirmed by HSQC-NOESY experiments.

Sattelle et al. investigated sugar ring flexing and its effect on conformation and bioactivity in eleven heparanome oligosaccharides (Figure 5.2). Heparanome glycans are known to play a role in diverse set of biological processes including cancer, coagulation, nervous system development, and inflammation.[224] Unbiased solution multi-microsecond MD simulations of eleven heparanome oligosaccharides using the GLYCAM06 force field suggest that polymerisation alters the ring pucker dynamics of IdoA2S residues present within the chain. The ring flexing of the IdoA2S residues is also shown to be sensitive to chain position and substitution of adjacent residues. They also demonstrate that incorporating this ring pucker behaviour into a coarse-grain model allows for improved accuracy in the simulation of macromolecular sugars.

Interestingly, Kong et al. assessed the ability of the machine learning potentials combined with semi-empirical quantum chemical methods to model ring pucker conformers of five monosaccharides and oxane in the gasphase.[225] They found that this approach provides reasonable accuracy and computational efficiency but that the models reproduce chair more accurately than non-chair geometries.

5.2 Modelling bacterial polysaccharide antigens

Polysaccharide folding on a bacterial cell membrane is typically very hard to reproduce in simulated systems.[18] Both the mechanism of cell-surface polysaccharide biosynthesis and well as the molecular heterogeneity of the outer membrane composition pose particular challenges to molecular modelling. For example, LPS molecules are synthesized on the inner membrane and transported to the cell surface in a limited number of special sites.[226,227] O-Ag chain folding starts before LPS transfer to the outer membrane and this preliminary stage determines further evolution of the O-Ag conformation in the membrane environment.

Furthermore, An epitope* presented by a bacterial cell, such as an immunodeterminant of a carbohydrate antigen, can elicit an array of antibodies, differing in amount, type, and avidity. The structure of a single antigen-antibody complex will likely not account for all families of antibodies and alternative epitopes that may be involved in such an interaction.[8] There are many number of different conformational carbohydrate antibody-antigen complexes. Thus it may be more efficient in terms of both resources and time to model the conformation and dynamics of the free, unbound polysaccharide antigen in solution.[8,19]

A systematic and incremental approach for the modelling of polysaccharide antigens in solution has proved effective in a number studies.[8,16,18] First the preferred orientations of the glycosidic linkages in the polysaccharide antigen are calculated by potential of mean force (PMF) calculations of representative disaccharide units.[8,228] These preferred orientations are then used to build an oligosaccharide for use as a starting structure in MD simulations. The MD simulations are performed with carbohydrate specific force fields in solution (using counterions for ionic antigens where required) for a duration of several hundred nanoseconds to a few microseconds. This approach allows for the study of the conformation and dynamics of the polysaccharide over time.

MD simulations of polysaccharide antigens can be challenging due to the significant computational resources required to model them accurately. To overcome this issue, it has been shown that using oligosaccharides with several repeating units can provide a good approximation of the conformation and flexibility of an entire polysaccharide antigen.[16,17,19,22–24] However, the accuracy of the simulation is highly dependent on factors such as the number of residues in a repeating unit, the constituent monosaccharides, and the nature of glycosidic linkages.

Achieving sufficient conformational sampling is a challenge in MD simulations of polysaccharides. To achieve convergence, a simulation must be run for a long enough period to obtain a representative statistical weighting of all low-energy conformations. The starting structure of a

---

*The region of an antigen molecule to which an antibody attaches.

polysaccharide antigen is crucial, as it can significantly impact the simulation time required to achieve adequate sampling and convergence. Because of the large number of degrees of freedom in a polysaccharide molecule and its potential for significant flexibility, relatively long simulations must be performed. Although it is not possible to determine exactly when convergence has occurred, statistical methods can provide estimates.

Polysaccharide antigens can adopt a single well-defined conformer or cluster into a family of conformations.[8,13,19,22–24,229] By analyzing a polysaccharide antigen MD simulation trajectory, it is possible to identify the preferred conformations and distributions of the antigen's repeating units and determine whether any famlties of cluster conformations exist. Other derived chemical and physical properties, such as hydrodynamic size, viscosity, hydrogen bonding, and ring conformations, can also be measured. It is important to note that the accuracy and validity of the carbohydrates modelled are subject to the limitations of molecular modelling techniques, as well as the force field and methodology employed. Therefore, it is critical to carefully consider these factors when interpreting the MD simulations results of carbohydrates.

Simulations of carbohydrates can be validated by comparison with experimental data. Where possible, conformational structures can be elucidated from X-ray crystallography and directly compared with those from modelling. A somewhat less direct validation can be performed by NMR nOe experiments to measure interatomic distances of molecules and compared to those obtained in modelled structures. Such approaches are quite common,[19,22,126,135,230,231] although, measurements can be limited due to constraints of the nOe effect. A core part of the three dimensional characterization of gylycans remains the ability to relate NMR observables to populations of oligo- or polysaccharides.[10]

Modelling of bacterial carbohydrate antigens including *Salmonella*, *Escherichia coli*, *Streptococcus pneumoniae*, *Neisseria meningitides*, and *Shigella flexneri*, has illustrated the broad range of conformations possible, as discussed below.

5.2.1 *Salmonella*

Molecular modelling has been performed on several CPS and OAg of *Salmonella* bacterium including *S. typhi* Vi, *S. typhimurium* serotype B, and *S. enteritidis*.

S. typhi **Vi**: [-α-D-GalNAc-]$_n$

S. typhimurium **serotype B**: →2)-[α-D-Abe(1→3)]α-D-Man*p*-(1→4)-α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→

S. enteritidis: →3)-α-D-Gal*p*-(1→2)[α-Tyv*p*-(1→3)]-α-D-Man*p*-(1→4)-α-L-Rha*p*-(1→

Salmonella typhi Vi causes typhoid fever and the bacterial CPS is an antigenic component. The CPS antigen is a homopolymer polysaccharide with an α-(1→4)-Gal*p*NAcA repeating unit and variable O-acetylation at C-3.[232,233] Legnani et al. modelled hexasaccharides corresponding to *S typhi* Vi CPS as well as analogs containing a zwitterionic motif with various degrees of acetylation.[234] They found that the CPS has two preferred conformations; both with extended geometries but without any tendency towards supercoiling. The most stable conformation was a clockwise helix and the less stable conformation an anticlockwise helix. 3-O-acetylation also confers rigidity to the oligosaccharide affecting both its flexibility and hydrophobic character.

More recent work on *Salmonella enterica* serovar Typhi and *Citrobacter freundii*, which both produce the same Vi polysaccharide, explored the "effects of O-acetylation on the physical and immunological characteristics" of the CPS.[235] The study revealed that the Vi polysaccharides from the two bacteria differ with respect to viscosity in water and de-O-acetylation susceptibility, even though NMR reveals them to be structurally identical.[235] MD simulations elucidate the conformational changes caused by de-O-acetylation on Vi polysaccharide viscosity and antigenicity. The modelling reveals that O-acetylated Vi polysaccharide is a more rigid helix in solution. In comparison, de-O-acetylation leads to the polysaccharide adopting a more flexible coil with increased sodium ion interaction, which may be a cause of the reduced O-acetylated polysaccharide viscosity. In addition, "partial de-O-acetylation unveils concealed epitopes, recognized by both human and sheep anti-Vi PS immune sera".[235]

Salmonella typhimurium is a zoonotic pathogen and major non-typhoidal Salmonella disease that can cause gastroenteritis in humans. The O-Ag is constructed from the tetrasaccharide repeating unit: →2)[α-D-Abe(1→3)]α-D-Man*p*-(1→4)-α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→. Modelling of a 12 RU *S. typhimurium* serotype B O-Ag were performed by Galochkina et al. using two force fields (GLYCAM and OPLS-AA).[18]

The 1 μs MD simulations reveal that the O-Ag repeating units can undergo reversible or irreversible coil-to-globule transitions depending on the force field or temperature (Figure 5.3). The mechanism of these transitions is related either to the rotation of the carbohydrate residues around the glycosidic linkages or to flips of the pyranose rings. Interestingly, they found that the presence of rhamnose in the O-Ag chain significantly increases its conformational mobility. Additionally they found that O-Ag ring conformations from the GLYCAM simulation differs significantly from that of the OPLS-AA model — while no mannose ring flips we observed with either force field, several transitions with long lifetimes of up to 200 ns in the $^1C_4$ state were observed for galactose.

S. enteritidis is a predominant cause of food-borne salmonellosis in humans.[236] The CPS repeating unit consists of a trisaccharide backbone and a 3,6-dideoxy-D-*arabino*-hexopyranosyl

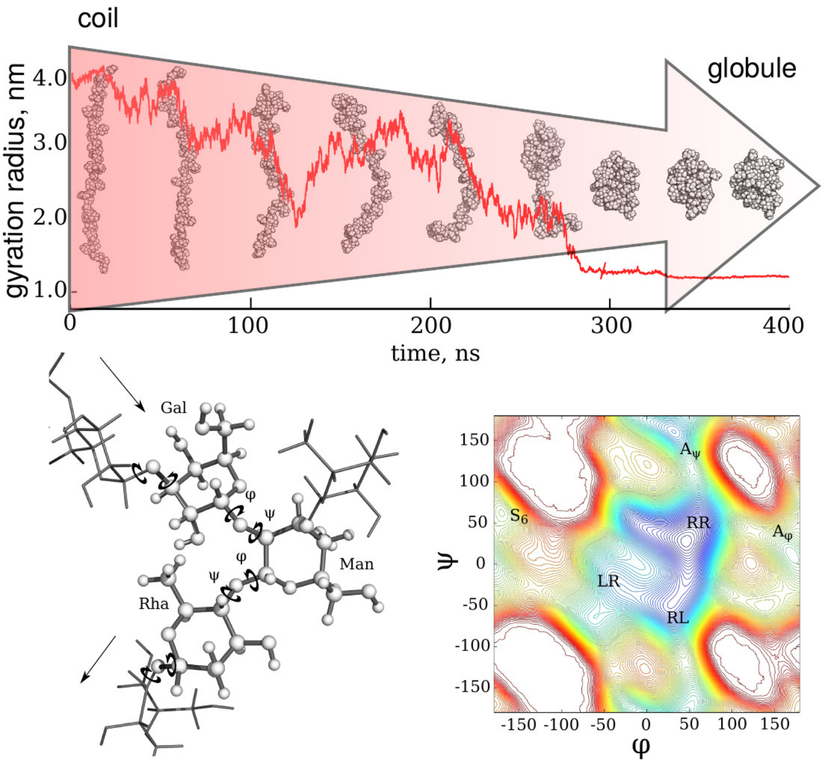


Figure 5.3: The long O-Ag chain from *Salmonella typhimurium* demonstrates qualitatively different conformational behavior depending on the temperature and the force field used. Novel simulation scenarios such as single O-Ag chain collapse into a dense globule or reversible coil-to-globule transitions cannot be deduced from the previous results for the shorter oligosaccharides. Reproduced with permission from [18] © NCBI.

side-chain: →3)-*α*-D-Gal*p*-(1→2)[*α*-Tyv*p*-(1→3)]-*α*-D-Man*p*-(1→4)-*α*-L-Rha*p*-(1→, where Tyv*p* is a 3,6-dideoxy-D-arabino-hexopyranosyl group.[237] Olsson et al. modelled the tetrasaccharide repeating unit from the *Salmonella enteritidis* CPS.[238] The MD simulations reveal two conformational famiilties and indicate that the chain flexibility may be greater when the mannosyl residue is vicinally disubstituted at O-2 and O-3 by the galactosyl and tyvelosyl groups respectively — even though this initially appeared to be a more sterically crowded and conformationally restricted oligosaccharide. Additionally, these results are consistent with the NMR and modelling measurements on *α*-D-Glc*p*-(1→3)[*β*-D-Glc*p*-(1→4)]-*α*-D-Glc*p*-OMe in which the anti-*ψ* conformational state at the (1→4)-linkage was significantly populated.[239] Notably, analysis of the ring pucker conformations of the tyvelosyl group showed that while the $^4C_1$ state is dominant both $^1C_4$ and other ring conformations are transiently observed.

5.2.2 *Escherichia coli*

Escherichia coli is the predominant facultative anaerobe of human colonic flora and includes both pathogenic and commensal strains.[240] Pathogenic strains are the most common cause of urinary tract infections and food poisoning.[9] *E. coli* is also responsible for many gastrointestinal diseases such as travelers' and infantile diarrhea.[227] The structures of approximately 197 *E. coli* O-Ags, with 27 subgroups have been elucidated to date.[126] The conformations of *E. coli* groups O5ab, O5ac, O6, O91, and O176 CPS were studied using NMR spectroscopy and molecular modelling and demonstrate that subtle variations in polysaccharide primary structure can lead to significant conformational differences.

O5ab: →4)-*β*-D-Qui*p*3NAc-(1→3)-*β*-D-Rib*f*-(1→4)-*β*-D-Gal*p*-(1→3)-*β*-D-Gal*p*NAc-(1→

O5ac: →2)-*β*-D-Qui*p*3NAc-(1→3)-*β*-D-Rib*f*-(1→4)-*β*-D-Gal*p*-(1→3)-*β*-D-Gal*p*NAc-(1→

O176: →4)-*α*-D-Man*p*-(1→2)-*α*-D-Man*p*-(1→2)-*β*-D-Man*p*-(1→3)-*α*-D-Gal*p*NAc-(1→

O6: →4)-*α*-D-Gal*p*NAc-(1→3)-*β*-D-Man*p*-(1→4)-*β*-D-Man*p*-(1→3)-*β*-D-Glc*p*NAc-(1→

O91: →4)-*α*-D-Qui*p*3NAc-(1→4)-*β*-D-Gal*p*-(1→4)-*β*-D-Glc*p*NAc-(1→4)-*β*-D-Glc*p*A6NGly-(1→3)-*β*-D-Glc*p*NAc(1→

The O-Ags of *E. coli* O5ab and O5ac are composed of very similar linear tetrasaccharide repeating units — with the linkage position of the *β*-D-Qui*p*3NAc residue as the only difference. For both O5ab and O5ac a family of repeating unit conformations were identified. O5ab may have four distinct conformations while the O5ac family may be smaller with two distinct conformations.

Modelling of *E. coli* LPS within a lipid bilayer have been performed for serotypes O6, O91, and O176 using the CHARMM36 force field.[136,138,139] Wu et al. constructed a model of an *E. coli* R1 (core) O6 (antigen) LPS molecule and performed several all-atom simulations of the LPS (lipid and O-Ag) in a membrane bilayer.[241] The simulations included varying lengths of O-Ag repeating units (0RU, 5RU, and 10RU) on the LPS molecule. The MD simulation results show that increasing LPS molecular length significantly influences LPS structure and dynamics as well as on the LPS bilayer properties (Figure 5.4). ^{1}H,^{1}H NOESY experiments were conducted on the O-Ag to measure interproton cross-relaxation rates and distances. They reveal that the effective ^{1}H-^{1}H distances show generally very good agreement between the NMR experiments and MD simulations of the O-Ag repeating units.

^{1}H-^{1}H NOESY and HSQC-NOESY experiments were used to calculate ^{1}H-^{1}H cross-relaxation rates and transglycosidic $^3J_{C,H}$ coupling in the *E.coli* O91 O-Ag. MD simulations of 10RU chains of the O-Ag were performed in solution as well as within an LPS bilayer. Comparison of the NMR measurements against those of solution MD simulations of a 10RU chain demonstrate

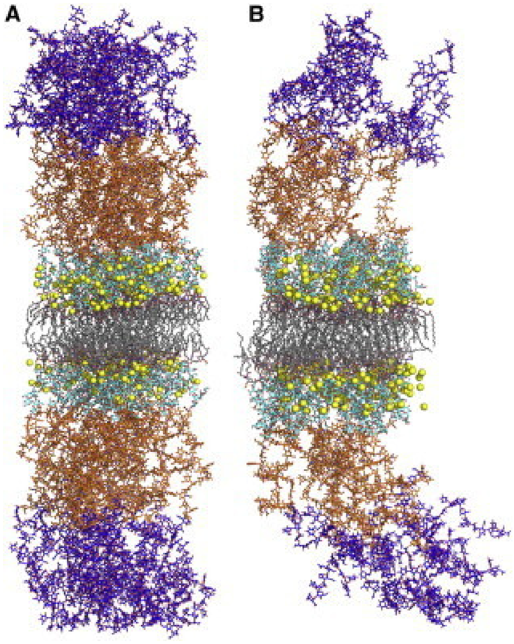


Figure 5.4: LPS lipid bilayer snapshots: (A) LPS10 (lipid A + R1 core + 10 units of O6 antigen) and (B) MIX10_50 (LPS0/LPS10 mixed bilayer with 50% LPS10). The coordinate sets are from the 100 ns time point. Lipid A, R1 core, O6 antigen repeating units 1–5, and repeating units 6–10 are coloured in gray, cyan, orange, and blue, respectively. Calcium ions are shown in yellow spheres. Water molecules and KCl are not shown for clarity. Reproduced with permission from [242] © American Chemical Society.

that modelling is able to reproduce the dynamical behavior of internal correlation times for the O-Ag. The simulations demonstrate that conformations of the O-Ag in solution and within the LPS bilayer are similar but not identical with respect to populated states. The O-Ag simulated within the LPS bilayers had more extended chains as a result of spatial limitations due to close packing.

Patel et al. used NMR spectroscopy and MD simulations to determine the conformational preferences of *E. coli* O176 O-Ag.[139] The O176 O-Ag consists of a tetrasaccharide backbone with three mannopyranose residues and a single N-acetylglucosamine residue. The experimentally derived internuclear distances of the O-Ag with ten repeating units showed very good agreement to those calculated from MD simulations of the O-Ag in solution. Analysis of the LPS bilayer simulations with five and with ten RUs revealed that, although similar with respect to populated states in solution, the O-Ag in LPS bilayers had more extended chains compared to when in directly solution.

The modelled O176 LPS bilayer conformations were compared to MD simulations of O-Ag polysaccharides from *E. coli* O6 and O91.[136,138] For all three O-Ags, the ensemble of structures present for the polysaccharides in solution were consistent with the results from their

^{1}H,^{1}H NOESY experiments. The similarities between the O-Ag on its own and as a constituent of the full LPS in bilayer environment show that it is possible describe the LPS conformation and dynamics from MD simulations in solution alone.

5.2.3 *Streptococcus pneumoniae*

The conformational properties of several *Streptococcus pneumoniae* polysaccharide antigens have been widely investigated.[17,27,78,243–247] *S. pneumoniae* is the main cause of invasive pneumococcal disease and is largely responsible for community acquired pneumonia and meningitis in children and the elderly. There are over a hundred different serotypes of *S. pneumoniae* and the bacterial CPS is essential for pathogenesis.[248] Below, we discuss the application of molecular modelling to explore vaccine cross-protection in three key pneumococcal serogroups, Pn19 and Pn6, and Pn 14:

Pn 19F: →4)-α-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→2)-α-L-Rha*p*-(1→P→

Pn 19A: →4)-α-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→P→

Pn 6A: →2)-α-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→3)-D-Rib-(5P→

Pn 6B: →2)-α-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→4)-D-Rib-(5P→

Pn 6C: →2)-α-D-Glc*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→3)-D-Rib-(5P→

Pn 6D: →2)-α-D-Glc*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→4)-D-Rib-(5P→

Pn 14: →6)-β-D-Gal*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Kuttel et al. explored the conformations and dynamics of the polysaccharide antigen repeating units from serotypes 19F and 19A in an attempt to understand the low cross-protection of 19F vaccine against 19A disease.[17,78] The CPS repeating unit of both serotypes are trisaccharides with N-acetlymannose, glucose and rhamnose residues: (19F) →4)-α-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→2)-α-L-Rha*p*-(1→P→; (19A) →4)-α-D-Man*p*NAc-(1→3)-α-D-Glc*p*-(1→2)-α-L-Rha*p*-(1→P→. The only difference in the primary chemical structures of the repeating units is the linkage position of the glycosyl residue to Rha*p*: 1→2 linked in Pn19F and 1→3 linked in Pn19A.

Free energy surfaces were calculated for the glycosidic linkages and compared with 100 ns solution MD simulations with the CHARMM36 and the GLYCAM06 force fields.[17,78] Although key differences were identified between the force fields, overall they were in agreement in predicting a 19F and 19A repeating unit conformations. The modelled CPS repeating units from both serotypes "form unstructured random coils, with inflexible repeat units linked by mobile phosphodiester linkages".[17] However, in 19F the rhamnose residue is nearly orthogonal to the backbone chain whereas in 19A it is more linear. This results in the 19F CPS adopting a wider range of conformational families than the more restricted 19A trisaccharide, which suggests a probable conformational difference between the 19F and 19A polysaccharides and may explain the low cross-protection of 19F vaccines against 19A disease.

A recent examination of serogroup 6 pneumococcal capsules revealed eight serotypes, with four of particular interest: Pn6A, Pn6B, Pn6C, and Pn6D, which all feature ribitol phosphate in a trisaccharide RU.[249] Multivalent polysaccharide vaccines demonstrate that vaccines wtih Pn6A are able to provide cross-protect against Pn6C, but Pn6B cannot. However, vaccines with Pn6B can cross-protect against Pn6A and Pn6D. MD simulations elucidated the conformations of serogroup 6 polysaccharides, revealing significant differences among serotypes.[8] Replacing galactose with glucose increased stability, while the substitution of sugar linkages impacted conformation. Pn6B displayed the greatest conformational diversity, partially explaining its ability to cross-protect against Pn6A and Pn6D, but not Pn6C.

Pneumococcal serotype 14 (Pn14) CPS is structurally similar to *Streptococcus agalactiae* serotype III (GBSIII PS), with the addition of sialic acid as on GBSIII as the only difference. However, they have unique immunogenicites.[250] MD simulations of Pn14 and GBII repeating units provide insight into the diverse conformations of the two antigens.[246] Interestingly, the simulations demonstrate that the two CPSs share a dominant solution conformation. The backbone of the CPSs are stabilised by their side chains but further from the branch-point there is more flexibility.

5.2.4 *Neisseria meningitides*

Neisseria meningitidis (meningococcus) is a gram-negative bacteria responsible for several life-threatening diseases including meningitis, meningococcemia, and pneumonia. It can cause both endemic and epidemic infection across all age groups and mortality rates can be high if the infection is not treated rapidly and appropriately. The widespread use of conjugate meninogococcal vaccines over the last decade, particularly in sub-Saharan Africa has seen a significant reduction in the number of reported cases.

There are twelve meningococcal serogroups, with the vast majority of all incidences globally attributed to six of these: A, B, C, W, X, and Y. The *N. meningitidis* CPS is the main virulence factor and a primary antigenic component. Structurally there are pairs of CPS from the different serogroups that a very similar. The CPSs from groups A and X respectively consist of phosphodiester linked α-D-Man*p*NAc and α-D-Glc*p*NAc homopolymers. Serogroup B and C CPSs are also homopolymers but of sialic acid residues that vary in their linkage positions and

O-acetylation patterns. The serogroup Y and W CPSs are formed by disaccharide repeating units of sialic acid and α-D-Glc*p* (for MnY) or α-D-Gal*p* (for MnW) residues.

MnA: →6)-α-D-Man*p*NAc(3/4OAc)-(1→OPO_3→

MnX: →4)-α-D-Glc*p*NAc-(1→OPO_3→

MnB: →8)-α-D-Neu*p*NAc-(2→

MnC: →9)-α-D-Neu*p*NAc(7/8OAc)-(2→

MnY: →6)-α-D-Glc*p*-(1→4)-α-D-Neu*p*NAc(7/9OAc)-(2→

MnW: →6)-α-D-Gal*p*-(1→4)-α-D-Neu*p*NAc(7/9OAc)-(2→

The CPS repeating units of serogroups Y and W are almost identical. Both are disaccharides, contain a sialic acid residue, and in the case of MnY an α-D-Glc*p* residue while MnW contains a α-D-Gal*p* residue. The only difference in the repeating units is the stereochemistry of the C-4 hydroxyl group of the Glc/Gal residue, which is equitorial in MnY and axial in MnW.

The structural similarity between the CPS repeating units of the two serogroup suggests the possibility of cross-protection between MnY and MnW vaccines. While cross-protection has not been deliberately tested in current licensed vaccines, a small scale clinical trial evaluated mono- and divalent MnY and MnW vaccine formulations.[251] The study showed that both the mono- and divalent vaccines were able to elicit a strong immune response and elicit antibodies against their respective antigens. Moreover, the monovalent was also able to elicit antibodies against the other serogroup but elicited different levels of cross-protection. Approximately 71% of volunteers given the MnY vaccine had antibody against MnW four weeks after taking the vaccine. Less than half of this number (30%) of volunteers given the MnW vaccine had antibody against MnY after the same period. Molecular modelling of the CPSs can provide a potential rationale for these observed differences.

Kuttel et al. modelled the conformation and dynamics of the meningococcal Y and W polysaccharide antigens. 3 RU strands of MnY and MnW had distinct differences in the orientations of the α-D-Neu*p*NAc-(2→6)-α-D-Glc*p* linkage in MnY and α-D-Neu*p*NAc-(2→6)-α-D-Gal*p* linkage in MnW. The α-D-Neu*p*NAc-(2→6)-α-D-Glc*p* linkage has a single orientation corresponding to the *gg* conformation whereas the α-D-Neu*p*NAc-(2→6)-α-D-Gal*p* linkage moved freely between the three staggered conformations: *gg*, *gt*, *tg*.

This behaviour results in MnY adopting a single preferred conformation while MnW exhibits a family of conformations (Figure 5.5). These results provide a potential explanation for the observed differences in the MnY and MnW vaccines. MnY CPS would likely raise antibodies

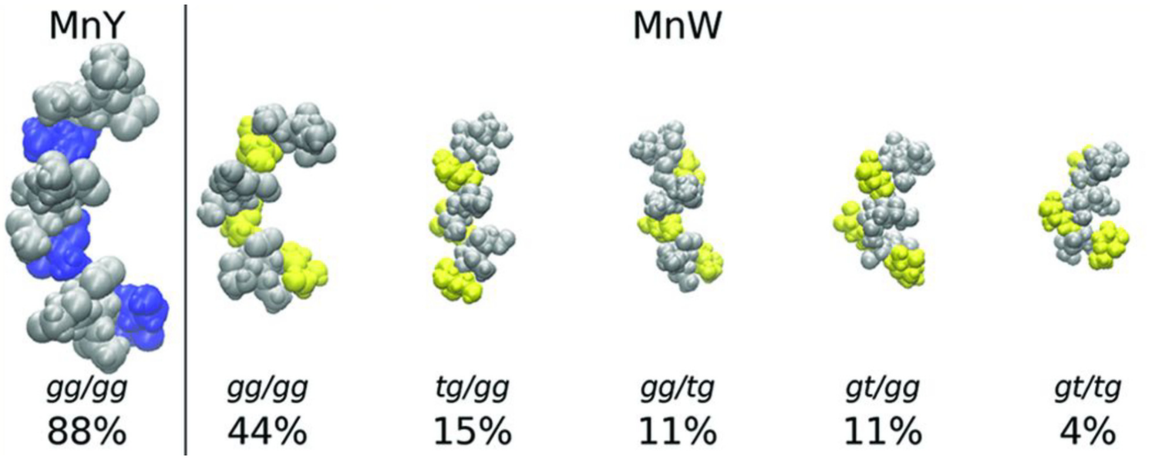


Figure 5.5: Representative structures from the dominant conformational families for the capsular polysaccharides in MnY (left) and MnW (right). Reproduced with permission from [8] © American Chemical Society.

against a single conformation only while MnW would raise antibodies against a family of conformations, only one of which would correspond to the MnY antigen.

Molecular modelling has also provided insights into the loading of sialic acid-containing meninogococcal CPSs (MnY, W, C) onto a carrier protein.[252] MD simulations with 6 RUs chains for 1 μs reveal significantly different molecular flexibilities (MnW > MnC > MnY) and that the shorter simulations with 3 RU strands do not adequately model the flexibility of the CPSs. Further modelling revealed a correlation between CPS flexibility and the number of chains terminally attached to the CRM^{197}carrier whereby increased chain flexibility enables accommodation of additional CPSs on the protein surface.

There are currently three licensed tetravalent conjugate vaccines against meningococcal serogroups A, C, Y and W.[253] The CPSs of MnA and MnX are both homopolymers of amino sugars α-D-ManNAc and α-D-GlcNAc featuring phosphodiester linkages at C-6 and C-4, respectively. The similar primary structure of these polysaccharides raises the possibility that vaccination targeting MnA could offer cross-protection against MnX diseases. MD simulations revealed that the MnA CPS behaves as a flexible random coil with significant population of compact S-bend conformations, which become less conformationally defined with increasing length.[20,21] The MnX CPS, in contrast, “forms a more stable regular helical structure” with four residues per turn and these results are supported by NMR analysis.[20,21] Moreover, the simulations indicate that the O-Ac groups in MnX are prominently exposed to the solvent, promoting more extended conformations compared to the relatively compact structures of MenA, which lack O-aAc. These findings carry potential implications for the strategic design of conjugate vaccines and suggest that O-acetylation may influence the efficacy of cross-protection.

5.2.5 *Shigella flexneri*

The O-Ag conformations for *S. flexneri* serotypes Y, X, 2, 3, 5 and 6 have been explored using NMR and MD simulations, which showed that, while serotype Y is highly flexible, glucosylation and O-acetylation in *S. flexneri* 2, 3, and 5 restricts O-Ag flexibility.[15,16,22–24] The O-Ag structures of serotypes 1 and 4 have also been studied but in less detail.[15] The conformation and dynamics of *S. flexneri* 7a and 7b O-Ags are yet to be investigated and are the focus of this study.

Theillet et al. performed studies of the conformations of w *S. flexneri* O-Ags[†] using a combination of experimental and modelling techniques.[15] MD simulations of 3RU O-Ags found "that in all O-Ags, but 1a and 1b, serotype-specific substitutions of the backbone do not induce any new backbone conformations compared with the linear type O-Ag Y."[15] corroboration in part by NMR nOe measurements, although, the uncertainty associated with the measured nOe distances is quite large and may correlate to different conformations than those obtained from the modelling.

Clement et al. investigated the conformations of *S. flexneri* 5a O-Ags as part of a Shigella vaccine program.[77] Modelling with the MM3 force field and comparison with ^{1}H-^{1}H distance measurements from NMR indicated a single conformation for the *S. flexneri* 5a O-Ag: a right-handed 3-fold helix with the glucosyl residue on Rha$^{\mathrm{II}}$ pointing outwards. Conformational analysis of the *S. flexneri* Y O-Ag[254] suggest it adopts a linear chain and the differences in conformation between *S. flexneri* Y and 5a is attributed to the branched glucosyl residue.

When modelling O-Ags it is important to consider chain length: shorter chains may not have enough molecular flexibility, while longer chains are computationally expensive to model. In the case of *S. flexneri*, studies in mice have shown that a 3RU chain adequately represents the O-Ag conformation.[15,22,23,47] Antibodies usually bind to small fragments of the O-Ag, typically between one and seven residues, which corresponds to 1–2 RU in the case of *S. flexneri*.[15,47] More recent modelling of *Sf* Y, 2, 3, 5, and 6 O-Ags were performed on 6 RU chains.[22–24]

The *S. flexneri* Y O-Ag of up to 6 RU was modelled with a well-defined, systematic approach by Kang et al.[16] These studies revealed that simulations of a few repeating units display polymer-like flexibility. Additionally, hairpin-like conformations facilitate the formation of metastable compact states.[16]

[†] O-Ags for serotypes 1a, 1b, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, X, and Y were modelled computationally. NMR conformational analysis were performed for serotypes 1a, 2a, 3a, and 5a O-Ags.[15]

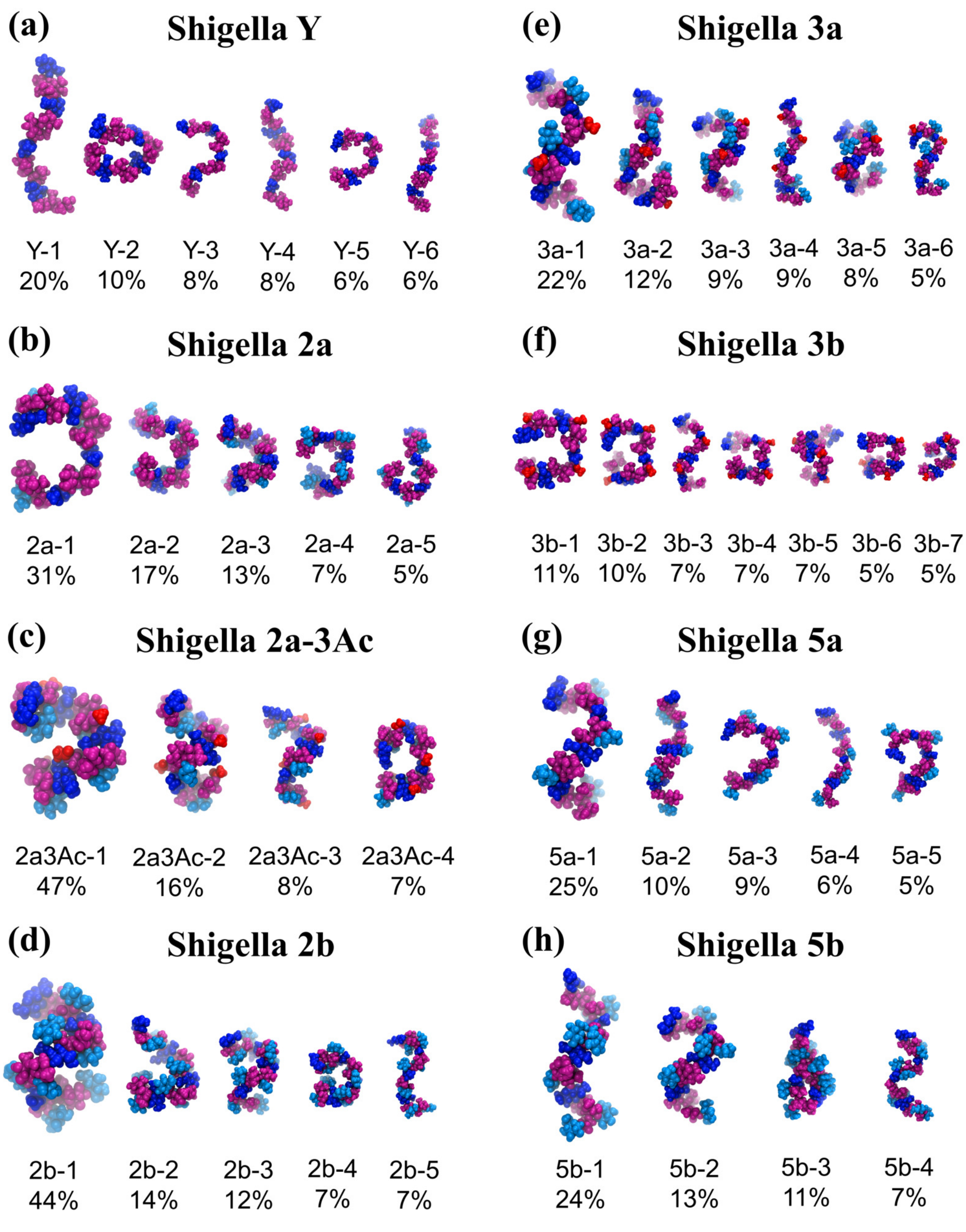


Figure 5.6: Dominant conformations of the central 4 RU of the 6 RU chains for (a) Y, (b) 2a, (c) O-3-acetylated 2a, (d) 2b, (e) 3a, (f) 3b, (g) 5a, and (h) 5b. Relative occupancies in the simulations (excluding the initial 200 ns) are indicated as percentages. Clusters of less than 5% are not shown. The sugars are coloured: pink for Rha, dark-blue for GlcNAc, cyan for Glc side chains, and red for O-acetyl groups. Reproduced with permission from [22].

These results were confirmed in-part with extended multi-microsecond length MD simulations (Figure 5.6 and Figure 5.7).[22,24] The studies demonstrated the extreme flexibility of the unbranched backbone with a high number of both extended and compact conformations. Moreover simulations with the GLYCAM06 force field showed formation of collapsed, globular conformations[16] that were absent when modelled with CHARMM36.[22,24]

MD simulations of *S. flexneri* serotypes 2, 3, and 5 revealed that all have flexible backbones, with substitutions affecting chain dynamics in different ways. Three general heuristics for the effects of substitution were proposed: "(1) substitution on rhamnose C (Rha^{I}) reduces the extension of the O-Ag chain; (2) substitution at O-3 of rhamnose A (Rha^{III}) restricts the O-Ags to predominantly helical conformations, (3) substitution at O-3 of rhamnose B (Rha^{II}) has only a slight effect on conformation".[22,23] Where possible we will look to extend the assumptions of the effect of substitution on backbone conformation.

Further conformational modelling of *S. flexneri* 6 O-Ag showed significant differences compared to *S. flexneri* Y with chains that are predominantly extended.[24] Moreover, it was found that O-acetylation has little effect on backbone conformation and hence may not be essential for antigenicity[†].[24] These results are corroborated in part by an in vivo study in mice, which indicate that O-acetylation of the *S. flexneri* 6 O-Ag does not have an impact on immune response.

These studies collectively suggest that a quadrivalent glycoconjugate vaccine containing serotypes 2a and 3a could provide coverage against *S. flexneri* serotypes 3b and 5 but not serotype 6.[22–24] This study will allow for the conformational comparison of *S. flexneri* 7a and 7b to these studied O-Ags. It may also be possible to form assumptions on bacterial antigenicity and potential vaccine cross-protection.

[†] Antigenicity is the generation of a specific antibody by immune cells in response to an antigen

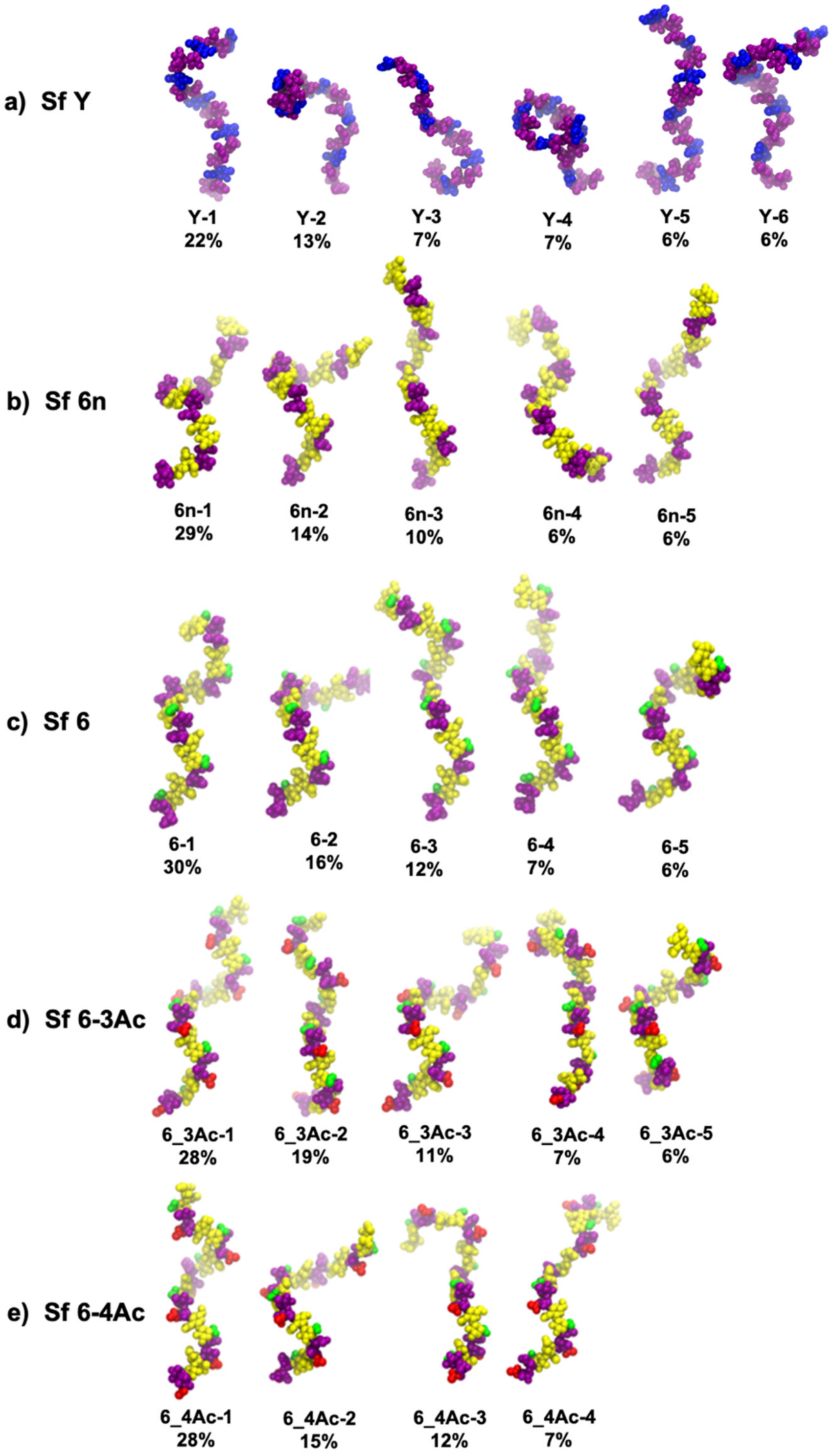


Figure 5.7: Dominant conformations of the middle four repeating units of the 6 RU O-Ag chains for *S. flexneri*: (a) Sf Y, (b) *Sf* 6n, (c) *Sf* 6, (d) *Sf* 6-3Ac, and (e) *Sf* 6-4Ac. The relative occupancies of the conformational clusters (excluding the initial 200 ns of equilibration) are shown as percentages with clusters less than 6% not shown. The colour scheme is as follows: blue for Glc and Glc*p*NAc; purple for Rha*p*; yellow for Gal*p*, Gal*p*A, and Gal*p*NAc; green for the COOH group of GalA; red for the O-acetyl groups. Reproduced with permission from [24]

Chapter 6. Molecular Modelling of *Shigella flexneri* 7a and 7b O-antigens

S. flexneri is a gram-negative bacterium and a causative agent of shigellosis; a gastrointestinal disease caused by the *Shigella* family of bacteria.[38,42,49] Shigellosis is among the leading causes of infant death in developing countries.[38–40] The *S. flexneri* O-Ag is a key immunogenic component and understanding its molecular conformation is of interest. Previous studies have modelled and analysed the conformations of *S. flexneri* Y, 1, 2, 3, 5, and 6.[15,16,22–25] This work focuses on comparing the conformations of structurally similar O-Ags from two subtypes of the *S. flexneri* bacterium: 7a and 7b (Figure 6.1). In this chapter we outline the methods used to model O-Ag repeating units of *S. flexneri* 7a and 7b and analyse the simulation results, with comparisons made to conformations of the *S. flexneri* Y O-Ag from previous studies.[22,24]

6.1 Methods

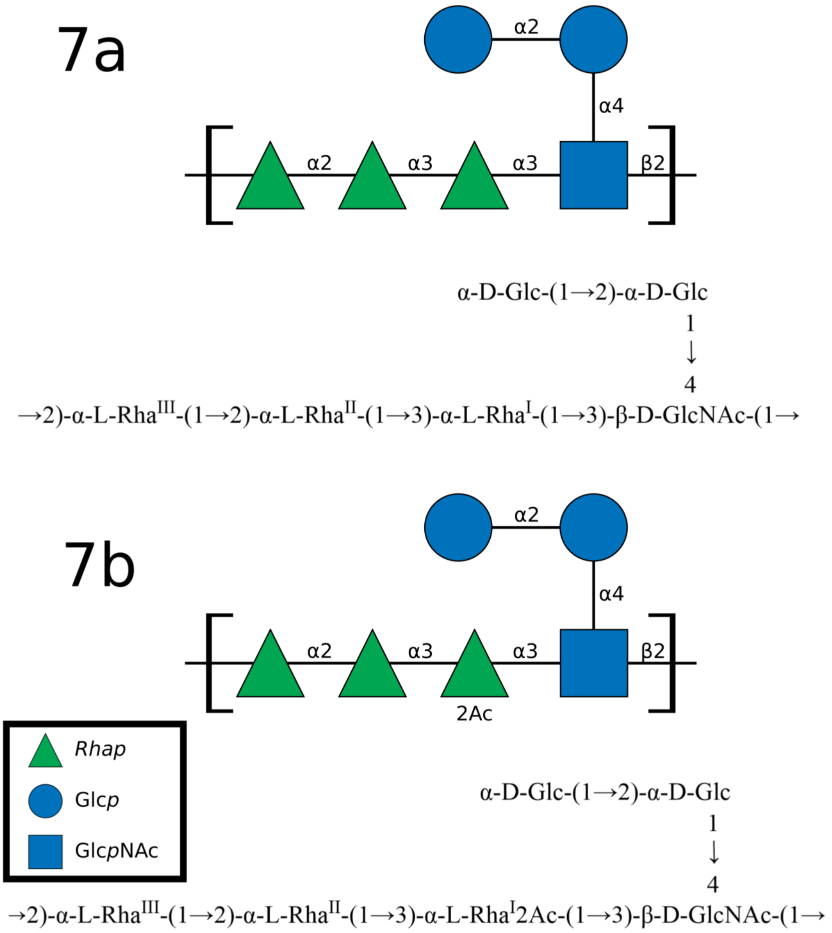
We follow a systematic approach that we have established for modelling polysaccharides.[8] We start by using preferred orientations of the glycosidic linkages for representative disaccharides present in the *Sf* 7a and 7b repeating units, precalculated using solution metadynamics simulations.[228] Using these values, and taking into account the glycosidic linkage torsion angles measured from the *Sf* Y 6 RU simulations,[22,24] we build a single 6 RU polysaccharide each of *Sf* 7a and 7b. These static structures were built using CarbBuilder v2.1.34[228,255] and are used as the starting model for the MD simulations.

6.1.1 MD simulations

We conducted independent all-atom, solution MD simulations[*] for *Sf* 7a and 7b in a 90 Å^3 water box. The simulations were 1 μs in length at a 1 fs time-step, each for 6 RU chains of

---

[*] All molecular modelling simulations were performed at the University of Cape Town High Performance Computing Facility.



$$\begin{array}{r} \alpha\text{-D-Glc-}(1\rightarrow2)\text{-}\alpha\text{-D-Glc}\\ 1\\ \downarrow\\ 4\\ \rightarrow2)\text{-}\alpha\text{-L-Rha}^{III}\text{-}(1\rightarrow2)\text{-}\alpha\text{-L-Rha}^{II}\text{-}(1\rightarrow3)\text{-}\alpha\text{-L-Rha}^{I}\text{-}(1\rightarrow3)\text{-}\beta\text{-D-GlcNAc-}(1\rightarrow \end{array}$$

$$\begin{array}{r} \alpha\text{-D-Glc-}(1\rightarrow2)\text{-}\alpha\text{-D-Glc}\\ 1\\ \downarrow\\ 4\\ \rightarrow2)\text{-}\alpha\text{-L-Rha}^{III}\text{-}(1\rightarrow2)\text{-}\alpha\text{-L-Rha}^{II}\text{-}(1\rightarrow3)\text{-}\alpha\text{-L-Rha}^{I}\text{2Ac-}(1\rightarrow3)\text{-}\beta\text{-D-GlcNAc-}(1\rightarrow \end{array}$$

Figure 6.1: Structures of the biological repeating units of *S. flexneri* 7a and 7b O-Ags. Each structure is represented in SNFG notation (top) and standard nomenclature (bottom).

Sf 7a and 7b. Simulations were performed with the CHARMM36 force field,[55–58,60] which is optimised for carbohydrate-based simulations.

All MD simulations were conducted in solution using NAMD v2.13.[256] The TIP3P water model was employed to simulate the aqueous solution.[171] Both prior to and after solvation the MD simulations followed a standard NAMD protocol consisting of 10,000 steps of energy minimization. Each system was equilibrated in the first 100 ns of the simulation. The CHARMM36[55–58,60] all-atom force field was used in the simulations.

The NAMD velocity-verlet integrator[163] was employed and long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) summation method, with a κ value of 0.20 Å and PME grid dimensions of 12 nm. Non-bonded interactions were truncated, and a switching function was applied between 1.2 nm and 1.5 nm for groups with integer charge. The 1–4 interactions were not scaled as per CHARMM36 force field recommendations.[257] All simulations used periodic boundary conditions. Temperature and pressure were controlled using the Nose-Hoover thermostat[258,259] and Langevin piston barostat[260] methods under the NPT ensemble at 300 K and 1 atm. The NPT ensemble was chosen over the NVT ensemble due to the recommendation of constant pressure for periodic simulations in NAMD.[261]

6.1.2 Analysis and visualisation

For visualising and analysing the data, we utilized Visual Molecular Dynamics (VMD) v1.9.2[262] or developed in-house Python scripts using the MDAnalysis library.[263,264] The initial 100 ns of all simulations were considered as an equilibration phase and were excluded from analysis, unless otherwise stated. We analyse the conformation and dynamics of the modelled glycans by several measurable parameters including block standard error (to determine simulation convergence), cluster conformations, glycosidic linkage torsion angles, and pyranose ring puckering. Direct comparison of all measured parameters are made against previously conducted simulations of *Sf* Y.[22,24] The molecular conformations were rendered using the Tachyon VMD engine and visualised with the van der Waals or Licorice representations. When necessary, the PaperChain visualisation algorithm[265,266] was employed to emphasize the glycan rings.

Simulation convergence

We use block averaging analysis to assess simulation convergence[267,268] in a manner similar to previous carbohydate modelling studies.[22–24,26,27,199,252] The block averaging algorithm divides a simulation trajectory consisting of N frames into M blocks, each containing n frames in length.

$$N = nM \tag{6.1}$$

A single measurable parameter (e.g. radius of gyration, root mean square deviation or end-to-end distance) is selected and an average is calculated for each block. The block length, n, is incrementally increased. At each value of n, the block averages are recalculated. We then compute the mean for each block, the mean of all the block means, and the standard error of all the block means (the block standard error — BSE). The BSE is plotted as a function of the block size and the simulation is indicated to have converged once the BSE asymptotes to a plateau.

Torsion angles

The atoms that form the glycosidic linkage torsion angles are defined as:

$$\phi = \mathrm{H}_1 \text{ - } \mathrm{C}_1 \text{ - } \mathrm{O}_1 \text{ - } \mathrm{C}'_x$$
$$\psi = \mathrm{C}_1 \text{ - } \mathrm{O}_1 \text{ - } \mathrm{C}'_x \text{ - } \mathrm{H}'_x$$

O1 and O6 rotamer conformations are defined by $C(n+1) - C(n) - O(n) - H$ as used in previous studies.[269] Hydroxyl orientations were binned as gauche (g$^-$), anti (t) or gauche (g$^+$) centred on -60, ±180 and $+60°$, respectively. Hydroxymethyl conformations (tg, gt and gg) were defined by the torsions O_6-C_6-C_5-O_5 as has also been done in similar studies.[63]

Molecular end-to-end distance

The end-to-end distance, r, of a modelled polysaccharide is a measure of chain extension and flexibility. We define r for all *S. flexneri* O-Ags as the distance between C2 of α-L-Rha$^{\mathrm{II}}$ in the first repeating unit of the polysaccharide backbone (atom i) and C1 of α-L-Rha$^{\mathrm{I}}$ in the terminal repeating of the backbone (atom j) as follows:

$$r = \left[\left(\mathbf{x}_i - \mathbf{x}_j \right)^2 \right]^{\frac{1}{2}} \tag{6.2}$$

where x_i and x_j are the positions atom i and atom j.

Pyranose ring pucker

The Cremer-Pople pucker parameters for each ring were calculated according to a protocol provided by Kooijman,[270] which is based on the original Cremer-Pople derivations.[71] The angle ranges for the Cremer-Pople azimuth (θ) angle used to distinguish between chair and boat/skew conformers for pyranose rings as used in this study are as follows: $^4C_1$: $0° <= \theta < 30°$, boat/skew: $75° <= \theta < 105°$, and $^1C_4$: $150° < \theta <= 180°$. We do not anaylse half-chair and envelope states as they are very short lived states and we do not sample the MD trajectory at a sufficiently high frequency to perform an adequate analysis.

Free energies of the β-D-Glc$^{\mathrm{a}}$nd β-D-Glc*p*NAc ring pucker were calculated from simulated equilibrium data of the trisaccharides studied using the CHARMM36 and GLYCAM06 force fields. The $\Delta G = -RTln|K_{eq}|$ relationship is used to compute the free energy differences; where ΔG is the change free energy, R is the gas constant, T is the Temperature, and K_{eq} is the equilibrium constant.

Conformational clustering

Conformational clustering is a method of grouping molecular structures obtained from different time steps of a simulation trajectory. This can be achieved by employing various clustering algorithms that identify structures with similar conformations based on one or more predefined, measurable criteria.[271,272]

We used a modified Quality Threshold Algorithm[273] to compute clusters. This algorithm involves three parameters: the measurement criteria, the maximum threshold, and the estimated number of clusters. The algorithm starts by selecting a random point as the centre of the first cluster. It then iteratively adds the closest points until the threshold distance is exceeded. This process is repeated for the remaining points until all points are assigned to clusters, with outliers grouped together.

The 6 RU O-Ag in each each frame of the MD trajectory was aligned along the ring and glycosidic linkage atoms using the root mean square difference (RMSD) metric, with a cut-off of 2 Å for the *Sf* Y O-Ag and 8 Å for the *Sf* 7a and 7b O-Ags. We extracted the centroid frame, representing the geometric mean, from each particular cluster as a representative conformation. Only clusters containing a proportion of simulation frames greater than 5% were considered for analysis.

6.2 Results

In analysing the MD simulations of *Sf* Y (as modelled in previous studies), 7a, and 7b, we first determine convergence of each of the simulations. We then compare and contrast the extension and flexibility of the saccharide chains, followed by identification and analysis of the dominant conformational families for each O-Ag. We subsequently evaluate all glycosidic linkage torsion angles in the repeating unit and finally the ring pucker conformations of the residues.

6.2.1 Simulation convergence

We used block averaging analysis of two metrics of chain flexibility – end-to-end distances and radius of gyration — to asses MD simulation convergence (Figure 6.2). The blocked standard error (BSE) of both metrics reaches a plateau for *Sf* Y indicating convergence of the structure over the course of the simulation. While the *Sf* 7a BSE begins to plateau there is still large variation in the BSE at small block sizes (approx. 40 ns), which suggests that convergence has not occurred. *Sf* 7b has not converged over the course of the simulation as the BSE does not plateau.

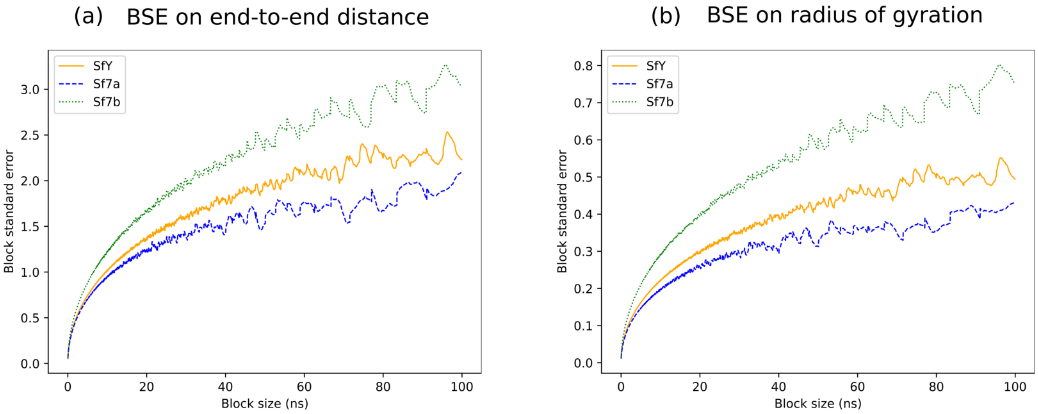


Figure 6.2: Block averaging analysis calculated for the time series of (a) end-to-end distance and (b) radius of gyration of the 6 RU O-Ags for serotypes Y, 7a, and 7b.

For *Sf* Y the correlation time indicates sufficient sampling at 50 ns — far less than the simulation time of 2000 ns. The correlation times for the *Sf* 7a and *Sf* 7b are over 50 ns — close to the simulation times of 1000 ns. The lack of convergence and the long correlation times in the serotype 7 simulations is unusual and unexpected. Similar 6 RU simulations of other *S. flexneri*[22–24] and other bacterial antigens reveal that simulations between 1 μs and 2 μs provide good conformational sampling. Extended simulations of *Sf* 7a and 7b may be required to allow for sufficient conformational sampling and convergence. Further analysis of the glycan chain flexibility, glycosidic linkage torsion angles, and ring conformations should provide further insight into reasons the simulations did not converge.

6.2.2 Chain extension and flexibility

The time series of polysaccharide end-to-end distance, r, provides a simple measure of chain extension and flexibility. The r time series of the three *S. flexneri* O-Ags (Figure 6.3) indicates that *Sf* Y is flexible, adopting both extended as well as compact conformations. The mean r for *Sf* Y is 39.6 Å (σ = 15 Å) with a range of 5 Å to 75 Å (Figure 6.3-b). The chain is highly flexible; repeatedly fluctuating between collapsed and extended forms. There are two clear peaks in the r distribution of *Sf* Y — at 26 Å and corresponding to a more collapsed molecular conformation and at 53 Å with a more extended chain.

In contrast, *Sf* 7a (Figure 6.3-c) fluctuates between extended and compact forms. While *Sf* Y has a bimodal distribution, *Sf* 7a is uni-modal with a single peak at 46.0 Å and left-skewed towards extended conformations. The mean length is 46.0 Å (σ = 11 Å). The disaccharide side chain to the 4-position of β-D-Glc*p*NAc significantly changes the conformation and dynamics of the O-Ag backbone.

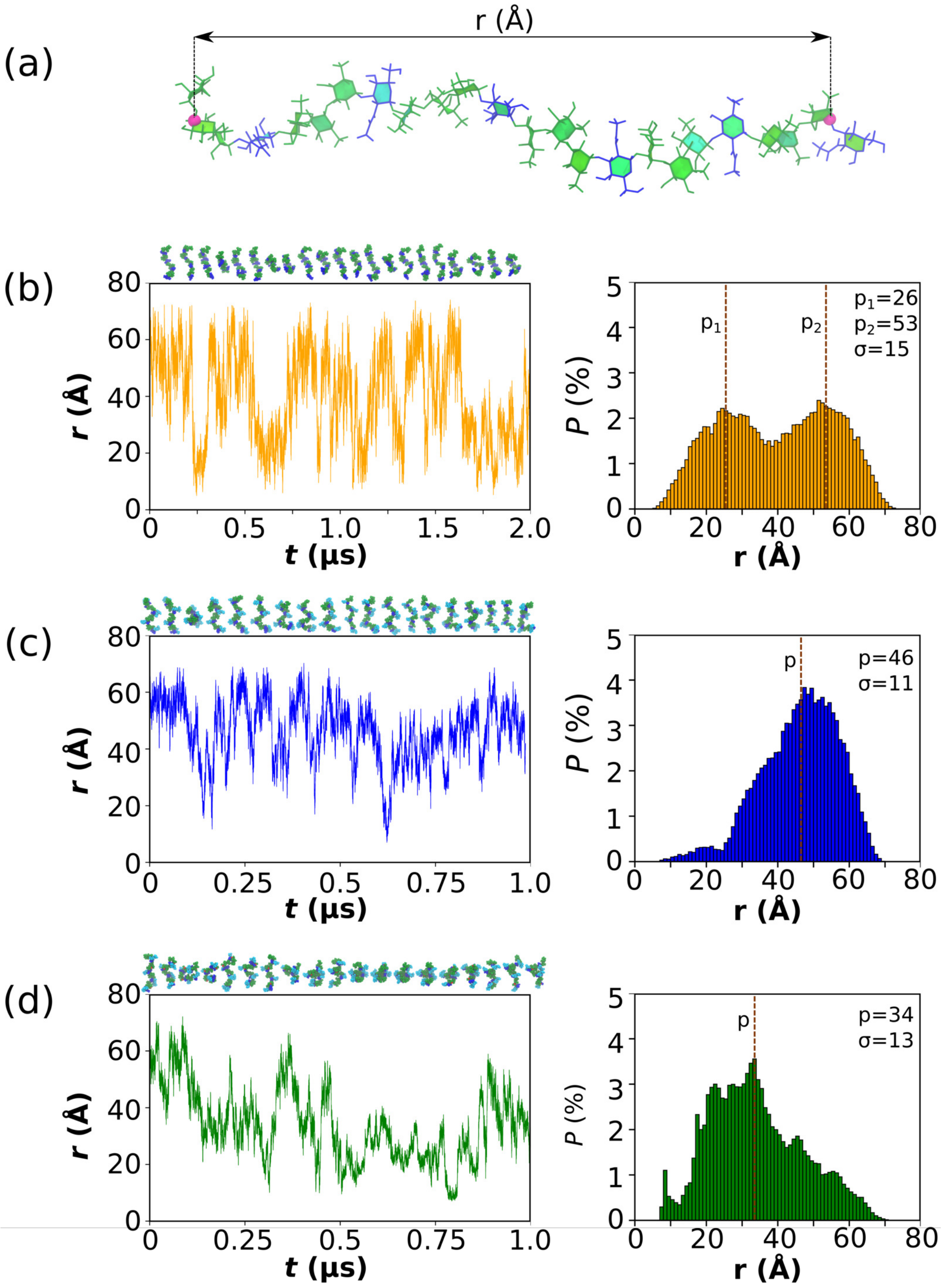


Figure 6.3: (a) The end-to-end distance, r, is indicated on the 6 RU *Sf* Y molecule and is measured between C2 of α-L-Rha$^{\text{II}}$ in the first repeating unit of the polysaccharide backbone and C1 of α-L-Rha$^{\text{I}}$ in the terminal repeating of the backbone. Time series graphs (left column) and corresponding histograms (right column) for the simulation trajectories are shown for O-Ags (b) *Sf* Y (2 μs), (c) *Sf* 7a (1 μs), (d) *Sf* 7b (1 μs). Conformational snapshots, at 100 ns intervals for *Sf* Y and 50 ns intervals for *Sf* 7a and 7b, are shown above the time series plots. The histograms are labeled with modal peak r value(s) (p) and standard deviations (σ).

The r time series for *Sf* 7b (Figure 6.3-d) also fluctuates between extended and compact conformations, but the latter predominates. The r histogram has a single right-skewed peak with a mean of 34 Å ($\sigma = 13$ Å). The single peak is at 26 Å. The shorter distances suggest that 2-OAc of α-L-Rha$^{\mathrm{I}}$, which is present only on *Sf* 7b, causes the backbone to adopt a more collapsed conformation. These more compact conformations occur in the second half of the simulation and may be closely correlated to specific glycosidic linkage orientations or ring conformations.

Sf 7a adopts extended conformations for longer periods than *Sf* Y and 7b, while *Sf* 7b maintains compact conformations for longer periods compared to the other two O-Ags. All three O-Ags share a common backbone repeating unit but the disaccharide side-chain causes *Sf* 7a and 7b to have significantly different flexibility to *Sf* Y by inducing changes in orientation of the backbone glycosidic linkages. However, *Sf* 7a and 7b have a small difference in their primary chemical structure (O-Ac on Rha$^{\mathrm{II}}$ of *Sf* 7b) yet there are large differences in the flexibility of their glycan chains. While end-to-end distance serves as a valuable yet imperfect measure of glycan flexibility, a more useful understanding of O-Ag conformation can be obtained through a comprehensive analysis of polysaccharide chain conformations and glycosidic linkage torsion angle rotations.[252]

6.2.3 O-antigen conformations and dynamics

We compare the central 2 RU dominant backbone chain conformations from the 6 RU *S. flexneri* 7a and 7b O-Ag simulations. All three polysaccharides have multiple conformational families.

Sf Y has a highly flexible unbranched backbone, with diverse conformational families.[22,24] The two peaks observed in the r histogram do not correspond to any particular conformation, but rather represent a broad range of both compact and extended as well as helical and non-helical forms (Figure 6.4: a-d). The primary cluster is an extended conformation as predicted earlier using the HSEA method,[77] and accounts for 62% of the simulation conformations. A smaller extended family accounts for a further 9% of the conformations. There are two additional compact conformations at 11% and 9%. These findings are consistent with a different MD study of serotype Y, which also found a variety of conformations in shorter chains with up to four repeating units.[16]

Our modelling of *Sf* 7a reveals a single dominant conformer (70%) with an extended chain (Figure 6.4-e) that corresponds to the single peak observed in the r histogram. The two minor conformers (13% and 7%) are more compact and a differently elongated chain respectively (Figure 6.4: f, g). This suggests that the disaccharide glucose side chain of *Sf* 7a causes the backbone to favour more extended forms compared to the unbranched *Sf* Y.

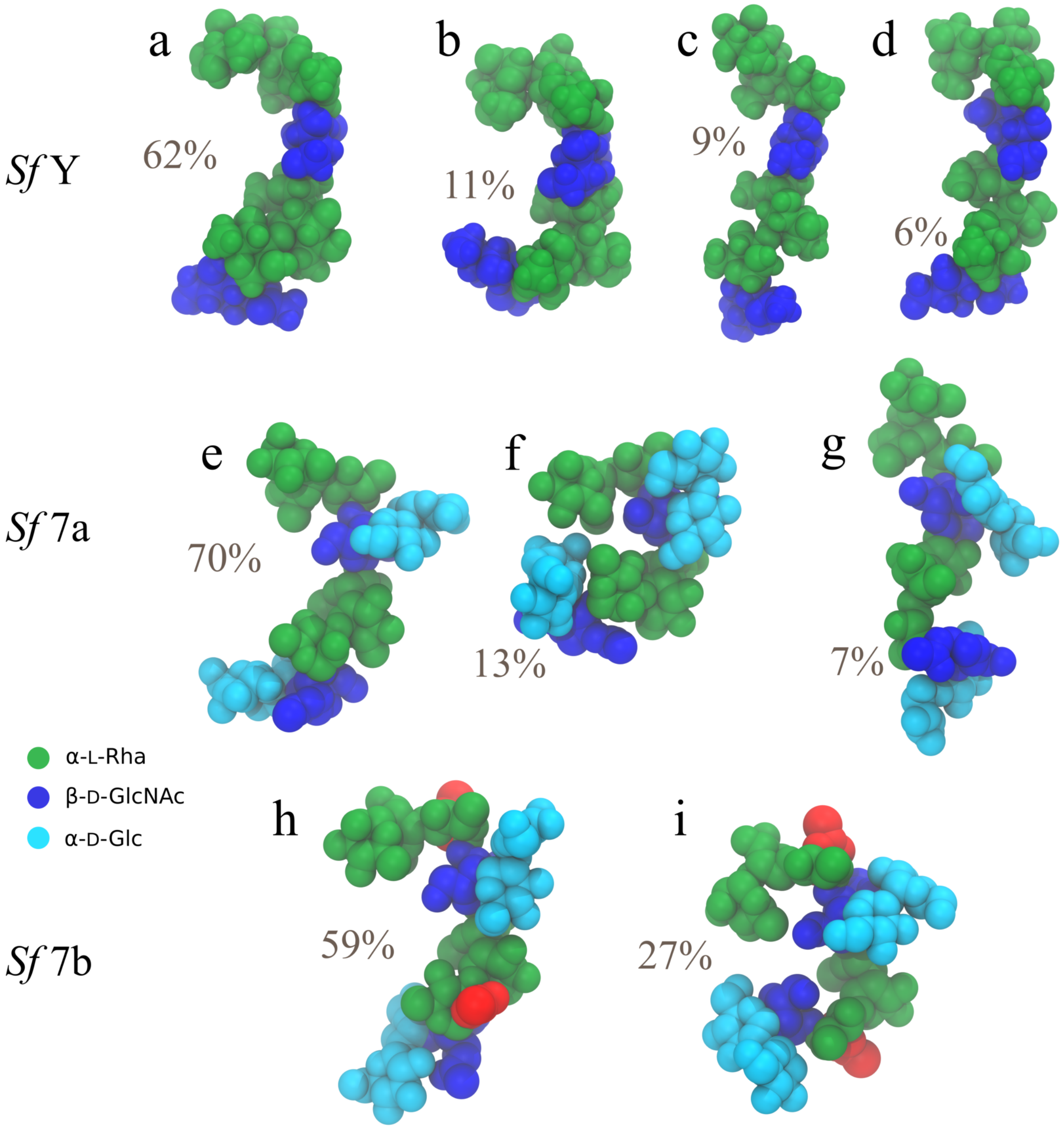


Figure 6.4: Dominant conformations of the central two repeating units in the 6 RU O-Ag for *S. flexneri* Y (top row), 7a (middle row), and 7b (bottom row) O-Ags. Relative occupancies in the MD simulations (excluding the initial 100 ns) are indicated as percentages. Clusters of less than 5% are not shown. The sugars are shown with the VdW representation in VMD and coloured as follows: green for Rha, dark-blue for GlcNAc, cyan for Glc, and red for O-acetyl groups.

Sf 7b, in contrast, has two conformational families — a moderately compact, loosely curved conformer (59%); and a compact helix (27%) (Figure 6.4: h, i). Overall, *Sf* 7b adopts more compact conformations than 7a indicating that 2-OAc on α-L-Rhap^{I} induces more compressed backbone conformations. The 2-OAc on α-L-Rhap^{I} is located on the outer edges of the helical turn, likely making it more available for binding. None of the conformations families in either *Sf* 7a and 7b adopt readily identifiable helical form.

Changes in polysaccharide conformation primarily occur due to two factors: changes in glycosidic linkage torsion angles and, to a lesser extent, changes in the pucker of the sugar rings. We observe puckering in the β-D-GlcNAc*p* residues into non-chair conformers in both *Sf* 7a and 7b but not *Sf* Y. This puckering likely directly influences the O-Ags conformations by altering the glycosidic linkage torsion angles.

6.2.4 Glycosidic linkage torsion angles

We analyse the glycosidic linkage torsion angles as they are a principle degree of freedom for polysaccharides. Figure 6.5 compares the glycosidic linkages for the four central repeating units in *Sf* Y, 7a, and 7b. The orientation of glycosidic linkages is determined by the two torsion angles, ϕ and ψ, as defined: ϕ = H_1 - C_1 - O_1 - C'_x and ψ = C_1 - O_1 - C'_x - H'_x.

The H1-Rhap^{III} – H2-Rhap^{II} (Figure 6.5: a-g) and H1-Glc*p*NAc – H2-Rhap^{III} (Figure 6.5: n-p) glycosidic linkages adopt very similar orientations across all three O-Ags while the H1-Rhap^{II} – H3-Rhap^{I} (Figure 6.5: h-j) and H1-Rhap^{I} – H3-Glc*p*NAc (Figure 6.5: k-m) linkages are significantly different.

The are two preferred orientations for the H1-Rhap^{III} – H2-Rhap^{II} and H1-Glc*p*NAc – H2-Rhap^{III} glycosidic linkages for all three O-Ags — $\phi, \psi \approx +45°, +15°$ and $\phi, \psi \approx +45°, +60°$. Similarly, there are two dominant orientations for the H1-Glc*p*NAc – H2-Rhap^{III} glycosidic linkages with populated regions at $\phi, \psi \approx +25°, +45°$ and $\phi, \psi \approx -50°, +30°$. These results suggest that the disaccharide side chain, present only in *Sf* 7a and 7b, does not significantly alter the orientation of these three glycosidic linkages.

In contrast, the H1-Rhap^{II} – H3-Rha$^{\mathrm{I}}$ and H1-Rhap^{I} – H3-Glc*p*NAc linkages are different in all three O-Ags. In *Sf* Y the linkage is predominantly populated at $\phi, \psi \approx +40°, +20°$. There is also a less populated at $\phi, \psi \approx +30°, +50°$ and a small anti ψ torsion. The H1-Rhap^{II} – H3-Rhap^{I} glycosidic linkage of *Sf* 7a populates similar torsion angles to *Sf* Y but has a different distribution. The glycosidic linkage torsion angle is primarily centred in a narrow range at $\phi, \psi \approx +40°, +15°$. There are also two additional population centres that are much smaller in size. The larger and broader grouping is at $\phi, \psi \approx +30°, -60°$ while the smaller and more

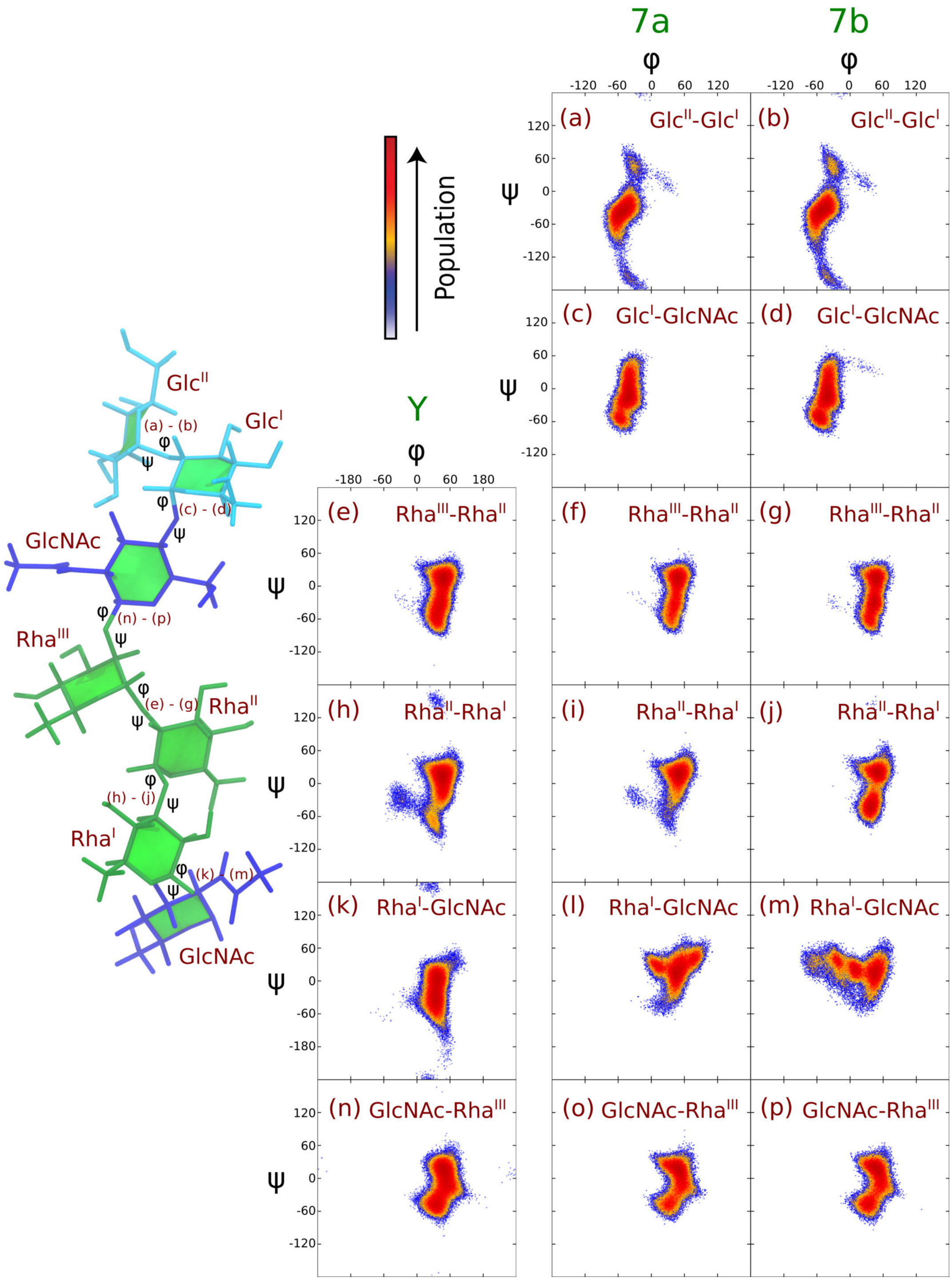


Figure 6.5: Scatter plots illustrated as heat maps for the glycosidic linkage torsion angles (ϕ, ψ) in the four central repeating units (RU2-RU5) of the modelled O-Ag polysaccharide of *Sf* Y (e, h, k, n), 7a (a, c, f, i, l, o), and 7b (b, d, g, j, m, p). ϕ, ψ defined as H_1 - C_1 - O_1 - C'_x and C_1 - O_1 - C'_x - H'_x respectively. The heat maps combine the points from the four central repeating units to broadly sample backbone behavior. The color scale on the upper-left indicates the relative occupancy of the torsion angles during the simulations. The structure on the left depicts the O-Ag residues and glycosidic linkages.

concentrated group is at $\phi, \psi \approx -40°, -20°$. However, the H1-Rha$^{\text{II}}$ – H3-Rha$^{\text{I}}$ glycosidic linkage of *Sf* 7b occupies two distinct regions: $\phi, \psi \approx +40°, +15°$ and $\phi, \psi \approx +30°, -60°$. There is a near equal distribution between the two centres. The H1-Rha$^{\text{II}}$ – H3-Rha$^{\text{I}}$ linkage adopts a broader conformational space compared to *Sf* Y, 2a, and 2b.[22] The presence of the 2-OAc on α-L-Rha$^{\text{I}}$ leads to a bi-modal distribution of the glycosidic linkage torsion angle as also observed in *Sf* 3a and 3b.[23]

The H1-Rha$^{\text{I}}$ – H3-Glc*p*NAc glycosidic linkage is different in the three O-Ags. In *Sf* Y only the ϕ region is populated between 0° and +60°. The ψ torsion angle ranges from −80° to 60°. The glycosidic linkage also spends a short period of time in the anti-orientation at $\phi, \psi \approx +20°, \pm 180°$.

The linkage in *Sf* 7a and 7b adopts different torsion angles in comparison. In *Sf* 7a the H1-Rha$^{\text{I}}$ – H3-Glc*p*NAc linkage covers a range between −20° and 10° in ϕ and between 0° and 80° in ψ. The torsion angle heat map forms an unusual shape with the predominant population centre at ϕ, ψ = 50°, 20° and a smaller centre at ϕ, ψ = 0°, 40°. In *Sf* 7b the torsion primarily populates three regions with the largest at $\phi, \psi \approx +50°, +20°$ and the two smaller regions at $\phi, \psi \approx +20°, +30°$ and $\phi, \psi \approx +50°, -20°$. Each of these centres are also smaller than those in *Sf* Y and 7a.

The H1-Glc$^{\text{I}}$ – H4-Glc*p*NAc linkage connects the disaccharide side chain to the backbone in *Sf* 7a and 7b (Figure 6.5: c, d). The glycosidic linkage torsion angle populated by both serotypes are similar — between −100° and −20° in ϕ and between −80° and 70° in ψ indicating that the linkage is quite flexible. There are two primary population centres in both serotypes at $\phi, \psi \approx -50°, -20°$ and $\phi, \psi \approx -60°, -60°$. There is also a much smaller populated region in the $\phi+$ region of *Sf* 7b but it does not appear significant.

The torsion angles of the H1-Glc$^{\text{II}}$ – H2-Glc$^{\text{I}}$ linkage are similar for *Sf* 7a and 7b (Figure 6.5: a, b). There is a single large centre at $\phi, \psi \approx -40°, -40°$ that spreads between $\psi \approx 0$ and $\psi \approx -80$. There is also a less populated region at $\phi, \psi \approx -30°, -40°$. In addition, the glycosidic linkage adopts an anti-orientation for short periods ($\phi, \psi \approx -20°, \pm 180°$), however, *Sf* 7a populates this space for more time than *Sf* 7b.

Both *Sf* 7a and 7b show unusual orientations of the H1-Rha$^{\text{I}}$ – H3-Glc*p*NAc and H1-Glc$^{\text{I}}$ – H4-Glc*p*NAc glycosidic linkages. Moreover, time series analysis of the torsion angles for the two linkages reveal a non-systematic alternation between states — selecting random samples of the ϕ and ψ torsion angles from the trajectory provides significant different values. Such behaviour may play a role in the observed non-convergence in the BSE sampling, although, other factors could also contribute. Therefore, the ring pucker conformations of the residues were analysed with initial focus on the β-D-Glc*p*NAc ring.

6.2.5 β-D-Glc*p*NAc ring pucker

The conformational differences between *Sf* 7a, 7b, and Y also appears to be associated with puckering of β-D-Glc*p*NAc and is further investigated. The ring of pyranose sugars can adopt a range of shapes, called puckers, but are generally in $^4C_1$. Here we employ the Cremer-Pople method, which uses three parameters (ϕ, θ, and Q) to fully define the ring pucker conformation of a pyranose ring (Section 2.1.1). The meridian (ϕ) and azimuth (θ) angles are used to identify the specific pucker state while the amplitude (Q) provides the magnitude of the pucker. The two chair conformers are denoted $^4C_1$ and $^1C_4$ with values of $\theta \approx 0°$ and $\theta \approx 180°$ respectively. For both chairs ϕ can adopt any value between 0° and 360°. For boat/skew conformers $\theta \approx 90°$ and the value of ϕ defines the specific type of boat or skew.

Time series of the Cremer-Pople θ parameters for β-D-Glc*p*NAc in the four central repeating units of the 6 RU *Sf* Y, 7a and 7b demonstrate the presence of $^4C_1$, boat/skew, and $^1C_4$ states (Figure 6.6). Polar heatmaps of the Cremer-Pople meridian (ϕ) and azimuth (θ) for the β-D-Glc*p*NAc residue in the four central repeating units of *Sf* Y, 7a and 7b illustrate the population distribution across the different ring pucker states (Figure 6.7).

In *Sf* Y the $^4C_1$ ($\theta \approx 0$) chair conformation of the β-D-Glc*p*NAc ring is dominant as expected.[22–24,63] Two β-D-Glc*p*NAc residues (Figure 6.6: b, d) adopt boat/skew conformation ($\theta \approx 90$) for 25 ns (about 1% of the simulation time). These correlate to $^{1,4}B$ and $^1S_5$ ring pucker states (Figure 6.7: b, d).

In contrast, the β-D-Glc*p*NAc ring in both *Sf* 7a and 7b adopts pucker states where boat/skew conformers are the most prevalent (Figure 6.6: e - h). Further, in RU4 the β-D-Glc*p*NAc residue adopts a $^1C_4$ conformer $\theta \approx 180$ about half-way through the simulation and maintains this pucker state until the end. The three common ring pucker states adopted by β-D-Glc*p*NAc in *Sf* 7a and 7b are $^4C_1$, $^{1,4}B$, $^1S_5$, and $^1C_4$ (Figure 6.8).

For both *Sf* 7a and 7b boat/skew conformers are dominant in β-D-Glc*p*NAc residues of RU3, RU4 and RU5 — present for approximately 62% of the simulation (Figure 6.6: e-l, Figure 6.7: e-l). Interestingly, in *Sf* 7a and 7b a single β-D-Glc*p*NAc residue enters the $^1C_4$ pucker conformation and remains in this state for about half of the simulation. In both O-Ags, once β-D-Glc*p*NAc adopts a $^1C_4$ state, it does not transition into any other conformer, remaining entirely in $^1C_4$. Additionally, the variation in β-D-Glc*p*NAc ring conformation significantly influences

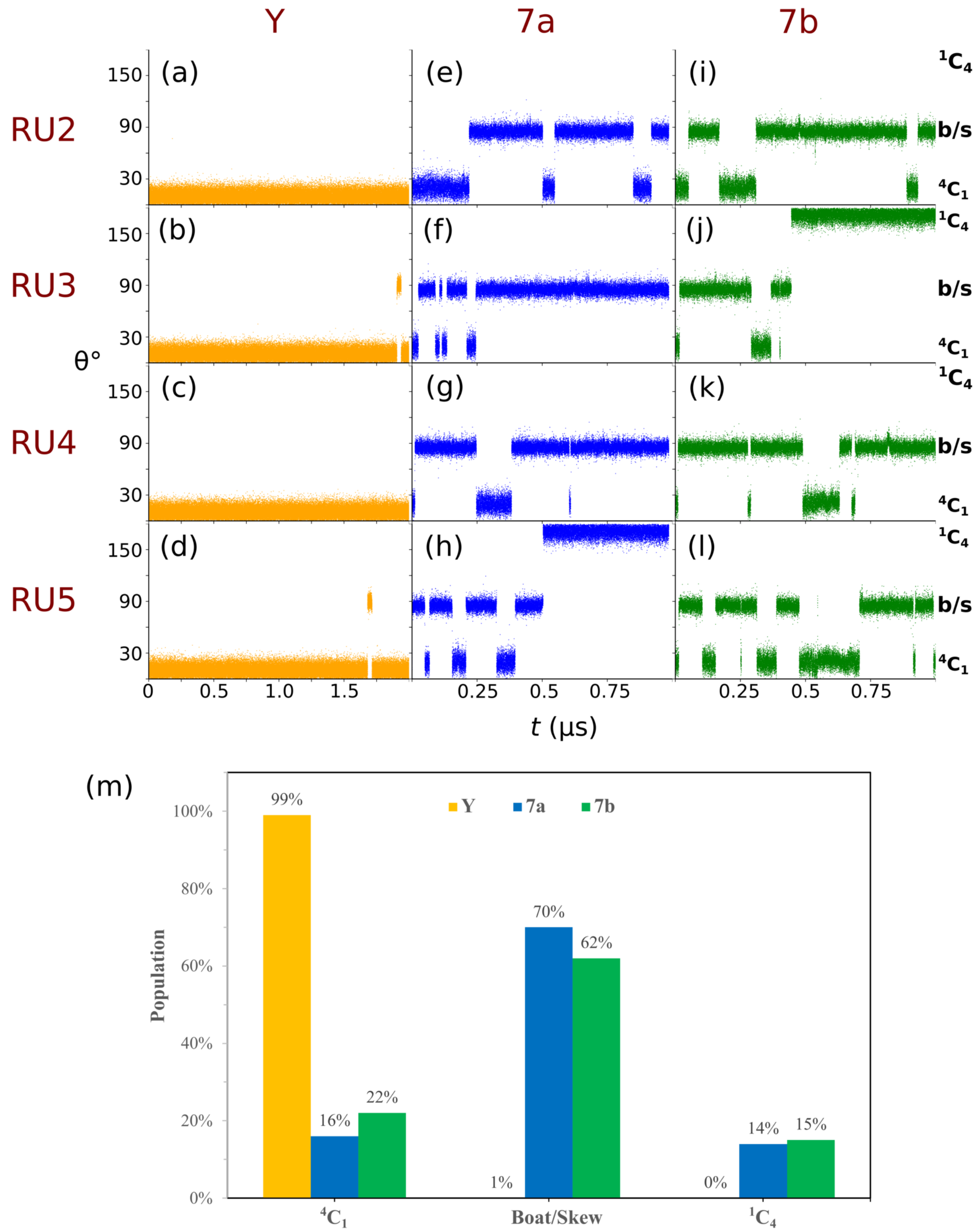


Figure 6.6: (a-l) Time series plots for the β-D-Glc$p$NAc Cremer-Pople ring pucker θ paramter in the four-central repeating units of *Sf* Y, 7a and 7b. Relationship between θ and ring pucker conformation: $^4C_1$ ($\theta \approx 0°$), boat/skew (b/s, $\theta \approx 90°$), $^1C_4$ ($\theta \approx 180°$). In all four central repeating units of *Sf* Y β-D-Glc$p$NAc remains entirely in $^4C_1$, with the exception of a short transition into boat/skew conformers. *Sf* 7a and 7b on average adopt boat/skew states for over half of the simulation time. The $^1C_4$ state is occupied for shorter periods across all the repeating units while the $^4C_1$ state is adopted only in a single repeating unit (RU5 in *Sf* 7a and RU3 in *Sf* 7b). (m) Distribution of β-D-Glc$p$NAc ring pucker conformations. Averaged from the four central repeating units in *Sf* Y, 7a, and 7b.

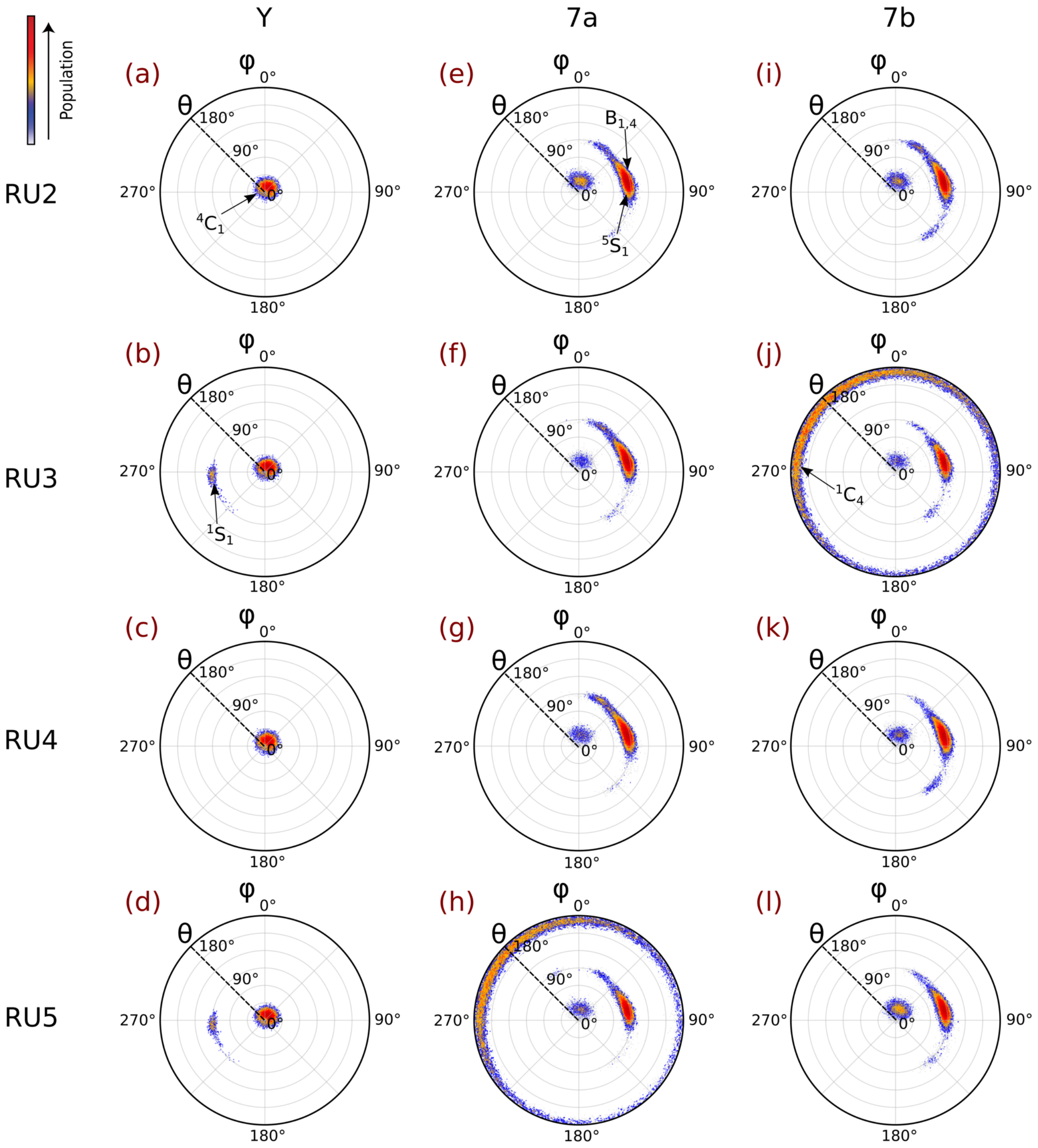


Figure 6.7: Polar heatmap plots for the β-D-Glc*p*NAc ring pucker θ angle in the four-central repeating units of the O-Ags of *Sf* Y, 7a and 7b. *Sf* Y almost entirely populates the $\theta \approx 0$ region (corresponding to $^4C_1$ ring conformer) across all four central repeating units with very minor scatterings at $\theta \approx 90$ (boat/skew) in two of the repeating units.*Sf* 7a and 7b primarily populate the $\theta \approx 90$ (boat/skew) region across all four repeating units with smaller centres at $\theta \approx 0$ ($^4C_1$). β-D-Glc*p*NAc in a single repeating unit (RU5 for 7a and RU3 for 7b) also populate $\theta \approx \pm 180$ ($^1C_4$) for extended periods.

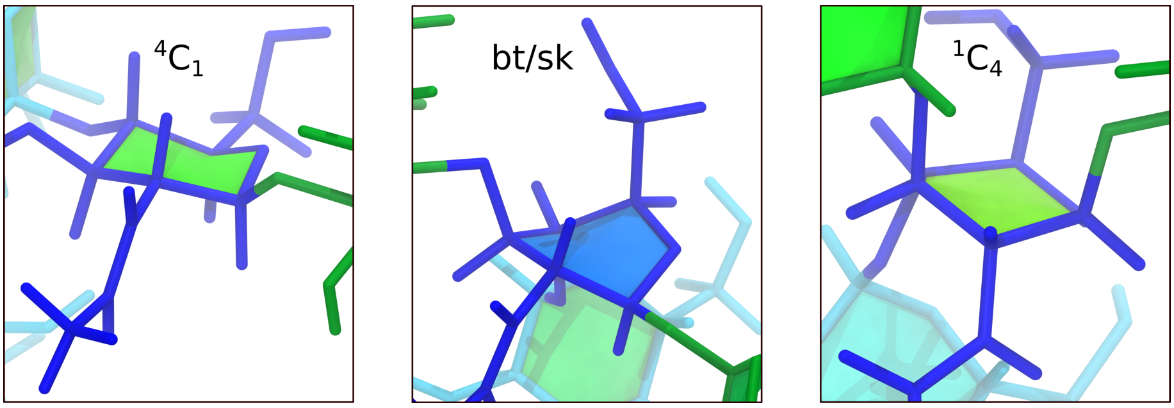


Figure 6.8: The three common ring pucker conformers of β-D-Glc*p*NAc in *Sf* 7a; represented by two chair conformers ($^4C_1$ and $^1C_4$) and a boat/skew (bt/sk) conformer ($^{1,4}B/^1S_5$). β-D-Glc*p*NAc in *Sf* 7b adopts the same pucker states. Structures shown using the Licorice representation in VMD with GlcNAc (blue), Rha (green), and Glc (cyan).

the overall O-Ag backbone conformation. This is most notable in *Sf* 7b, which mains a compact conformation through most of the second half of the simulation; this closely correlates to, and is likely caused by, a β-D-Glc*p*NAc in a central repeating unit adopting a $^1C_4$ state through the same period.

The β-D-Glc*p*NAc rings rarely adopt the idealised pucker conformations as defined by the Cremer-Pople parameters. For example, the $^4C_1$ state rarely adopts a $\phi = 0°$ value. Rather it fluctuates within a range relatively close to this number. This is also the case for the $^1C_4$ and boat/skew conformations observed. Interestingly, the *Sf* Y pucker values are evenly distributed around the origin ($\phi = 0°, \theta = 0°$). However, in *Sf* 7a and 7b they rather populate a region between $\phi = 320°$ and $\phi = 40°$. This suggests that even when β-D-Glc*p*NAc is in $^4C_1$ it adopts a conformation that more closely aligns with the $^OH_5$, O_E, and $^OH_1$ states. Similarly the $^1C_4$ conformer is closer to the $\phi = 270°$ and $\phi = 360°$ regions corresponding to $^1H_2$ and ^3E respectively.

The large variations observed in the backbone conformation as well as the glycosidic linkage torsion angles can, at least in part, be attributed to puckering of the β-D-Glc*p*NAc rings across the different repeating units. Puckering into boat/skew and inverted chair conformers is only observed in the β-D-Glc*p*NAc residues and not in the α-L-Rha*p* or α-D-Glc*p* moieties. Further, analysis of both the *Sf* 7a and 7b trajectories do not reveal any intra- or intermolecular hydrogen bonding or hydrophilic/hydrophobic interactions that may induce puckering of the β-D-Glc*p*NAc ring.

6.3 Discussion

The simulations of *Sf* 7a and 7b showed significant β-D-Glc*p*NAc ring puckering across all the repeating units in both O-Ags, which affects the orientations of the H1-Rhap^{II} – H3-Rhap^{I} and H1-Rhap^{I} – H3-Glc*p*NAc glycosidic linkages. These results suggest that the presence of the disaccharide glucose side chain in *Sf* 7a and 7b has a localized effect on the backbone linkage orientations, which is consistent with effect of side chains and O-Ac substitutions of other *S. flexneri* O-Ags.[22–24].

Further, there is a relationship between glycosidic linkage orientation and the β-D-Glc*p*NAc ring pucker. The H1-Glc$^{\text{I}}$ – H3-Glc*p*NAc and H1-Glc$^{\text{I}}$ – H4–Glc*p*NAc torsion angles and the β-D-Glc*p*NAc pucker conformation is correlated for both *Sf* 7a and 7b O-Ags. In the torsion angle heat map there is a narrow primary population centred at $\phi, \psi \approx -50°, -60°$, which correlates to a $^4C_1$ β-D-Glc*p*NAc state. Additionally, there is a broad, more sparsely distributed stretch occupied between $\phi, \psi \approx -30°, +30°$ that occurs in conjunction with a boat/skew β-D-Glc*p*NAc pucker. In the MD simulation trajectory the H1-Glc$^{\text{I}}$ – H4-β-D-Glc*p*NAc torsion angle is more flexible when β-D-Glc*p*NAc is in a non-chair conformation.

MD simulations of 3RU *S. flexneri* 1a and 1b O-Ags,[15] which have the same β-D-Glc*p*NAc 3,4-disubstitution as *Sf* 7a and 7b, confirm that this specific glucosylation pattern has a major influence on the conformation of the H1-Rhap^{I} – H3-Glc*p*NAc linkage located within the same RU. The mean lengths of the *Sf* 1a and 1b O-Ag modelled are also shorter and r histogram profiles have a large range than those of other investigated *Sf* O-Ag segments. Although there is no analysis on ring puckering within the repeat units.

Puckering of β-D-Glc*p*NAc has not been observed to such an extent in other polysaccharide antigens.[8,15–24] Microsecond MD simulations of the β-D-Glc*p*NAc monosaccharide with the GLYCAM06 force field showed that the ring puckers, transitioning into the $^1S_3$, $^1S_5$, ^2S_O, $^OS_2$, and $^1C_4$ states.[63] However, the amount of time spent in non-$^4C_1$ states and the number of pucker conformers explored decreased with increased ring substitution. In *Sf* 7a and 7b the disaccharide glucose attached to the 4-position of β-D-Glc*p*NAc likely provides the necessary torque to pucker the ring and affects which pucker states are entered.

NMR measurements of *S. flexneri* O-Ags, in contrast, have not indicated any β-D-Glc*p*NAc ring pucker, although, the primary purpose of these studies were structure elucidation and not conformational analysis.[15,51,90] This would initially suggest that the unusual MD simulation conformations of *Sf* 7a and 7b may not be a realistic phenomenon but rather occur due to artifacts of the force field. However, an NMR and modelling study of 3,4-disubstituted glucopyranoside trisaccharides revealed "larger conformational changes with multiple energy minima" for the α-(1→4)-linkage compared to the α-(1→3)-linkage.[54], which is similar to results from the *Sf* 7a and 7b simulations. To further explore and better understand these observations we perform MD simulations on a series of 3,4-disubstituted β-D-Glc*p*NAc trisaccharides using two different carbohydrate optimised force fields.

Chapter 7. Conformation and Dynamics of 3,4-disubstituted β-D-Glc*p*NAc trisaccharides: A Comparison of Two Force Fields

In Chapter 6 we revealed unusual ring puckering of the β-D-Glc*p*NAc residues in the MD simulations of *S. flexneri* 7a and 7b O-Ags. Puckering has not been observed in simulations of Shigella O-Ags or other bacterial polysaccharides.[15,16,22–25] While pyranose ring pucker has been widely modelled with QM,[64,216,274,275] and MD simulations[63,66,73,155,222,276,277] we find no immediate rationale for the β-D-Glc*p*NAc ring pucker observed in *Sf* 7a and 7b. We hypothesis that that the ring flips may be due to the 3,4-disubstitution pattern of β-D-Glc*p*NAc, which causes an increase in steric strain around the residue and induces puckering of the ring.

In this chapter we further investigate the observations in the *Sf* 7a and 7b simulations to determine the conditions that cause β-D-Glc*p*NAc ring pucker. We compare the conformations, dynamics, and ring pucker free-energies of selected 3,4-disubstituted β-D-Glc*p*NAc trisaccharides with two carbohydrate force fields: CHARMM36 and GLYCAM06. The trisaccharides include β-D-Glc*p* anomers linked at the 3- and 4- positions of either β-D-Glc*p* or β-D-Glc*p*NAc (Figure 7.1). The 3- and 4- positions are substituted with β-D-Glc*p*, as the residue has more mature and better defined force field parameters, instead of α-L-Rha*p* as found in *Sf* 7a and 7b, which has known inconsistencies.[199,214]

We model the central glucose residue with and without a 2-NAc substitution as well as all possible α and β anomeric configurations. This was done to measure and compare the conformation, dynamics, and energies of the 3,4-disubstituted trisaccharides across three metrics: (1) different force fields; (2) varying anomeric configurations; and (3) with and without 2-N-acetylation on β-D-Glc*p*NAc. We also model β-D-Glc*p*NAcOMe (1-O-methylation) in aaGN to compare the effect of a fully substituted residue on ring pucker conformation.

Metadynamics simulations of the trisaccharides will allow us to calculate PMF energies between the different pucker states and compare them to the atypical ring pucker conformations

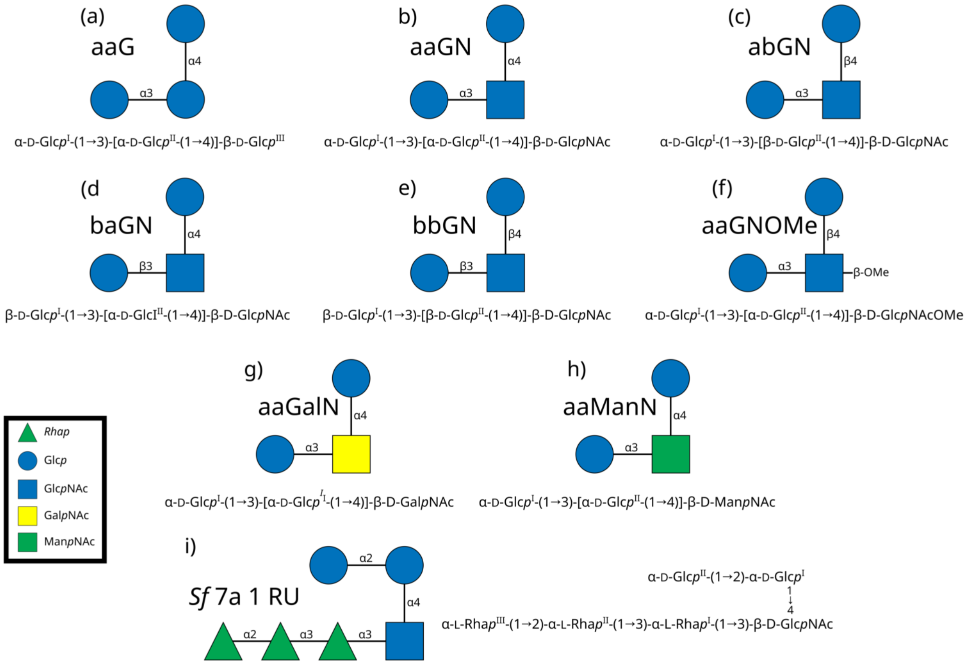


Figure 7.1: Structures of the trisaccharides and *S. flexneri* 7a O-Ag 1 RU modelled. Each structure is represented in SNFG notation and standard nomenclature.

and associated free energies observed in the MD simulations. The results should also provide a better understanding of the current force field parameters and how they affect the energies of, and therefore transitions between, different pucker states. Furthermore, the study may also give further insight into *β*-D-Glc*p*NAc conformations observed in the MD simulations and could potentially aid in a future force field optimisation process.

Additional simulations are performed with 3,4-disubstituted *β*-D-Gal*p*NAc and *β*-D-Man*p*NAc trisaccharides. These residues are compared with *β*-D-Glc*p*NAc for two reasons: (1) they occur frequently in O-Ags and other cell surface polysaccharides and (2) each residue only differs in the orientation of a single group when compared to *β*-D-Glc*p*NAc — the 4-position in *β*-D-GalNAc is axial and so is the 2-position for *β*-D-ManNAc. Finally, we perform MD simulations on a single repeating unit of *Sf* 7a with CHARMM36 and GLYCAM06, to compare and contrast the effects of the different force fields on repeating unit conformation.

7.1 Methods

We follow a similar approach for the trisaccharide MD simulations as we did for the modelling of *Sf* 7a and 7b. We use preferred orientations for the glycosidic linkages, precalculated using solution metadynamics simulations,[228,278] to build the trisaccharides. For simulations using the CHARMM36 force field, structures were built using CarbBuilder v2.1.34.[228,255] Structures modelled with the GLYCAM06 force field were constructed using the glycam.org carbohydrate builder[278] with the default glycosidic torsion angles.

7.1.1 Simulation setup and parameters

All simulations* (MD and metadynamics) were conducted in solution using NAMD v2.13.[256] The TIP3P water model was employed to simulate the aqueous solution.[171] Prior to and after solvation, both metadynamics and MD simulations followed a standard NAMD protocol consisting of 10,000 steps of energy minimization. An all-atom force field, either a modified version of CHARMM36[55–58,60] or GLYCAM06,[62] was used in the simulations.

The NAMD velocity-verlet integrator[163] and long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) summation method, with a κ value of 0.20 Å and PME grid dimensions of 8 nm for the simulations. All simulations were performed with a 34 Å^3 water box. Non-bonded interactions were truncated, and a switching function was applied between 1.2 nm and 1.5 nm for groups with integer charge. The 1-4 interactions were not scaled in the CHARMM36 simulations as recommended by the CHARMM36 force field.[257] All simulations used periodic boundary conditions. Temperature was controlled with the Nose-Hoover thermostat[258,259] method at 300 K

7.1.2 MD simulations

All-atom, unbiased solution MD simulations with a time step of 1 fs, were run on 3,4-disubstituted β-D-Glc and β-D-Glc*p*NAc trisaccharides. Pressure was controlled with the Langevin piston barostat[260] method under NPT ensemble at 1 atm. The NPT ensemble was chosen over the NVT ensemble due to the recommendation of constant pressure for periodic simulations in NAMD.[261]

Simulations were initiated with all pyranose rings in the $^4C_1$ conformer. All trisaccharide MD simulations were performed for 2 μs, with the exception of aaGN, which was modelled for 4 μs. A further all-atom, unbiased MD simulation was performed on a 1RU polysaccharide of

---

*All molecular modelling simulations were performed at the University of Cape Town High Performance Computing Facility.

Sf 7a for 1 μs using CHARMM36 and GLYCAM06 force fields for comparative purposes. Each system was equilibrated in the first 100 ns of the simulation.

7.1.3 Metadynamics simulations

Metadynamics simulations were used to calculate the Cremer-Pople ϕ, θ potential of mean force of puckering for the β-D-Glc*p* ring (in aaG) and β-D-Glc*p*NAc ring (in aaGN, baGN, abGN, and bbGN). The simulations were performed with NAMD v2.13[256] compiled with the open-source, community-developed PLUMED v2.6 library.[279–281]

Post solvation of the trisaccharide each system was equilibrated in a two-step process. First, the system was heated in a cycle from 10 K to 300 K at 10 K increments, with a 10000 step energy minimisation and a 100 ps NPT MD simulation at each step. The second equilibration step was a 2 ns MD simulation under NVT. The metadynamics simulations were performed at a 0.1 fs timestep for between 25 ns and 50 ns, under NVT as per PLUMED recommendations,[279–281] until no new minima were found and the existing minima did not shift by 5° for a period of at least 5 ns.

The PLUMED sum-hills algorithm[279–281] was used to convert the energies from the hills file into potential of mean force values with *mintozero* enabled, $bin = 179$ in both dimensions, with minimum and maximum values of $0, 0$ and $2pi, pi$ respectively.

7.1.4 Analysis and visualisation

For visualising and analyzing the data, we used Visual Molecular Dynamics (VMD) v1.9.2[262] or Python scripts using the MDAnalysis library.[263,264] The molecular conformations were rendered using the Tachyon VMD engine and visualised using the van der Waals or Licorice representations. When necessary, the PaperChain visualisation algorithm[265,266] was employed to emphasize the glycan rings.

The initial 100 ns of all simulations were considered as an equilibration phase and were excluded from subsequent analyses, unless otherwise stated. When analysing pyranose ring pucker conformations we do not consider half-chair and envelope states as they are very short lived and we do not sample the MD trajectory at a sufficiently high frequency to perform an adequate analysis.

Simulation convergence

We use block averaging analysis to assess simulation convergence[267,268] in a manner similar to previous carbohydate modelling studies.[22–24,26,27,199,252] The block averaging algorithm divides a simulation trajectory consisting of N frames into M blocks, each containing n frames in length.

$$N = nM \tag{7.1}$$

A single measurable parameter (e.g. radius of gyration, root mean square deviation or end-to-end distance) is selected and an average is calculated for each block. The block length, n, is incrementally increased. At each value of n, the block averages are recalculated. We then compute the mean for each block, the mean of all the block means, and the standard error of all the block means (the block standard error — BSE). The BSE is plotted as a function of the block size and the simulation is indicated to have converged once the BSE asymptotes to a plateau.

Torsion angles

The atoms that form the glycosidic linkage torsion angles are defined as:

$$\phi = \mathrm{H_1 - C_1 - O_1 - C'_x}$$
$$\psi = \mathrm{C_1 - O_1 - C'_x - H'_x}$$

O1 and O6 rotamer conformations are defined by $C(n+1) - C(n) - O(n) - H$ as used in previous studies.[269] Hydroxyl orientations were binned as gauche (g^-), anti (t) or gauche (g^+) centred on -60, ±180 and $+60°$, respectively. Hydroxymethyl conformations (tg, gt and gg) were defined by the torsions O_6-C_6-C_5-O_5 as has also been done in similar studies.[63]

Pyranose ring pucker

The Cremer-Pople pucker parameters for each ring were calculated according to a protocol provided by Kooijman,[270] which is based on the original Cremer-Pople derivations.[71] The ranges for the Cremer-Pople azimuth (θ) angle used to distinguish between chair and boat/skew conformers for pyranose rings as used in this study are as follows: $^4C_1$: $0° <= \theta < 30°$, boat/skew: $75° <= \theta < 105°$, and $^1C_4$: $150° < \theta <= 180°$. Transition between ring pucker states is calculated as a change from chair to boat/skew (or vice-versa) and where such a change is maintained for at least 200 ps. We do not anaylse half-chair and envelope states as they are very short lived states and we do not sample the MD trajectory at a sufficiently high frequency.

Free energies of the β-D-Glc*p*$^{\mathrm{III}}$ and β-D-Glc*p*NAc ring pucker were calculated from simulated equilibrium data of the trisaccharides studied using the CHARMM36 and GLYCAM06 force fields. The $\Delta G = -RTln|K_{eq}|$ relationship is used to compute the free energy differences; where ΔG is the change free energy, R is the gas constant, T is the Temperature, and K_{ew} is the equilibrium constant.

The number of pyranose ring pucker transitions and average lifetime of a transition were calculated for the trisaccharide simulations. A transition between ring pucker states is defined here as a change from chair to boat/skew (or vice-versa) and where such a change is maintained for at least 200 ps. For each simulation we counted the number of transitions per ring. The average lifetime (the time a pyranose ring spends in a particular pucker state) was measured by calculating the time between ring pucker transitions.

7.2 Results

We analyse the MD simulations of the trisaccharides across several measures. We first determine simulation convergence, followed by comparison of the glycosidic linkage torsion angle orientations, and then analysis of the ring pucker dynamics and conformations. We continue with the metadynamics simulation analysis of the trisaccharides by studying the ring pucker free energies and transition states. Finally, we compare and contrast the CHARMM36 and GLYCAM06 MD simulations of 1 RU of *Sf* 7a.

7.2.1 Simulation convergence

We calculate the Block Standard Error (BSE) as a function of the radius of gyration to analyse simulation convergence (Figure 7.2). The BSE for aaG reaches a plateau at block sizes well below 100 ns in both the CHARMM36 and GLYCAM06 simulations, which suggests that the simulations converge. The BSE for aaGN does not plateau within the same block size for both force fields; indicating that these simulations have not converged.

The remaining trisaccharides (baGN, abGN, bbGN) reach convergence during their respective simulations (Figure A.2). Interestingly, however, bbGN from the GLYCAM06 simulation has a higher BSE and takes longer to plateau within a 100 ns block size compared to the other trisaccharides.

Typically, the BSE for small molecules (such as the trisaccharides studied) would be expected to converge during 1-2 μs MD simulations. The failure of the aaGN simulations to do so is likely due to their highly flexible glycosidic linkages and extensive puckering of the β-D-Glc*p*NAc ring. We extended the aaGN simulations to 4 μs but still did not achieve convergence

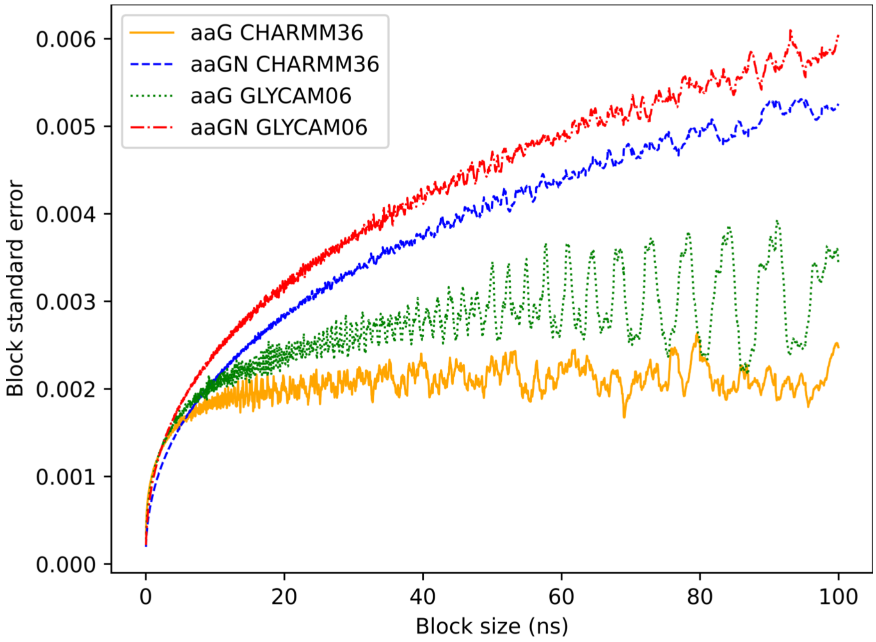


Figure 7.2: Block Standard Error (BSE) of aaG and aaGN from the CHARMM36 and GLYCAM06 simulations. The aaG simulations (green, orange) converge. The aaGN simulations (red, blue) do not converge.

within this time period. Longer simulations may be required for the β-D-Glc*p*NAc to achieve equilibrium and simulation convergence.

7.2.2 Glycosidic linkage torsion angles

Figure 7.3 provides a comparison of the glycosidic linkage torsion angles for the trisaccharides modelled and both force fields. The 1→3 and 1→4 glycosidic linkages of the trisaccharides adopt orientations that are different for all the molecules as well as across force fields. Both linkages consistently populate multiple energy minima regions, although, they appear to be more flexible in aaG and aaGN and more constrained in baGN, abGN, and bbGN.

The ϕ, ψ torsion angles for the aaG H1-Glc$^{\mathrm{I}}$ – H3-Glc$^{\mathrm{III}}$ glycosidic linkage are similar between the two force fields (Figure 7.3: a, c). In both, ϕ primarily remains within the $+30^\circ$ to $+60^\circ$ range. The ψ angle is more flexible; fluctuating between $+20^\circ$ and $+80^\circ$ (CHARMM36), and -30° and $+60^\circ$ (GLYCAM06). The ϕ, ψ glycosidic linkage torsion angle time series (Figure A.3: a, d) also highlight a small number of transitions to the anti-orientation in the simulations of both force fields.

In contrast, the ϕ, ψ torsion angles for the H1-Glc$^{\mathrm{II}}$ – H4-Glc$^{\mathrm{III}}$ glycosidic linkage (Figure 7.3:

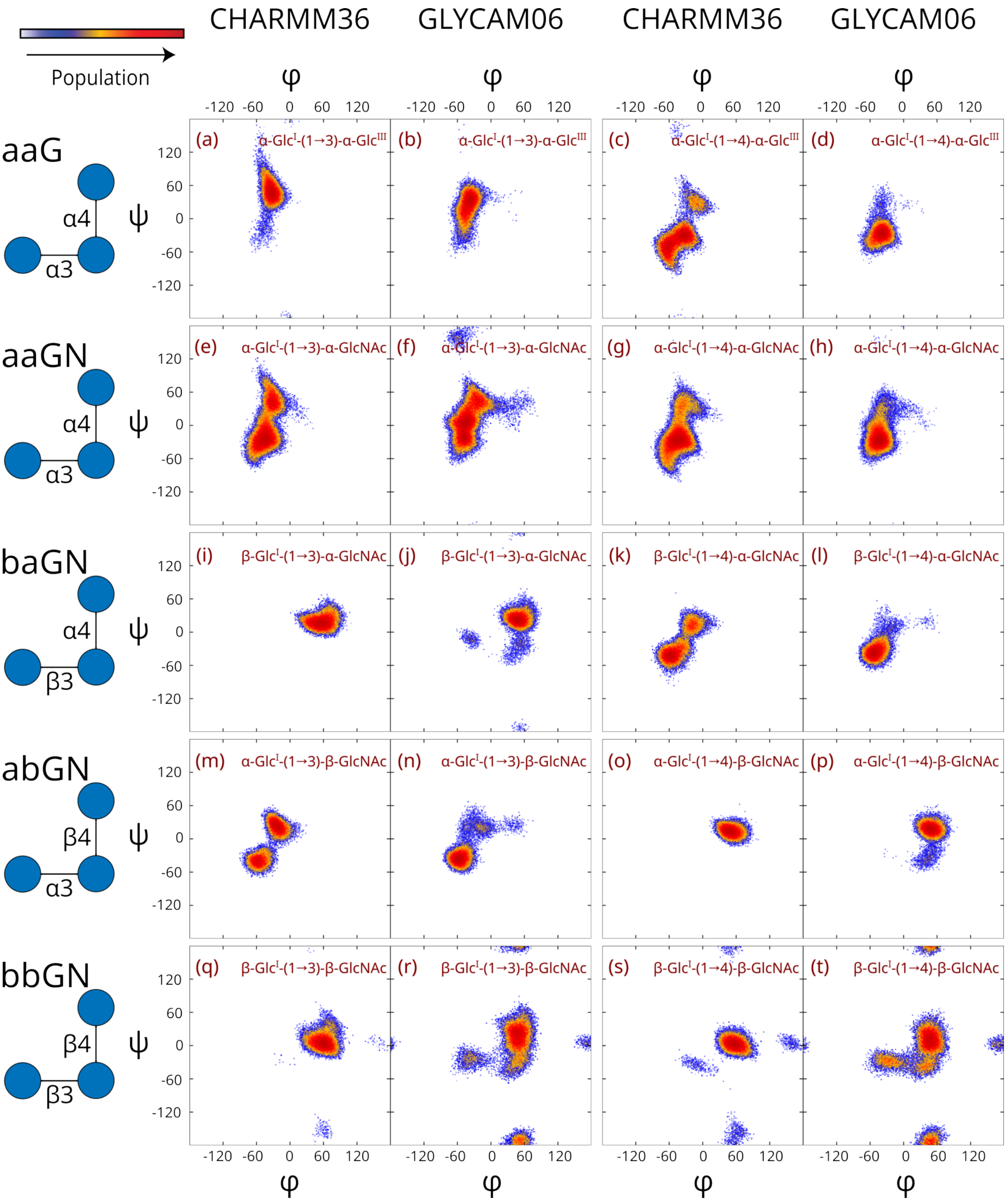


Figure 7.3: Scatter plots for the glycosidic linkage torsion angles (ϕ, ψ) in aaG, aaGN, baGN, abGN and bbGN. The colour scale on the upper-left indicates the relative occupancy of the torsion angles during the simulations.

b, d) are different between the two force fields. In the CHARMM36 simulation the region between $\phi, \psi = -90°, -90°$ and $\phi, \psi = 0°, 90°$ is highly populated. There is also a less populated region at $\phi, \psi = -20°, +30°$. In the GLYCAM06 simulation, the glycosidic linkage adopts a single orientation of $\phi, \psi = -40°, -30°$.

There are large differences in the aaGN H1-Glc$^{\text{I}}$ – H3-Glc*p*NAc ϕ, ψ glycosidic linkage torsion angles between the two force fields (Figure 7.3: e, g). The ϕ torsion in the CHARMM36 simulation remains relatively constant throughout, maintaining a value between $\phi = 0°$ and $\phi = -60°$. With GLYCAM06 the ϕ torsion also largely remains within the same range but there are significantly more transitions to the $+\phi$ region (Figure A.3: e, g). The ψ torsion is more flexible in both force fields and occupies the region between $-60°$ and $-60°$.

The aaGN H1-Glc$^{\text{I}}$ – H4-Glc*p*NAc linkage adopts very similar orientations to the 1→3 linkage in the CHARMM36 simulation (Figure 7.3: f). In the GLYCAM06 it is more constrained adopting a single orientation at $\phi, \psi = -60°, -50°$ (Figure 7.3: h).

The H1-Glc$^{\text{I}}$ – H3-Glc*p*NAc linkage of baGN populates a single region at $\phi, \psi = 60°, 30°$ for both force fields (Figure 7.3: i, k). The H1-Glc$^{\text{II}}$ – H4-Glc*p*NAc linkage for baGN has two population centres with CHARMM36 (Figure 7.3: j). A major one at $\phi, \psi = -60°, -45°$ and a minor one at $\phi, \psi = -20°, +20°$. The H1-Glc$^{\text{II}}$ – H4-Glc*p*NAc linkage for the GLYCAM06 simulation populates a single region at $\phi, \psi = -60°, -30°$ (Figure 7.3: l).

The ϕ, ψ torsion angles of the abGN H1-Glc$^{\text{I}}$ – H3-Glc*p*NAc and H1-Glc$^{\text{II}}$ – H4-Glc*p*NAc glycosidic linkages also adopt a narrow, well defined region (Figure 7.3: l, p). The ϕ, ψ torsion angle of the H1-Glc$^{\text{I}}$ – H4-Glc*p*NAc glycosidic linkage has two population centres with the CHARMM36 simulation centred at $\phi, \psi = -60°, -40°$ and $\phi, \psi = -20°, +30°$. With GLYCAM06 the ϕ, ψ torsion angle is located at the same regions, although, the centre at $\phi, \psi = -20°, +30°$ is very sparsely populated. The abGN H1-Glc$^{\text{II}}$ – H4-Glc*p*NAc glycosidic linkage is centred at $\phi, \psi = +60°, +20°$ and $\phi, \psi = +50°, +30°$ for CHARMM36 and GLYCAM06 respectively.

The ϕ, ψ torsion angles for the bbGN glycosidic linkages occupy similar space for both force fields (Figure 7.3: q, t). The H1-Glc$^{\text{I}}$ – H3-Glc*p*NAc glycosidic linkage is at $\phi, \psi = +60°, +30°$ and for H1-Glc$^{\text{II}}$ – H4-Glc*p*NAc at $\phi, \psi = +60°, +20°$. There is also populated region at $\phi = -180°$ but this is small, very sparsely populated and correlates to a non-$^4C_1$ β-D-Glc*p*NAc conformation.

Across all the trisaccharides simulated observations suggest a correlation between the $1 \rightarrow 3$ glycosidic linkage torsion angle with β-D-Glc*p*NAc ring pucker. However, the orientation of the $1 \rightarrow 4$ linkage does not appear to be as closely related to ring conformation. This dynamic is consistent with the relationship between glycosidic torsion angles and ring pucker conformations from the *Sf* 7a and 7b simulations. The anomeric configuration of both linkages also appears to be important. Further analysis of the ring pucker dynamics and conformation will allow us to compare pucker states of the different trisaccharides and identify the importance of glycosidic linkage orientation and anomeric configuration.

7.2.3 Ring pucker of β-D-Glcp^{III} and β-D-Glc*p*NAc

Ring pucker dynamics of β-D-Glcp^{III} and β-D-Glc*p*NAc

We begin by comparing the ring pucker dynamics of β-D-Glcp^{III}in aaG with β-D-Glc*p*NAc in aaGN, which have distinct differences. For both force fields the β-D-Glc*p*NAc ring puckers to a significantly greater extent than the β-D-Glcp^{III} ring. The ring pucker dynamics of β-D-Glc*p*NAc in baGN, abGN, and bbGN are similar across the three trisaccharides but all differ greatly to both β-D-Glcp^{III} in aaG and β-D-Glc*p*NAc in aaGN (Figure 7.4 and Table 7.1).

In both the CHARMM36 and GLYCAM06 simulations of aaG equilibrium of the β-D-Glcp^{III} pucker is achieved well within the 2 μs simulation time. The $^4C_1$ chair is the overwhelmingly dominant conformer for both force fields. Boat/skew and $^1C_4$ are adopted for substantially shorter periods.

In the aaG CHARMM36 simulation (Figure 7.4-b) the $^4C_1$ conformer ($\theta \approx 15°$) occurs approximately 98.4% of the time. Boat/skew ($\theta \approx 88°$) makes up the remainder and there are no $^1C_4$ conformers. There are a total of 6 $^4C_1$ $\leftrightarrow$boat/skew transitions[†] at an average forward and backward rate[‡] of 3 μs^{-1}. The average lifetime of the $^4C_1$ pucker is approximately 253 ns - and 5 ns for the boat/skew conformers. During the short periods that β-D-Glcp^{III} is in the boat/skew pucker state ϕ adopts two values: 90° ($^5S_1$) and 150° (^3S_O).

In the GLYCAM06 simulation of aaG (Figure 7.4-d), β-D-Glcp^{III} adopts $^4C_1$ for about 87.5% of the time, with an average of $\theta \approx 13°$. It also populates the boat/skew and $^1C_4$ for extended periods with approximate θ averages of 86° and 167° respectively. There are 32 forward $^4C_1$ $\leftrightarrow$boat/skew transitions at a rate of 18 μs^{-1} and 27 backward transitions at a rate of 13.5 μs^{-1}. In comparison, there are fewer $^1C_4$ $\leftrightarrow$boat/skew transitions, with 13 forward transitions in total, at a rate of 6.5 μs^{-1}, and 10 backward transitions at a rate of 5 μs^{-1}.

These results highlight the differences of β-D-Glcp^{III} modelling between the two force fields. As there are fewer boat/skew transitions in the CHARMM36 simulations it suggests that the

---

[†] A transition between ring pucker states is defined here as a change from chair to boat/skew (or vice-versa) and where such a change is maintained for at least 200 ps.

[‡] A forward transition is the movement from a pucker state to a different pucker state for a defined transition. A backward transition is the reverse movement for the same defined transition.

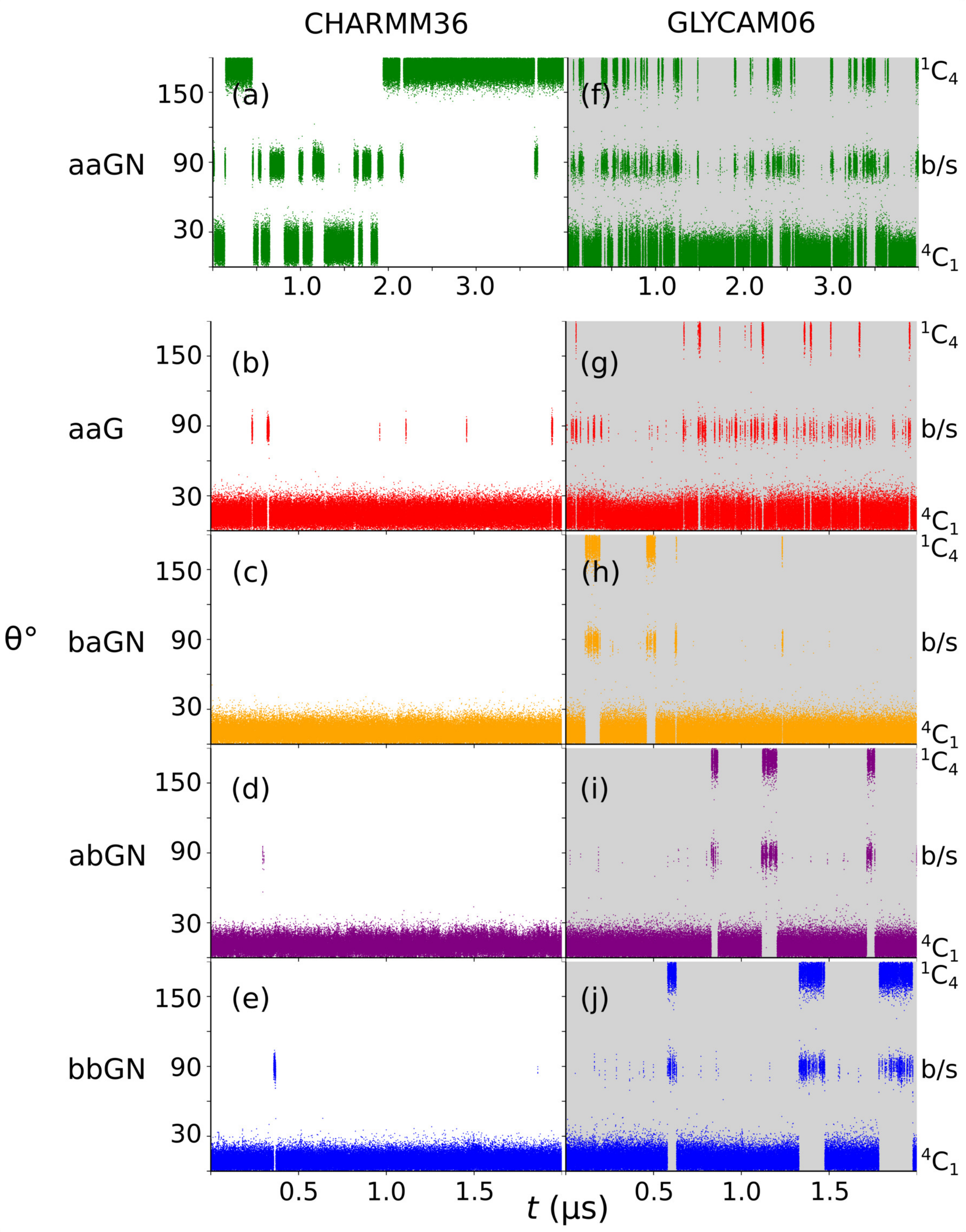


Figure 7.4: Cremer-Pople pucker azimuth (θ) parameter for β-D-Glcp^{III} (in aaG) and in β-D-Glc$p$NAc (in aaGN, baGN, abGN, bbGN) from the CHARMM36 and GLYCAM06 force fields. Simulation time is 4 μs for aaGN and 2 μs for for all other trisaccharides. Note that when θ is at 0° or 180° (i.e. the poles of the Cremer-Pople sphere) there is no direct correlation between ϕ and pucker conformation.

Table 7.1: Ring pucker parameters for β-D-Glc*p*$^{\text{III}}$ in aaG and β-D-Glc*p*NAc in aaGN, baGN, abGN, and bbGN. The number of transitions between the different pucker states, the rate of transition, the average lifetime of each pucker conformer, and the average Cremer-Pople ring pucker amplitude (Q) are provided.

Pucker Parameters	**CHARMM36**					**GLYCAM06**				
	aaG	aaGN	baGN	abGN	bbGN	aaG	aaGN	baGN	abGN	bbGN
Number of $^4C_1 \rightarrow$ boat/skew transitions	6	12	0	1	1	32	28	4	3	5
Number of boat/skew $\rightarrow$ $^4C_1$ transitions	6	10	0	1	1	27	52	4	3	3
Number of $^1C_4 \rightarrow$ boat/skew transitions	0	3	0	0	0	13	16	2	3	9
Number of boat/skew $\rightarrow$ $^1C_4$ transitions	0	3	0	0	0	10	15	2	3	8
Average rate of $^4C_1 \rightarrow$ boat/skew transitions (μs^{-1})	3	3	0	0.5	0.5	18	7	2	1.5	2.5
Average rate of boat/skew $\rightarrow$ $^4C_1$ transitions (μs^{-1})	2.5	2.5	0	0.5	0.5	13.5	13	2	1.5	1.5
Average rate of $^1C_4 \rightarrow$ boat/skew transitions (μs^{-1})	0	8.5	0	0	0	6.5	8	1	1.5	4.5
Average rate of boat/skew $\rightarrow$ $^1C_4$ transitions (μs^{-1})	0	7.5	0	0	0 0	5	7.5	1	1.5	4
Average lifetime of $^4C_1$ (μs)	25.3	0.6	2	2	1.95	47	0.54	0.73	0.56	0.52
Average lifetime of $^1C_4$ (μs)	0	0.89	0	0	0.45	2.7	50	0.03	0.04	0.2
Average lifetime of boat/skew (μs)	5	0.08	0	0	0.05	1.3	0.42	0.05	0.05	0.8
Average Q	0.57	0.53	0.56	0.57	0.59	0.36	0.56	0.53	0.53	0.54

energy barrier between these states is greater when compared to GLYCAM06. Further, the $^1C_4$ conformer is not accessed at any point in the CHARMM36 simulation. In the GLYCAM06 simulation, however, β-D-Glc*p*$^{\text{III}}$ repeatedly enters the $^1C_4$ state for significant periods of time. This also indicates that the energy barrier for the $^1C_4$ $\leftrightarrow$boat/skew transition in the GLYCAM06 simulations is significantly lower than that of CHARMM36. The lower interconversion barriers for GLYCAM06 is consistent with previous simulations of β-D-glucopyranose.[217]

In contrast, β-D-Glc*p*NAc in aaGN has a lower population of $^4C_1$ chair (Table 7.1). In the CHARMM36 simulation (Figure 7.4: e, f), between 0 μs and 2.1 μs β-D-Glc*p*NAc makes a single transition into $^1C_4$ and remains in this conformer for approximately 0.3 μs. It then regularly fluctuates between $^4C_1$ and boat/skew pucker states. Within the first 2.1 μs there are a total of 12 forward $^4C_1 \leftrightarrow$boat/skew transitions and 5 backward transitions. The average $^4C_1$ and boat/skew lifetime during this period is 0.26 μs and 0.23 μs respectively. In the last 1.9 μs of the simulation there is a significant change in the pucker dynamics of β-D-Glc*p*NAc; it does not adopt the $^4C_1$ state and there are therefore no $^4C_1 \leftrightarrow$boat/skew transitions. During the latter half of the simulation β-D-Glc*p*NAc spends over 90% of the time in the $^1C_4$ conformer with an average lifetime of 0.96 μs. There are also only 2 forward and backward $^1C_4 \leftrightarrow$boat/skew transitions.

The GLYCAM06 simulation of aaGN (Figure 7.4: f, h) show markedly different ring pucker dynamics in β-D-Glc*p*NAc. Unlike in the CHARMM36 simulation the $^4C_1$ conformer is dominant with an average lifetime of 0.54 μs. There are 28 forward $^4C_1\leftrightarrow$boat/skew transitions at a rate of 7 μs^{-1}. The are 52 backward transitions at a rate of 13 μs^{-1}.

The Cremer-Pople pucker amplitudes (Q) in the CHARMM36 and GLYCAM06 simulations are similar for β-D-Glc*p* in aaG (Figure A.6). While in the $^4C_1$ state Q is in the range of 0.5-0.6 (approximately 88% of the simulation time). While in the boat/skew conformer state Q is approximately between 0.7 and 0.8 (approximately 11% of the simulation time). The distribution of Q between these ranges is similar to that of the pucker conformer distribution itself. The aaGN simulation shows similar absolute values in the ranges of Q for both force fields. When in the $^4C_1$ or $^1C_4$ canonical pucker Q lies within the 0.5 - 0.6 range and when in a boat/skew state Q is within the 0.7 - 0.8 range.

Simulations of baGN, abGN and bbGN show some similarities and a few differences in the ring pucker dynamics of β-D-Glc*p*NAc when compared to that in aaGN. In the CHARMM36 simulation of baGN, β-D-Glc*p*NAc remains solely in the $^4C_1$ state. However, in the GLYCAM06 simulation β-D-Glc*p*NAc adopts both chair conformations as well as boat/skew pucker states. It maintains a $^4C_1$ conformer for approximately 90% of the simulation time white boat/skews (4%) and the $^1C_4$ chair (6%) make up the remainder.

In the GLYCAM06 simulation of baGN there are a total of 7 forward $^4C_1\leftrightarrow$boat/skew transitions at an average rate of 3.5 transitions μs^{-1} and 6 backward transitions at an average of 3 μs^{-1}. There are 4 forward and an equal number of backward $^1C_4\leftrightarrow$boat/skew transitions at an average rate of 2 μs^{-1}. However, it should be noted that all but one $^4C_1$ $\leftrightarrow$boat/skew transition occurred within the first microsecond of the simulation indicating that β-D-Glc*p*NAc ring pucker may have taken a few hundred nanoseconds to equilibrate. Although, this cannot be confirmed unless the simulation time is extended for at least several more microseconds, which is outside the scope of this work.

In the abGN CHARMM36 simulation, β-D-Glc*p*NAc makes a single short-lived transition to the boat/skew pucker states; while in the bbGN CHARMM36 simulation β-D-Glc*p*NAc similarly makes two short lived $^1C_4\leftrightarrow$boat/skew transitions. The results indicate that the ring pucker dynamics of the baGN, abGN and bbGN are all similar with either no transitions or short-lived transitions out of the $^4C_1$ chair.

The β-D-Glc*p*NAc pucker dynamics in the GLYCAM06 simulations of abGN and bbGN are also similar to that of baGN. The $^4C_1$ chair remains the overwhelmingly dominant conformer but is present a little less frequently, at about 80-85% of the time. This is in stark contrast to the pucker dynamics off aaG and aaGN GLYCAM06 simulations where the boat/skew and $^{1}C_{4}$ conformers are much more prevalent.

Ring pucker conformers of β-D-Glc*p*$^{\mathrm{III}}$ and β-D-Glc*p*NAc

The ring pucker conformation and dynamics of β-D-Glc*p*NAc in aaGN is notably different to that of β-D-Glc*p*$^{\mathrm{III}}$ in aaG and to β-D-Glc*p*NAc in the other N-acetylated trisaccharides, adopting similar pucker conformations to those observed in the *Sf* 7a and 7b simulations. Figure 7.5 illustrates the percentage occupancies for the boat/skew pucker states in the trisaccharides and Figure 7.6 depicts the most common canonical puckers of β-D-Glc*p*$^{\mathrm{III}}$ in aaG and β-D-Glc*p*NAc in aaGN and the pseudorotational transition pathways.

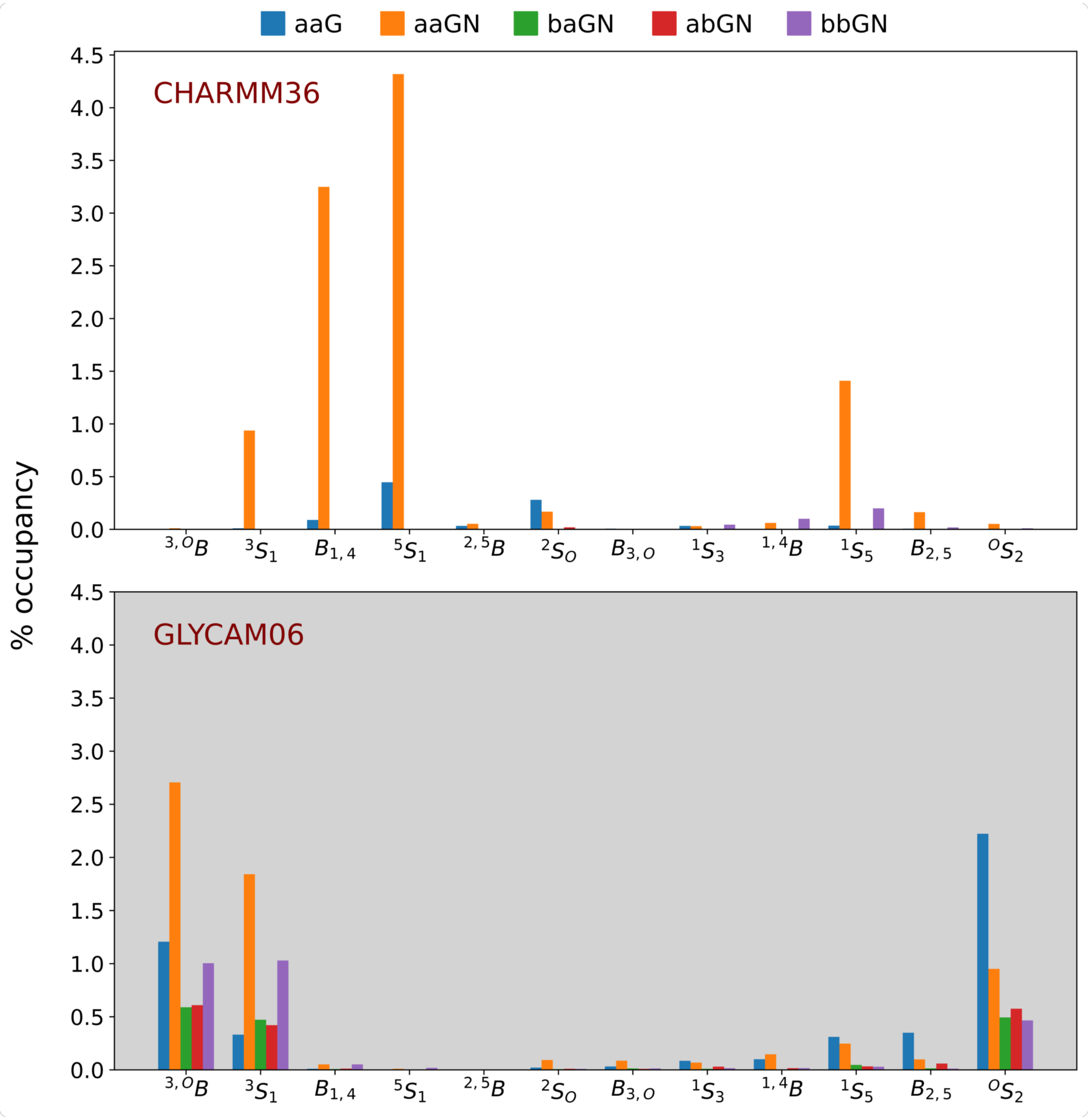


Figure 7.5: Occupancies of different boat/skew pucker conformers of β-D-Glc*p*NAc in aaG (blue), aaGN (orange), baGN (green), abGN (red), and bbGN (purple). In the CHARMM36 simulations β-D-Glc*p*NAc in aaGN has the greatest occupancy of boat/skew states, significantly more so than the other trisaccharides. In the GLYCAM06 simulations β-D-Glc*p*NAc of aaGN also has the greatest boat/skew occupancy but the other trisaccharides adoption of boat/skew conformations are significant as well.

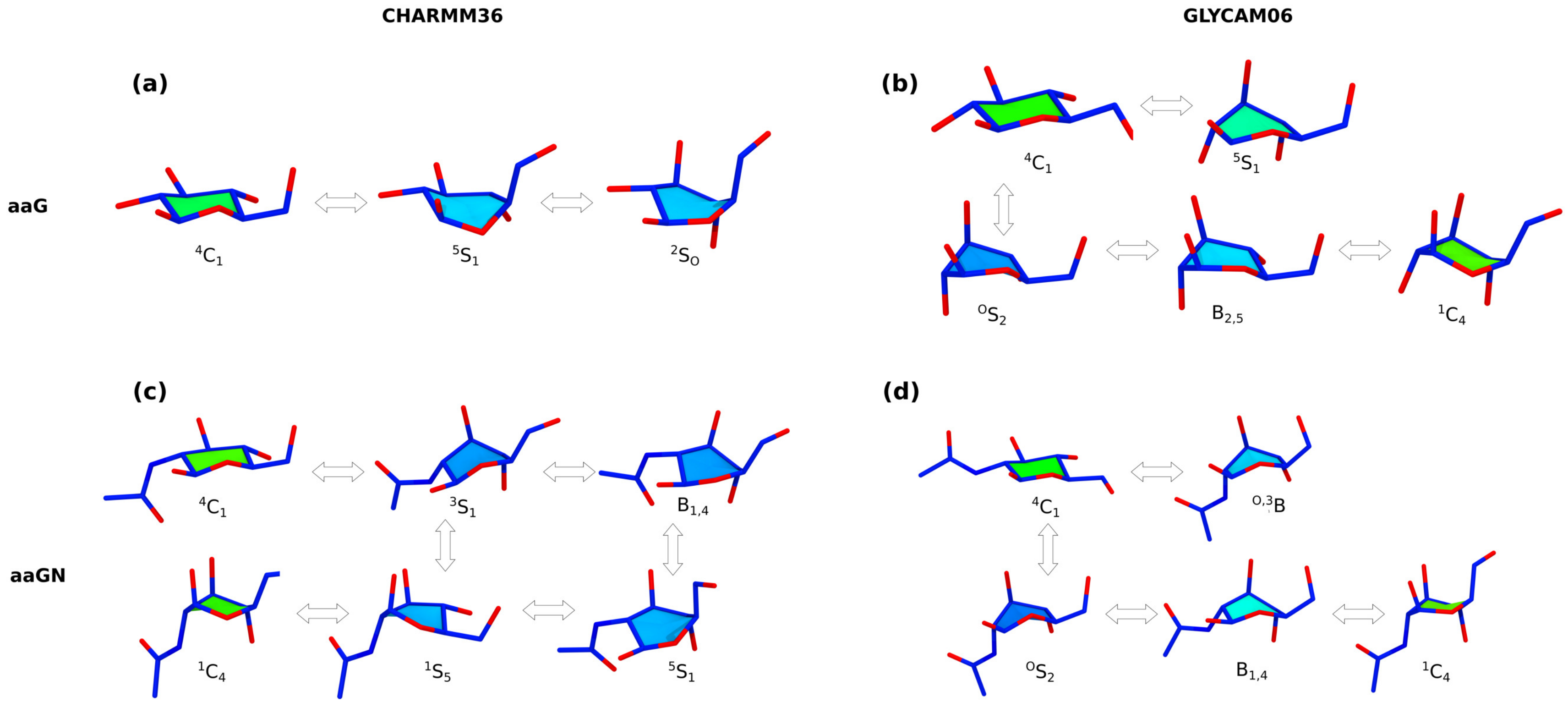


Figure 7.6: The most common canonical puckers of β-D-Glcp^{III} in aaG and β-D-Glc$p$NAc in aaGN and the pseudorotational transition pathways from the CHARMM36 and GLYCAM06 simulations. Structures shown using the Licorice representation in VMD.

The β-D-Glc$^{\mathrm{III}}$ ring pucker in aaG primarily maintains ${}^4\mathrm{C}_1$ (the lowest energy conformation) across both force fields. Different boat/skew pucker states as well as ${}^1C_4$ are also populated to differing extents. In the CHARMM36 simulations of aaG, the primary boat/skew puckers of β-D-Glc$^{\mathrm{III}}$ are ${}^5S_1$ ($\phi \approx 90°$, $\Delta G = +1.50$ kcal mol^{-1}) and ${}^2S_1$ ($\phi \approx 150°$, $\Delta G = +1.70$ kcal mol^{-1}). Transitions to and from the ${}^4\mathrm{C}_1$ chair occur primarily via the ${}^5S_1$ conformer by the ${}^4C_1 \leftrightarrow {}^5S_1 \leftrightarrow {}^2S_O \leftrightarrow B_{1,4}$ path (Figure 7.6-a).

In the GLYCAM06 simulation of aaG the ${}^4\mathrm{C}_1$ conformer of β-D-Glc$^{\mathrm{III}}$ transitions to different boat/skew conformers and ${}^1\mathrm{C}_4$ at a higher frequency (Figure 7.6-b). The more frequently populated non-chair conformers are ${}^OS_2$ ($\phi \approx 330°$, $\Delta G = +0.61$ kcal mol^{-1}) and ${}^{3,O}B$ ($\phi \approx 0/360°$, $\Delta G = +0.95$ kcal mol^{-1}). $B_{2,5}$, ${}^1S_5$, ${}^{1,4}B$, ${}^1S_3$, and $B_{3,O}$ are less occupied ($\Delta G \geq 1.3$ kcal mol^{-1}). The primary transition paths is by the ${}^4\mathrm{C}_1 \leftrightarrow {}^OS_2 \leftrightarrow {}^{3,O}B$ and ${}^4\mathrm{C}_1 \leftrightarrow {}^OS_2 \leftrightarrow B_{2,5} \leftrightarrow {}^1\mathrm{C}_4$ route indicating that transitions to the inverted chair occurs via the less dominant $B^{2,5}$ conformer.

Comparing the CHARMM36 simulation of aaGN, the lowest energy β-D-Glc*p*NAc ring conformer is between $B_{1,4}$ and ${}^5S_1$ ($\phi \approx 77°$). There are two primary pathways by which β-D-Glc*p*NAc in aaGN transitions between ${}^4\mathrm{C}_1$ ($\Delta G = +0.54$ kcal mol^{-1}) and ${}^1\mathrm{C}_4$ ($\Delta G = +0.32$ kcal mol^{-1}). The first, and more direct route, includes two skew intermediates: ${}^4C_1 \leftrightarrow^3 S_1 \leftrightarrow^1 S_5 \leftrightarrow^1 C_4$. The second and longer route occurs via a boat conformer: ${}^4C_1 \leftrightarrow^3 S_1 \leftrightarrow B_{1,4} \leftrightarrow^5 S_1 \leftrightarrow^1 S_5 \leftrightarrow^1 C_4$ (Figure 7.6-c). These unusually small differences between the free energies of the chair and non-chair conformers, relative to that of aaG, indicate that N-acetylation significantly changes the energy profile of the trisaccharide.

In contrast, the GLYCAM06 simulation of aaGN has ${}^4\mathrm{C}_1$ as lowest the energy conformer and ${}^1\mathrm{C}_4$ is $\Delta G = +0.55$ kcal mol^{-1} (Figure 7.6-d). The transition between the two pucker chair conformers occurs via a single path: ${}^1C_4 \leftrightarrow B_{2,5} \leftrightarrow^O S_2 \leftrightarrow^{3,O} B \leftrightarrow^1 C_4$. All boat and skew conformers are visited, and with the exception of the ${}^{2,5}B$, all are within $+2.00$ kcal mol^{-1} of ${}^4\mathrm{C}_1$.

The ring puckers of β-D-Glc*p*NAc in baGN, abGN, and baGN do not adopt the same conformations as aaG and aaGN for both force fields. These three trisaccharides either do not pucker out of ${}^4C_1$ or only do so for shorter periods of time. In the CHARMM36 simulations of baGN and abGN β-D-Glc*p*NAc remains entirely in ${}^4\mathrm{C}_1$ for the length of the simulation. However, bbGN makes a single transition to ${}^1S_5$ and ${}^{1,4}B$ for approximately 5 ns before returning to ${}^4\mathrm{C}_1$.

In the GLYCAM06 simulations of baGN the only notable non-chair conformations are ${}^3S_1$ ($\Delta G = +1.10$ kcal mol^{-1}) and ${}^{3,0}B$ ($\Delta G = +0.10$ kcal mol^{-1}) although, these states are populated for less than 2% of the total simulation time. The pathway from ${}^4\mathrm{C}_1$ to ${}^1\mathrm{C}_4$ occurs via the higher energy ${}^1S_5$ pucker: ${}^4C_1 \leftrightarrow^{3,0} B \leftrightarrow^1 S_5 \leftrightarrow^1 C_4$.

The confomers adopted by β-D-Glc*p*NAc in the both the CHARMM36 and GLYCAM06 simulations of abGN are very similar to that of baGN. However, the computed relative energies of $^{3,0}B$ (+0.87 kcal mol^{-1}) and $^{1}S_5$ (ΔG = +0.96 kcal mol^{-1}) in the GLYCAM06 simulation are greater in comparison to CHARMM36 while $^{O}S_2$ (ΔG = +0.96 kcal mol^{-1}) and $^{1}C_4$ (ΔG = +0.87 kcal mol^{-1}) are lower in energy.

The bbGN simulations show a greater degree of β-D-Glc*p*NAc puckering for both force fields. The most prevalent boat/skew states in the CHARMM63 simulation are $^{3}S_1$ (ΔG = +1.70 kcal mol^{-1}), $B_{1,4}$ (ΔG = +1.94 kcal mol^{-1}), and $^{5}S_1$ (ΔG = +1.70 kcal mol^{-1}). For GLYCAM06 the most frequented boat/skew conformers are $^{3}S_1$ (+0.57 kcal mol^{-1}) and $B_{1,4}$ (ΔG = +1.63 kcal mol^{-1}), which are both closer in energy to the ground state than in the CHARMM36 simulation. The $^{1}C_4$ chair (ΔG = +0.76 kcal mol^{-1}) is also populated in the GLYCAM06 simulation but not in the CHARMM36 simulation.

Figure 7.7 depicts relative occupancies of the different pucker states for all trisaccharides studied on flattened Cremer-Pople spheres. Figure 7.8 illustrates and Table A.2 lists the free energies of the observed chair, boat and skew conformers using the standard $\Delta G = |RTlnK_{eq}|$ relationship.

Similar to the *Sf* 7a and 7b simulations the ϕ and θ pucker values of β-D-Glc*p*$^{\text{III}}$ (in aaG) and β-D-Glc*p*NAc (in aaGN) are not evenly distributed around the origin when the rings are in the $^{4}C_1$ state. The rings populate the region between $\phi, \theta = 320°, 0°$ and $\phi, \theta = 40°, 0°$, which correspond to $^{O}H_5$, O_E, and $^{O}H_1$. In contrast, the ϕ and θ pucker values of β-D-Glc*p*NAc are evenly distributed around the origin for the CHARMM36 simulations of baGN, abGN, and bbGN, as well as all the GLYCAM06 simulations. This suggests that the $^{4}C_1$ conformer of β-D-Glc*p*NAc in these trisaccharides is not more closely aligned conformationally to a particular half-chair or envelope state for these trisaccharides.

These results indicate that the energy differences between the different β-D-Glc*p*NAc pucker states is reduced when both Glc*p*$^{\text{I}}$ and Glc*p*$^{\text{II}}$ have the same anomeric configuration. When both linkages are α-linked the relative energy difference between $^{4}C_1$ and boat/skew pucker states are lower in energy than when both residues are β-linked. If the anomeric configurations for the two residues are different (one α and the other β) the relative energies between the $^{4}C_1$ and boat/skew conformers are significantly greater, resulting in reduced puckering of the β-D-Glc*p*NAc ring.

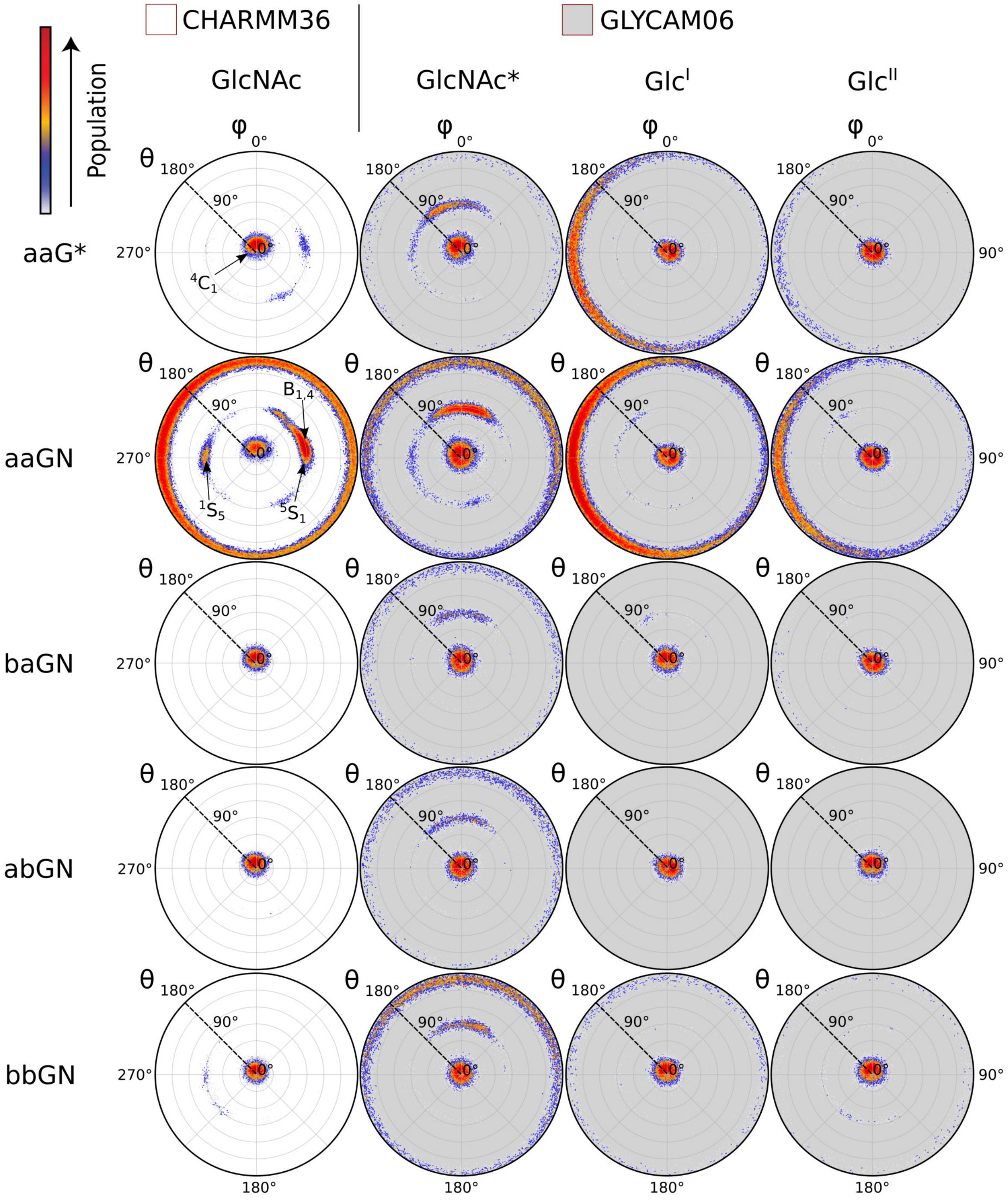


Figure 7.7: Polar projections of the Cremer-Pople puckering sphere for aaG, aaGN, baGN, abGN and bbGN as simulated with the CHARMM36 and GLYCAM06 force fields. Polar projections for $\mathrm{Glc}^{\mathrm{I}}$ and $\mathrm{Glc}^{\mathrm{II}}$ from the CHARMM36 simulations are now shown as both residues for all trisaccharides remain in $^4C_1$ for the entire simulation. The north-pole ($^4C_1$) lies at the centre of the ring and the south pole ($^1C_4$) is at the perimeter. Simulation time for is 4 μs for aaGN and 2 μs for all other trisaccharides. *Note aaG contains the $\mathrm{Glc}^{\mathrm{III}}$ residue and not Glc*p*NAc.

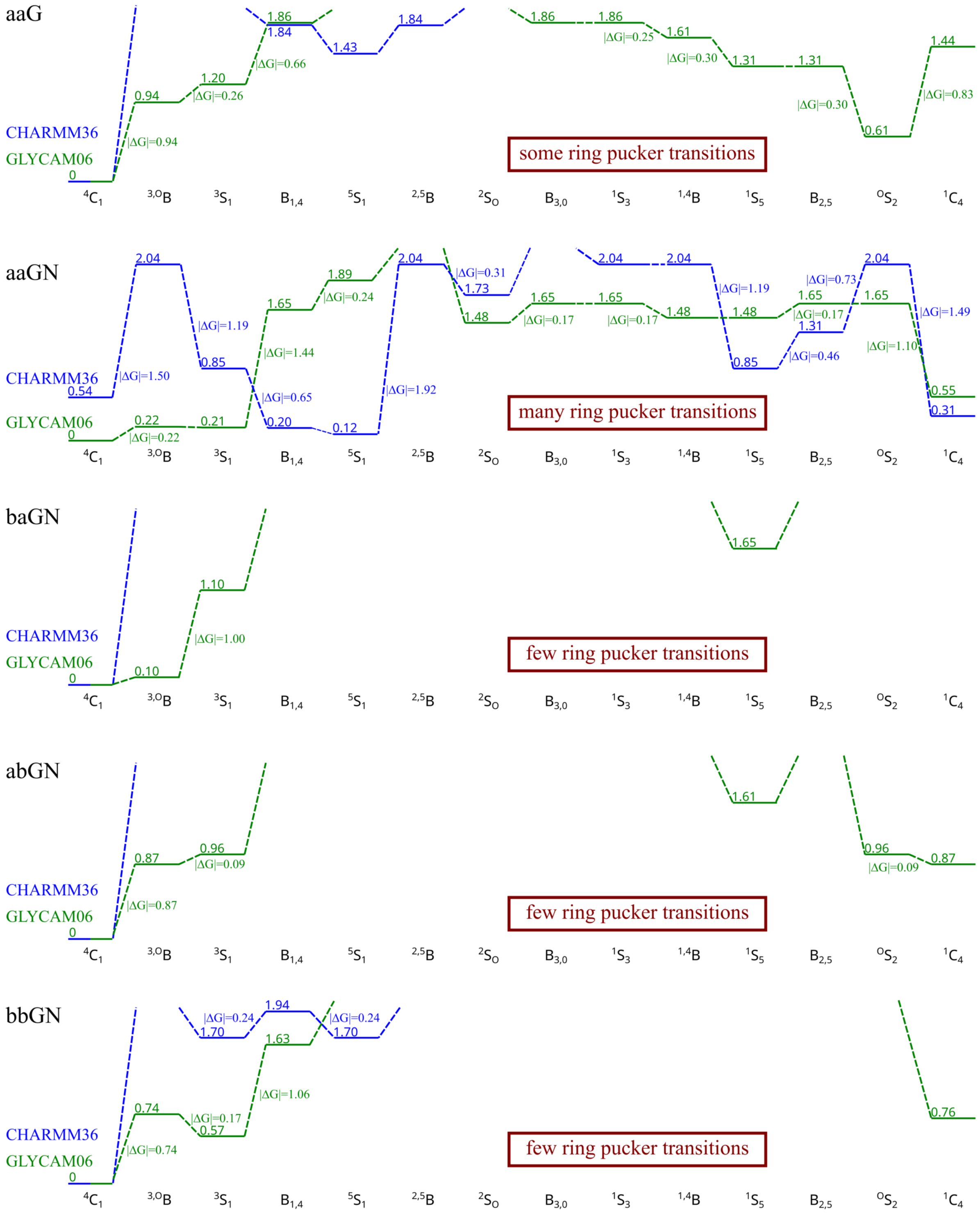


Figure 7.8: Free energies of the *β*-D-Glc$^{\mathrm{III}}$ (in aaG) and *β*-D-Glc*p*NAc (in aaGN, baGN, abGN, and bbGN) ring pucker as calculated from simulated equilibrium data of the trisaccharides for the CHARMM36 and GLYCAM06 force fields computed using the standard $\Delta G = -RTln|K_{eq}|$ relationship. Where no value is specified the free-energy could not be calculated. The minimum energy ring pucker conformer for *β*-D-Glc$^{\mathrm{III}}$/*β*-D-Glc*p*NAc for all but one trisaccharide is $^4C_1$ - in aaGN it is $^5S_1$.

Effect of 1-O-methylation on ring pucker of aaGN

The β-D-Glc*p*NAc residue in *Sf* 7a and 7b is fully substituted, with anomeric linkages at the 1-, 3-, and 4- positions. The trisaccharides modelled thus far do not include substitution at the anomeric position. We methylate the anomeric position of β-D-Glc*p*NAc and simulated α-D-Glc*p*$^{\text{II}}$ -(1→3)-[α-D-Glc*p*$^{\text{I}}$ -(1→4)]-β-D-Glc*p*NAcOMe as an initial model of polymerisation and to better understand the ring pucker conformations of a fully substituted a β-D-Glc*p*NAc that is 3,4-α-disubstituted.

The pucker conformers adopted by β-D-Glc*p*NAcOMe are similar to those as β-D-Glc*p*NAc in aaGN for both force fields. However, the amount of time spent in each of these differs significantly (Figure 7.9: a-d). In the CHARMM36 simulation the time series plot immediately shows that the β-D-Glc*p*NAc residue quickly inverts from $^4C_1$ to $^1C_4$ and remains in this state for almost the entire duration of the simulation. Only a single transition is made, for a brief period of about 100 ns, into a boat/skew state. At no point does it return to $^4C_1$. Similarly, in the GLYCAM06 simulation β-D-Glc*p*NAcOMe spends significantly more time ($\approx$ 75%) in $^1C_4$ than in $^4C_1$ or the boat/skew states.

These results suggest that O-methylation of the anomeric position leads to further deviation away from the $^4C_1$ conformer (Figure 7.9: e-h). The addition of a bulky -OMe group to the equatorial position of β-D-Glc*p*NAc to form (β-D-Glc*p*NAcOMe) may further increase the steric interaction of the already highly strained region around the ring. It appears that to alleviate some of this strain puckering occurs, placing the substituent groups in an axial position and further apart for extended periods of time even at the expense of increased 1,3-diaxial interactions in the substituents.

Finally, it should also be noted that there is no significant change in the torsion angles adopted by the 1→3 and the 1→4 linkages when compared to the β-D-Glc*p*NAc simulations. Additionally, the α-D-Glc*p*$^{\text{I}}$ and α-D-Glc*p*$^{\text{II}}$ residues maintain a $^4C_1$ conformation throughout both simulations.

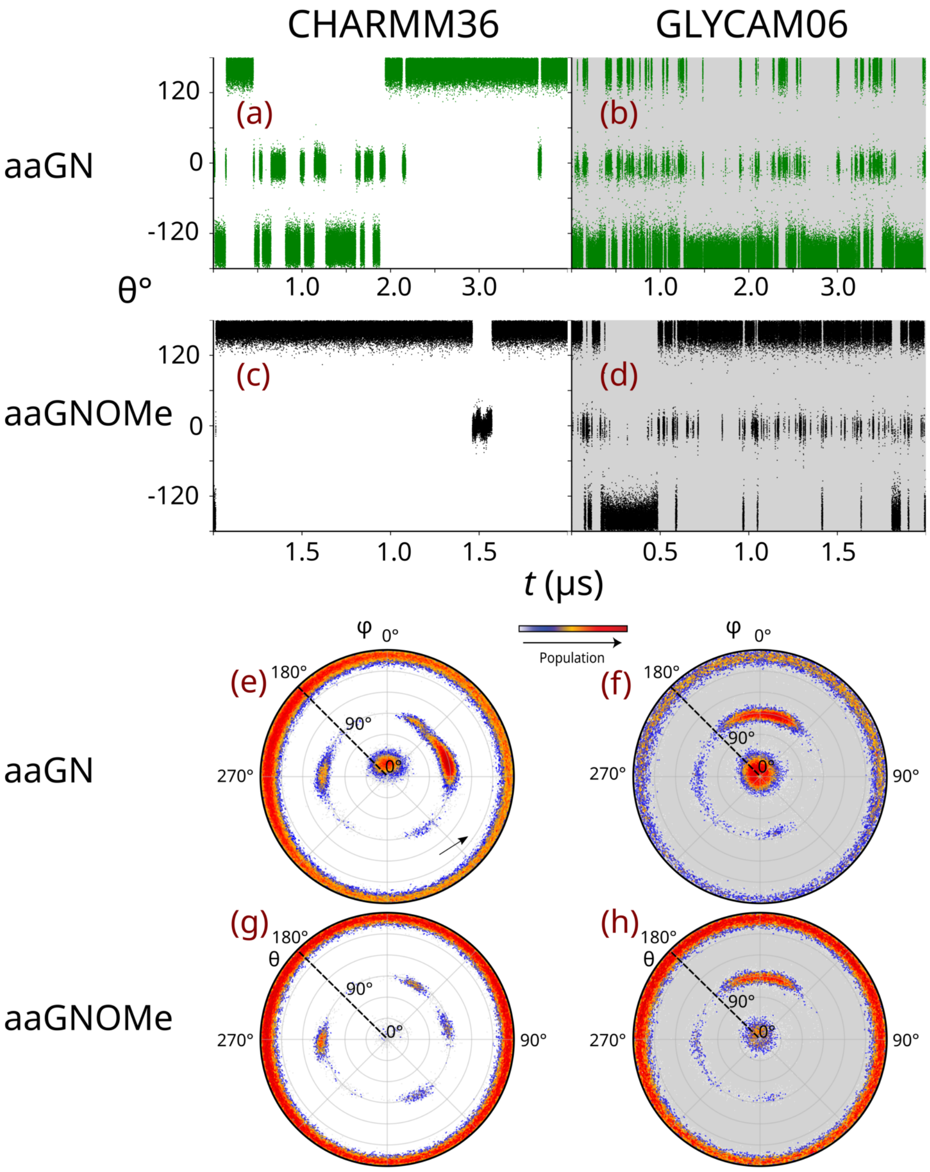


Figure 7.9: Time series of the Cremer-Pople azimuth pucker, θ, parameter for β-D-Glc*p*NAc in aaGN (a, b) and β-D-Glc*p*NAcOMe in aaGNOMe (c, d). Polar projections of the Cremer-Pople puckering sphere β-D-Glc*p*NAc in aaGN (e, f) and β-D-Glc*p*NAc in aaGNOMe (g, h). Simulations with CHARMM36 have a white background and with GLYCAM06 a grey background.

Ring pucker of Glcp^{I} and Glcp^{II}

Thus far we have only discussed the ring puckering behaviour of β-D-Glcp^{III} and β-D-Glc$p$NAc. However, analysis of the Glcp^{I} and Glcp^{II} residue puckers can also provide further insight into trisaccharide conformation.

In the CHARMM36 simulations of all five trimers (aaG, aaGN, baGN, abGN, and bbGN) the Glcp^{I} and Glcp^{II} residues remain entirely in the $^4C_1$ conformer ($\theta \approx 0°$). At no point during the simulations do either of the two Glc$p$ residues deviate away from this state (the CHARMM36 plots in Figure 7.10). On the other hand, in the GLYCAM06 simulations, Glc$^{\mathrm{I}}$ and Glc$^{\mathrm{II}}$ display different ring pucker behaviour for all the trisaccharides (the GLYCAM06 plots in Figure 7.10 and Figure 7.11).

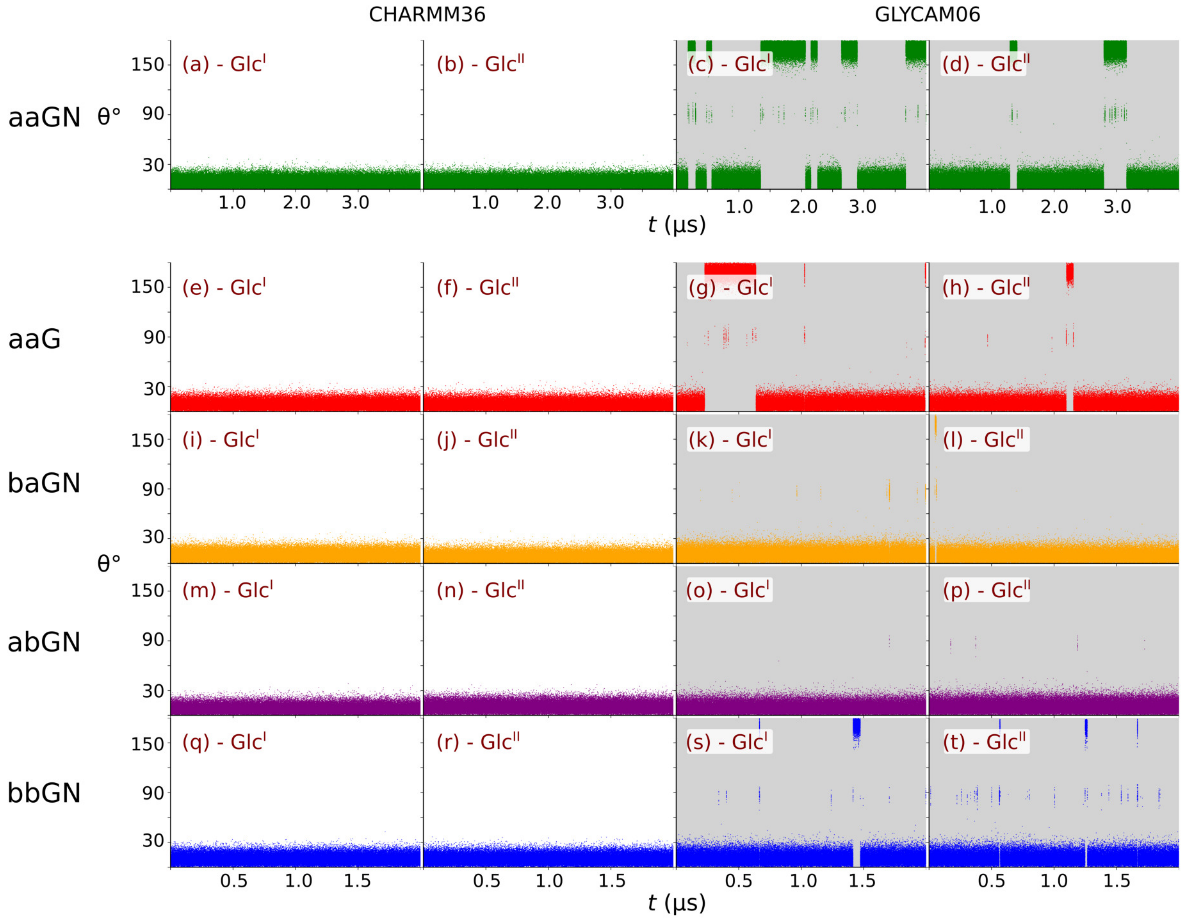


Figure 7.10: Time series of the Cremer-Pople pucker azimuth (θ) for α-D-Glc$^{\mathrm{I}}$ and α-D-Glc$^{\mathrm{II}}$ in aaGN (4 μs) as well as aaG, baGN, abGN, and bbGN (2 μs) using CHARMM36 and GLYCAM06 force fields. Ring puckering of these two saccharides are relatively rigid in the CHARMM36 simulations while it is more flexible in the GLYCAM06 simulations.

In the GLYCAM06 simulation of aaG both α-D-Glc residues transition to boat/skew conformers (where the $^{O}S_2$ conformer is adopted most frequently) as well as to $^1C_4$ (Figure 7.10: g, h). α-D-Glc$^{\mathrm{I}}$ makes several boat/skew $\leftrightarrow$ $^1C_4$ transitions and in total spends $\approx$ 400 ns in $^1C_4$. However, almost all of this time is within the first 1 μs of the simulation. In the second half of the simulation α-D-Glc$^{\mathrm{I}}$ maintains a $^4C_1$ almost entirely, with only two transitions to boat/skew and $^1C_4$. α-D-Glc$^{\mathrm{II}}$ also remains in $^4C_1$ for close to the entire simulation with only a few brief transitions to non-$^4C_1$ conformers.

In contrast, both α-D-Glc$^{\mathrm{I}}$ and α-D-Glc$^{\mathrm{II}}$ pucker significantly in the GLYCAM06 simulation of aaGN. Glc$^{\mathrm{I}}$ undergoes $^4C_1$ $\leftrightarrow$ boat/skew $\leftrightarrow$ $^1C_4$ transition relatively frequently (Figure 7.10: c, d). It is either in $^4C_1$ or $^1C_4$ for most of the simulation, spending approximately the same amount of time in each state. As is common with the other GLYCAM06 simulations observed thus far α-D-Glc$^{\mathrm{I}}$ adopts boat/skew states for only short periods of time as it transitions between the two chair conformers. α-D-Glc$^{\mathrm{II}}$ remains in $^4C_1$ for $\approx$ 90 % of the simulation with two transitions boat/skew and $^1C_4$. The average lifetime of the $^1C_4$ chair is 273 ns and 103 ns respectively for α-D-Glc$^{\mathrm{I}}$ and α-D-Glc$^{\mathrm{II}}$ respectively. The average lifetime of the boat/skew residues is approximately 2 ns for both α-D-Glc$^{\mathrm{I}}$ and α-D-Glc$^{\mathrm{II}}$.

Interestingly the notable puckering of Glc$^{\mathrm{I}}$ and Glc$^{\mathrm{II}}$ observed in aaGN is not seen in baGN and abGN (Figure 7.10: k, l, o, p). The Glc*p* residues in these trisaccharides make a very small number of short-lived transitions into either a boat/skew or $^1C_4$ state. They remain stable in $^4C_1$ for almost the entire simulation. However, Glc*p*$^{\mathrm{I}}$ and Glc*p*$^{\mathrm{II}}$ in bbGN show greater dynamic behaviour, puckering to slightly greater extents into both boat/skew or $^1C_4$ conformers (Figure 7.10: s, t).

It is increasingly evident that when β-D-Glc*p*NAc is either both α- or β-linked at the 3- or 4-positions there is increased puckering in all residues. This may be attributed to potentially greater steric strain between adjacent substituents on the β-D-Glc*p*NAc ring. Furthermore, these steric interactions may be larger when both the 3- and 4- positions are α-linked when compared to both being β-linked, which is why we see a significant degree of puckering in aaGN and less so in bbGN. The baGN and abGN residues show the least amount of puckering (or none at all depending on the force field) as the different orientations at the 3- and 4-linkages potentially help reduce the steric strain around the β-D-Glc*p*NAc ring.

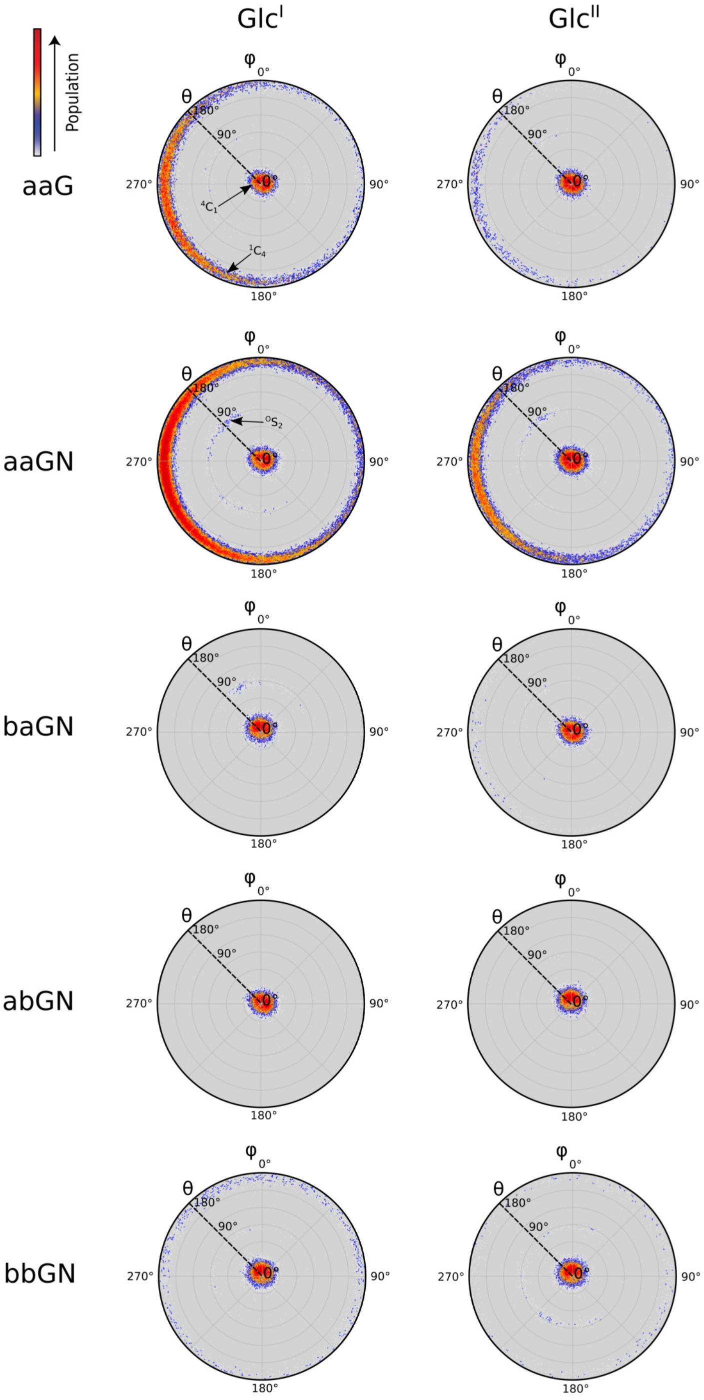


Figure 7.11: Polar projections of the Cremer-Pople puckering sphere of Glc^{I} and Glc^{II} for aaG, aaGN, baGN, abGN and bbGN as simulated with the GLYCAM06 force field. The north-pole ($^{4}C_{1}$) lies at the centre of the ring and the south pole ($^{1}C_{4}$) is at the perimeter. Projections from the CHARMM36 simulations are not shown as both Glc^{I} and Glc^{II} for all trisaccharides remain in $^{4}C_{1}$ for the entire simulation time. Simulation time is 4 μs for aaGN and 2 μs for all other structures.

7.2.4 The free energy of ring puckering in β-D-Glc*p*$^{\text{III}}$/β-D-Glc*p*NAc

We expand on the MD simulations by conducting a PMF analysis on the trisaccharides. We measure the energy differences between various pucker conformations of β-D-Glc$^{\text{III}}$ (in aaG) and β-D-Glc*p*NAc (in aaGN, baGN, abGN, and bbGN). We also conduct a comparative analysis of the energy differences for the pucker interconversion pathways for both force fields.

Energy landscape of the β-D-Glc*p*$^{\text{III}}$/β-D-Glc*p*NAc ring

We calculate PMF energies for the different pucker conformations of β-D-Glc*p*$^{\text{III}}$ and β-D-Glc*p*NAc in the trisaccharides (Figure 7.12). We measure the energy differences between different pucker states and compare them to the unusual pucker conformations observed in the MD simulations. The simulations of both force fields show significant differences in the free-energy profiles of the β-D-Glc*p*$^{\text{III}}$ and β-D-Glc*p*NAc pucker states. The anomeric configuration at the 3- and 4-positions of β-D-Glc*p*NAc also affects the ring pucker transition energies, with α disubstitution having the lowest energy barriers.

For the CHARMM36 PMF simulation the global minimum energy calculated for the β-D-Glc$^{\text{III}}$ ring pucker in aaG (Figure 7.12-a) resides at $\phi, \theta = 16°, 13°$; corresponding to $^4C_1$. The energy-well below 1 kcal mol^{-1} relative to the global minima covers the region from $0° - 70°$ and $290° - 360°$ in ϕ, and $10° - 25°$ in θ. Note that this minima is not located at $\theta = 0°$, which would correspond to an ideal $^4C_1$ state. Instead it is close to ^OE, $^OH_1$, and $^2H_1$, which is in partial agreement with observations from the CHARMM36 aaG MD simulations. At $\theta = 0°$ the ring pucker conformer is approximately $\Delta G = +5.5$ kcal mol^{-1} greater in energy.

There is a single local minimum that is less than 1 kcal mol^{-1} greater in energy at $\phi, \theta = 82°, 86°$ ($\Delta G = +0.80$ kcal mol^{-1}), consistent with a $^5S_1$ pucker. There are three other notable local minima corresponding to boat/skew conformers at $\phi, \theta = 148°, 90°$ (^2S_O, $\Delta G = +1.3$ kcal mol^{-1}); $\phi, \theta = 213°, 87°$ ($^1S_3$, $\Delta G = +2.6$ kcal mol^{-1}); $\phi, \theta = 269°, 88°$ ($^1S_5$, $\Delta G = +2.7$ kcal mol^{-1}). The $^4C_1$, $^5S_1$ conformers occur frequently in the CHARMM36 aaG MD simulations but the other states are not observed for significant periods of time. The energy-well in the $^1C_4$ region ($\theta \approx 360°$) is above $\Delta G = +4$ kcal mol^{-1} and $^1C_4$ is not populated in the MD simulation.

In the GLYCAM06 PMF simulation of aaG (Figure 7.12-b) the global minimum for the β-D-Glc$^{\text{III}}$ ring pucker is at $\phi, \theta = 323°, 15°$; also corresponding to a $^4C_1$ conformer. There are three local minima in the boat/skew region at $\phi, \theta = 0°, 87°$ ($^{3,0}B$, $\Delta G = +1.5$ kcal mol^{-1}); $\phi, \theta = 278°, 87°$ ($^1S_5$, $\Delta G = +1.9$ kcal mol^{-1}); and $\phi, \theta = 310°, 83°$ ($^OS_2$, $\Delta G = +1.8$ kcal mol^{-1}). There is a single energy-well in the $^1C_4$ area at $\phi, \theta = 80°, 170°$ ($\Delta G = +2.7$ kcal mol^{-1}). These are also the most frequented β-D-Glc*p*NAc pucker states in the aaG GLYCAM06 MD

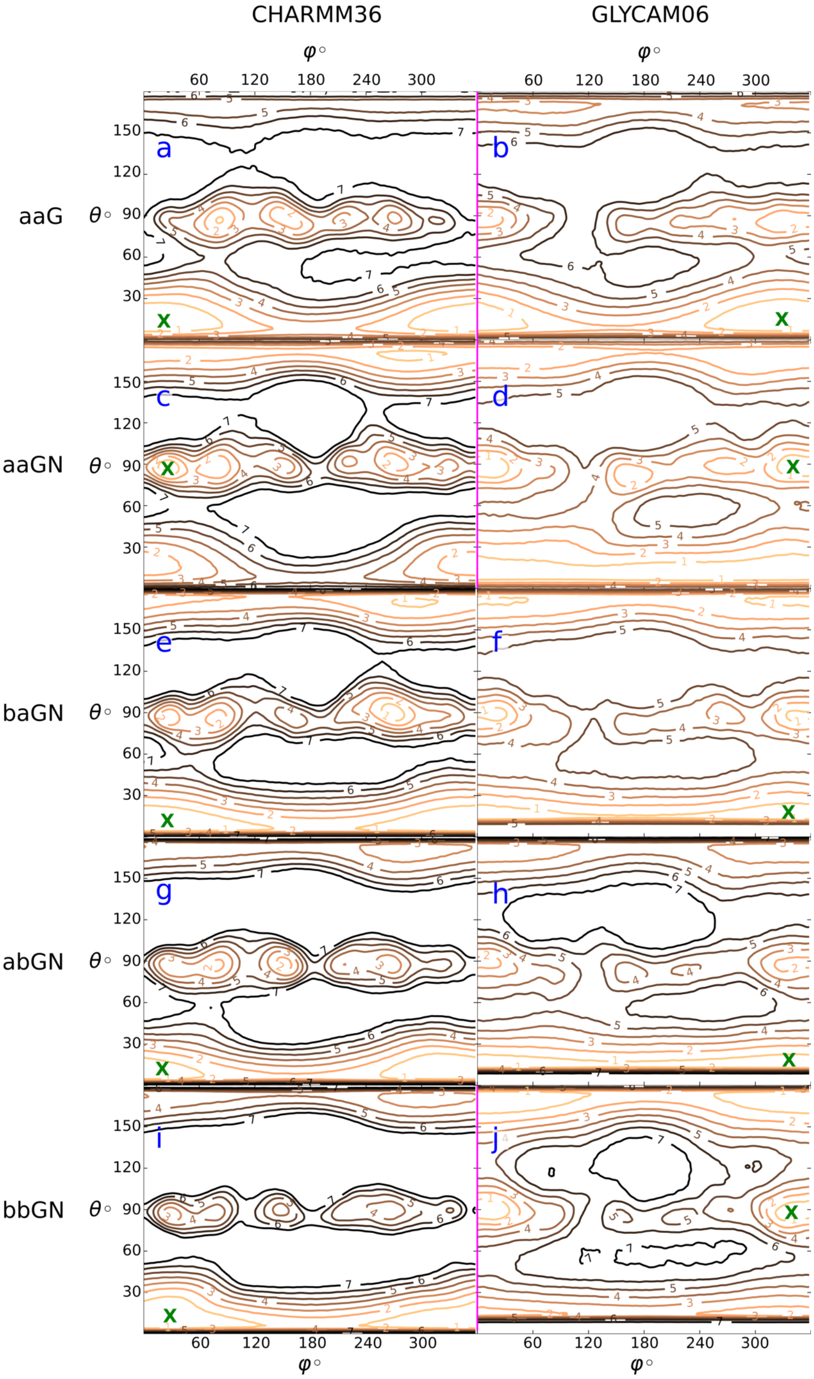


Figure 7.12: Mercator plot of the Cremer-Pople pucker meridian (ϕ) and azimuth (θ) parameters from potential of mean force calculations for the ring pucker of β-D-Glc$^{\mathrm{III}}$ in aaG and β-D-Glc*p*NAc in aaGN and baGN. For all trisaccharides, with the exception of aaGN, and across both force fields the lowest energy β-D-Glc*p*NAc ring conformer corresponds to the $^4C_1$ state. For aaGN modelled with CHARMM36 the lowest energy β-D-Glc*p*NAc conformer is $^3S_1$ while with GLYCAM06 it is $^OS_2$. Contours labels provide energy differences in kcal mol^{-1}. The location of the global energy minimum designated by X on each plot.

simulation — $^4C_1$ is the overwhelmingly dominant conformer with others present for shorter periods.

In contrast, the β-D-Glc*p*NAc ring pucker energies from the aaGN CHARMM36 PMF simulations are significantly different (Figure 7.12-c). The global energy minimum resides at $\phi, \theta = 28°, 87°$, which correlates to $^3S_1$. There is a single minima for $^4C_1$ at $\phi, \theta = 2°, 16°$ ($\Delta G = +1.5$ kcal mol^{-1}). In comparison, the $^1C_4$ conformer has a single, smaller energy-well and its minimum is lower in energy (+0.8 kcal mol^{-1}). There are multiple local minima in the boat/skew region and these are will within +2 kcal mol^{-1} of the global minimum. These results are similar to the those of the aaGN CHARMM36 MD simulations - the $^4C_1$ conformer is higher in energy than both $^1C_4$ and several boat/skew states. However, the PMF calculations indicate that $^3S_1$ is the lowest energy pucker overall whereas in the MD simulations $^5S_1$ is the most frequented followed by $B_{1,4}$ and $^3S_1$.

The aaGN GLYCAM06 PMF simulation energy landscape (Figure 7.12-d) shows a distinctly different profile to that of the CHARMM36 simulation. There are broad energy-wells in the $^4C_1$ ($\Delta G = +0.2$ kcal mol^{-1}) and $^1C_4$ (+1.0 kcal mol^{-1}) regions. The global minimum corresponds to $^OS_2$ and is located at $\phi, \theta = 343°, 88°$. There are two other local minima in the boat/skew region, which are fewer than those observed in the CHARMM36 simulation. These are found at $\phi, \theta = 159°, 81°$ (^2S_O, 1.1 kcal mol^{-1}) and $\phi, \theta = 272°, 87°$ ($^1S_5$, 1.5 kcal mol^{-1}). The aaGN GLYCAM06 MD simulation also suggests small energy differences with frequent transitions between the chair and boat/skew states, although, the $^4C_1$ state has the lowest free energy.

The energy landscape of the β-D-Glc*p*NAc from the PMF simulations of baGN, abGN, and bbGN show similar profiles to those of aaG for the respective force fields. The energy minima for the boat/skew conformers are also located in the same regions. For the CHARMM36 simulations (Figure 7.12: e, g, i) these correspond to $^3S_1$, $B_{1,4}$, ^2S_O, $^1S_3$, and $^1S_5$. While in the GLYCAM06 simulations (Figure 7.12: f, h, and j) the minima correlate to the $^OS_2$, ^2S_O, and $^1S_5$ conformers. The global minima for both force fields is located in the $^4C_1$ region. The local minima for the boat/skew conformers are approximately +1.0 kcal mol^{-1} to +3.0 kcal mol^{-1} higher in energy while the $^1C_4$ pucker is within 1.0 kcal mol^{-1} of $^4C_1$. These results also correspond with the MD simulations of the trisaccharides for both force fields.

The main difference between the energy profiles of aaGN and the other N-acetylated trisaccharides is the $^4C_1$ ↔boat/skew and $^1C_4$ ↔boat/skew transition paths. In the CHARMM36 simulations, these paths in aaGN are as low as approximately 4 kcal mol^{-1} while in the baGN, abGN, and bbGN it is 6 kcal mol^{-1}. The energy differences between the chair and boat/skew conformers are lower in the simulations at 4 kcal mol^{-1} for all structures.

These results correlate well with those observed in the MD simulations. In the CHARMM36 simulations β-D-Glc*p*NAc tends to spend longer periods in boat/skew while in the GLYCAM06 simulations the chair conformers are preferred, with rapid transitions into and out of boat/skew. The energies of the ring pucker transition pathways provide further insight into the β-D-Glc$^{\text{III}}$ and β-D-Glc*p*NAc ring pucker behaviour.

Transition energies of the β-D-Glc*p*$^{\text{III}}$/β-D-Glc*p*NAc ring

We map the transition pathways between the chair and boat/skew conformers of β-D-Glc$^{\text{III}}$ (in aaG) and β-D-Glc*p*NAc (in aaGN, baGN, abGN, and bbGN) to determine the energy differences between different pucker states and to identify potential ring pucker pathways for the MD simulations. The energy differences between adjacent pucker states along each path are provided in modified Stoddart diagrams. Figure A.11 illustrates the energy differences along the paths between chair and boat/skew conformers but does not include the half-chair/envelope pucker states that lie between them. Figure 7.13 depicts the half-chair/envelope along the paths only where the energy differences between adjacent pucker states are less than or equal to $+5.0$ kcal mol^{-1} from the respective chair conformer. The ring pucker transition energy pathways shows clear contrasts between the two force fields.

The Stoddart diagram for the aaG CHARMM36 PMF simulation shows that energy difference between $^4C_1$ and the boat/skew conformers are almost all less than 5 kcal mol^{-1}. The only exception is the $^4C_1 \leftrightarrow ^{3,O} B$ transition, which is 6.5 kcal mol^{-1}. All boat/skew conformers are within 3.2 kcal mol^{-1} of $^1C_4$ and within 2.8 kcal mol^{-1} of adjacent boat/skew conformers. The lowest energy boat/skew transition is 0.6 kcal mol^{-1} ($B_{3,O} \leftrightarrow ^1 S_3$ and $B_{2,5} \leftrightarrow ^O S_2$) and the highest energy transition is 2.8 kcal mol^{-1} ($^{3,O}B \leftrightarrow {}^3S_1$). There are three $^4C_1 \leftrightarrow$ half-chair/envelope $\leftrightarrow$ boat/skew transition paths that are less than 5 kcal mol^{-1}: $^4C_1 \leftrightarrow {}^OH_1 \leftrightarrow {}^3S_1$; $^4C_1 \leftrightarrow E_1 \leftrightarrow B_{1,4}$; and $^4C_1 \leftrightarrow {}^2H_1 \leftrightarrow {}^5S_1$ (Figure 7.13: a). The lowest energy transition is $^4C_1 \leftrightarrow ^O H_1$ (+4.4 kcal mol^{-1}) and $^OH_1 \leftrightarrow ^3 S_1$ (-0.8 kcal mol^{-1}). The transitions between $^4C_1\leftrightarrow$boat/skew states observed in the aaG MD simulations correspond to these low energy pathways.

The GLYCAM06 aaG Stoddart diagrams show that the energy differences between adjacent chair/boat-skew pucker states along the given paths are, with one exception, all within 5 kcal mol^{-1}. The only transition along a given path that are greater than this is $^4C_1 \leftrightarrow {}^{2,5}B$ at 5.9 kcal mol^{-1}. Moreover, the energy differences between $^4C_1$, $^1C_4$, and boat/skew pucker states that are neighbours on the Stoddart transition paths are lower than almost all of those in the CHARMM36 PMF simulation.

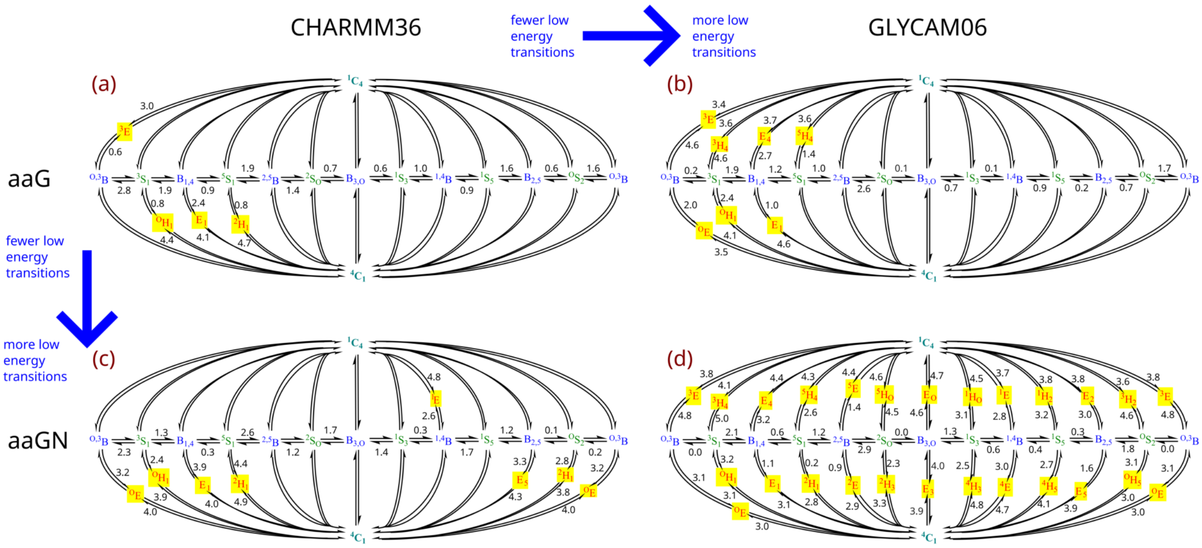


Figure 7.13: Transition state interconversion pathways as calculated from potential of mean force calculations for the ring pucker of β-D-Glc$^{\mathrm{III}}$ in aaG and β-D-Glc*p*NAc in aaGN. Only half-chair and envelope transitions that are less than or equal to 5 kcal mol^{-1} are shown in yellow. For aaG only several selected pathways are below 5 kcal mol^{-1} for both CHARMM36 and GLYCAM06 and these correspond closely with the pathways observed in the corresponding MD simulations of these trisaccharides. For aaGN there are a few more pathways below 5 kcal mol^{-1} for CHARMM36 but there are many more for GLYCAM06, indicating that N-acetylation reduces the energy required to transition out of $^{4}C_{1}$ and that this change is greater in GLYCAM06.

There are also a greater number of $^{4}C_{1}$ ↔ half-chair/envelope ↔ boat/skew transition paths that are below 5 kcal mol^{-1} in the GLYCAM06 simulation (Figure 7.13-b). The transition energies to and from the half-chair/envelope states are also mostly lower than the CHARMM36 PMF simulation. For example, the frequent $^{4}C_{1} \leftrightarrow\ ^{O}S_{2}$ transition via $^{O}H_{5}$ has an energy barrier of over 5 kcal mol^{-1} in the CHARMM36 simulation while the same transition energy barrier is 3.2 kcal mol^{-1} in the GLYCAM06 simulation. Interestingly, the common $^{4}C_{1} \leftrightarrow\ ^{3}S_{1} \leftrightarrow B_{1,4}$ transition has the same energy barrier for both force fields at 4.1 kcal mol^{-1} and 1.9 kcal mol^{-1} respectively. Although, these are the sole exceptions. The overall lower energy barriers between to the half-chair/envelope states provides an explanation for the greater frequency of $^{4}C_{1}$ ↔ boat/skew and boat/skew ↔ $^{1}C_{4}$ transitions seen in the GLYCAM06 MD simulations.

In contrast to β-D-Glc$^{\mathrm{III}}$ in aaG, the energy differences between the pucker states of β-D-Glc*p*NAc in aaGN are generally smaller. For both force fields there are no adjacent pucker states along the defined paths that are more than 5 kcal mol^{-1}. In the CHARMM36 PMF simulation there are seven $^{4}C_{1}$ ↔ half-chair/envelope ↔boat/skew paths less than 5 kcal mol^{-1} (Figure 7.13-c). The lowest energy path is $^{4}C_{1} \leftrightarrow\ ^{O}H_{1}$ (+3.9 kcal mol^{-1}) and $^{O}H_{1} \leftrightarrow\ ^{3}S_{1}$ (+2.4 kcal mol^{-1}). In contrast, there is only a single such path from $^{1}C_{4}$: $^{1}C_{4} \leftrightarrow\ ^{1}E$ (+4.8 kcal mol^{-1}) and $^{1}E \leftrightarrow\ ^{1,4}B$ (+2.6 kcal mol^{-1}). These low energy transition pathways provide a rationale for the conformations observed in the aaGN CHARMM36 MD simulations.

The Stoddart diagrams for the GLYCAM06 PMF simulation of aaGN illustrate that all $^{4}C_{1}$ ↔half-chair/envelope and boat/skew↔half-chair/envelope conformers are below 5 kcal mol^{-1} and the energies along these paths are also less than those in the CHARMM36 simulation. The lowest energy path from $^{4}C_{1}$ is along the $^{4}C_{1} \leftrightarrow {}^{2}H_{1}$ (+2.8 kcal mol^{-1}) and $^{1}H_{1} \leftrightarrow {}^{5}S_{1}$ (+0.3 kcal mol^{-1}) route (Figure 7.13: d), which is approximately 1 kcal mol^{-1} less in energy than the lowest energy route out of $^{4}C_{1}$ from the CHARMM36 simulation. All paths into and out of $^{1}C_{4}$ are less than 5 kcal mol^{-1}. Similar to the CHARMM36 simulation the lowest energy path is via the ^{1}E intermediate along $^{1}C_{4} \leftrightarrow^{1} E$ (+3.7 kcal mol^{-1}) and $^{1}E \leftrightarrow^{1,4} B$ (+2.8 kcal mol^{-1}).

The transition pathways of baGN, abGN, and bbGN reveal larger energy difference between the chair and boat/skew when compared to aaGN for both force fields. However, the energies between different pucker states are consistently higher in the CHARMM36 PMF simulations compared to the GLYCAM06 PMF simulations, which accounts for the increased dynamic behaviour of the ring pucker in the GLYCAM06 MD simulations.

These results provide further insight into the ring pucker dynamics of β-D-Glc*p*NAc observed in the MD simulations of aaGN. In the CHARMM36 MD simulations there are fewer transitions into and out of $^{4}C_{1}$. The boat conformers are also preferred and are adopted for greater periods of time, while the $^{1}C_{4}$ state is overwhelmingly dominant. This can be attributed to a number of factors. The $^{3}S_{1}$, $^{5}S_{1}$, $B_{3,O}$, and $^{1}C_{4}$ conformers are all lower in energy than $^{4}C_{1}$. Moreover, there is a relatively low $^{4}C_{1}$ ↔half-chair/envelope↔boat/skew barrier. These two components contribute to the low $^{4}C_{1}$ states observed in the MD simulation. Furthermore, the higher energy barriers between $^{4}C_{1}$ and half-chair/envelope pucker states, and the lower energy barriers into and out of boat/skew states, means that boat/skew conformers will be favoured.

Energy landscape of the Glc*p*$^{\text{I}}$ and Glc*p*$^{\text{II}}$ rings

The trisaccharide MD simulations indicate that the behaviour of the Glc*p*$^{\text{I}}$ and Glc*p*$^{\text{II}}$ ring pucker may be force field dependant. No notable puckering of these residues are observed in the CHARMM36 simulations while a small but significant amount is present in the GLYCAM06 simulations.

The PMF simulations also reveal no pucker of either of the Glc*p*$^{\text{I}}$ and Glc*p*$^{\text{II}}$ rings. This is not particularly surprising as the simulation time is relatively short and the timescales observed for the ring puckering is in the order of a several hundred nanoseconds; as observed from the MD simulations.

Similar behaviour is observed in the PMF simulations of all other trisaccharides. In short, these simulations are not sufficiently long to probe the relationship between the ring pucker of β-D-Glc*p*NAc and that of Glc*p*$^{\mathrm{I}}$ or Glc*p*$^{\mathrm{II}}$. The results indicate that the energies measured are a function β-D-Glc$^{\mathrm{I}}$/β-D-Glc*p*NAc ring pucker and are not affected by puckering of the Glc*p*$^{\mathrm{I}}$ or Glc*p*$^{\mathrm{II}}$ residues.

Comparisons with the energy landscape of 3,4-disubstituted β-D-Gal*p*NAc and β-D-Man*p*NAc rings

We simulated 3,4-disubstituted β-D-Gal*p*NAc and β-D-Man*p*NAc trisaccharides to probe whether the puckering behaviour observed in 3,4-disubstituted β-D-Glc*p*NAc is present in other monosaccharides. The residues only differ with β-D-Glc*p*NAc in the orientation of a single group — the 4- and 2- positions are axial in β-D-Gal*p*NAc and β-D-Man*p*NAc respectively. In β-D-Glc*p*NAc both are equatorial. Figure 7.14 provides the potential of mean force energies, illustrated as contour plots of Cremer-Pople ring pucker parameters, for different pucker conformations.

The global minimum for the β-D-Gal*p*NAc ring pucker in the CHARMM36 simulation (Figure 7.14-a) corresponds to $^4C_1$ and is located at $\phi,\theta = 70°, 11°$. Boat/skew conformers are all greater than 4 kcal mol^{-1} in energy with all $^4C_1 \leftrightarrow$ boat/skew transition paths over 5 kcal mol^{-1}. The $^4C_1 \leftrightarrow B_{1,4}$ path is lowest in energy between $5-6$ kcal mol^{-1}. All other paths are over 6 kcal mol^{-1} in energy. The $^1C_4$ state has a minimum at $\phi,\theta = 305°, 168°$ (+1.3 Kcal mol^{-1}).

The transition paths in the GLYCAM06 simulation of β-D-Gal*p*NAc (Figure 7.14-b) all appear to be of greater energy than those in CHARMM36. In the 50 ns metadynamic simulation the ring only entered two boat/skew conformational states - $\phi,\theta = 189°, 80°$ ($B_{3,O}$, +4.0 kcal mol^{-1}) and $\phi,\theta = 205°, 81°$ ($^1S_3$, +3.2 kcal mol^{-1}). The global minimum also corresponds to a $^4C_1$ pucker state at $\phi,\theta = 217°, 14°$. No other boat/skew or chair conformers are traversed during the simulation and the minimum transition to all any of them from $^4C_1$ is 7 kcal mol^{-1}.

The CHARMM36 PMF simulation of aaManN (Figure 7.14-c) shows that the energy profile of the β-D-ManNAc residue is also significantly different to that of β-D-Glc*p*NAc in aaGN. The energy barrier between $^4C_1$ and boat/skew conformations is over 7 kcal mol^{-1}. Only a single boat/skew conformation is traversed during the simulation at is located at $\phi,\theta = 213°, 84°$

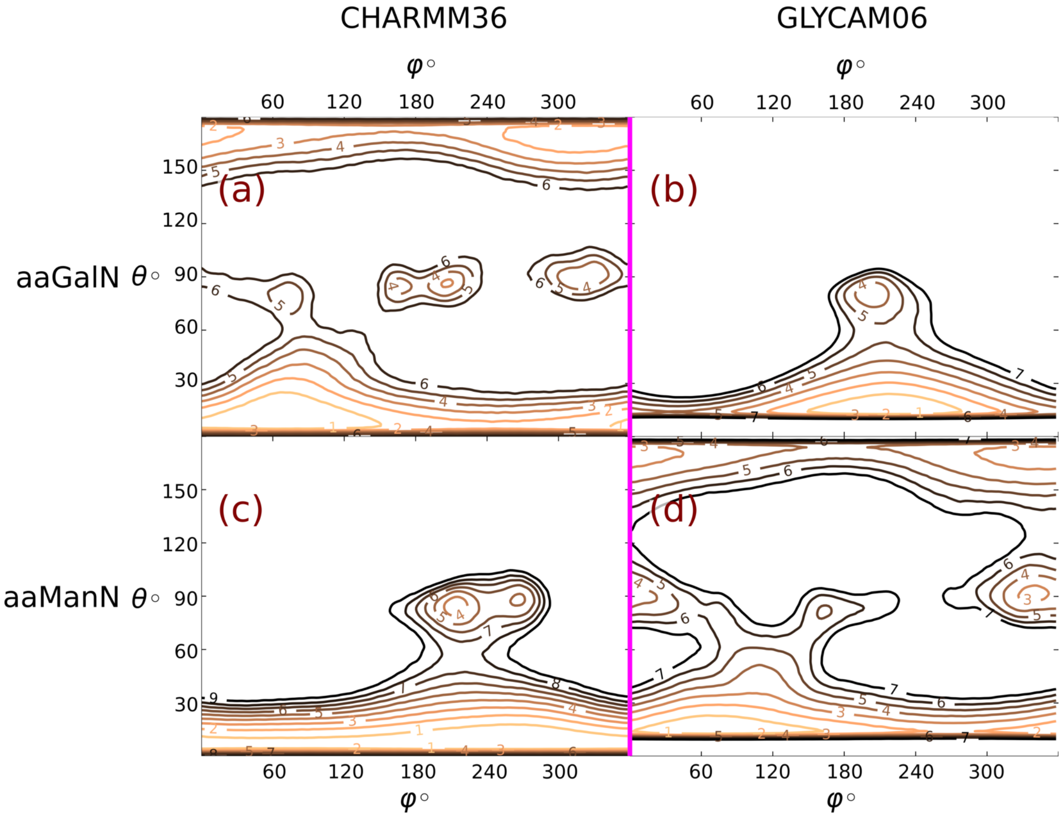


Figure 7.14: Mercator plot of the Cremer-Pople pucker meridian (ϕ) and azimuth (θ) parameters from potential of mean force calculations for the ring pucker of β-D-Gal*p*NAc in aaGalN and β-D-Man*p*NAc in aaManN. For both trisaccharides and force fields $^4C_1$ is the lowest energy ring pucker conformer. Contours labels provide energy differences in kcal mol^{-1}.

($^1S_3$, +3.3 kcal mol^{-1}). The relative energy of the $^1C_4$ conformer can also not be determined as β-D-Man*p*NAc does not enter the pucker state at any point during the 50 ns simulation.

The GLYCAM06 PMF simulation of aaManN (Figure 7.14-d) reveals that the $^4C_1 \leftrightarrow$ boat/skew transition paths of β-D-Man*p*NAc all have an energy barrier greater than 6 kcal mol^{-1}. The $^4C_1$ conformer is also the most stable state with a minimum at $\phi, \theta = 68°, 14°$. The minimum region corresponding to $^1C_4$ is located at $\phi, \theta = 2°, 168°$ (+2.6 kcal mol^{-1}).

The results from the aaGalN and aaManN PMF simulations indicate that the change in orientation of the hydroxyl at either the 4- or 2-position significantly changes the ring pucker energy profile. The single reorientation of the from equatorial to axial appears to greatly stabilise the $^4C_1$ pucker state.

7.2.5 MD simulations of the *S. flexneri* 7a O-Ag repeating unit with CHARMM36 and GLYCAM06

The trisaccharide MD and PMF simulations of 3,4-disubstituted β-D-Glc*p*NAc highlighted significant differences between the CHARMM36 and GLYCAM06 force fields. Particularly with respect to the dynamic behaviour and conformations adopted by the β-D-Glc*p*NAc ring. To further compare and contrast the differences between the two force fields we perform solution MD simulations on a single repeating unit of *Sf* 7a with each force field.

The results show that the β-D-Glc*p*NAc ring adopts a $^1C_4$ conformation for extended periods of time in the simulations of both force fields (Figure 7.15: c). In the CHARMM36 simulation the β-D-Glc*p*NAc ring spends approximately 40% of the simulation time in $^1C_4$. The β-D-Glc*p*NAc ring enters the $^3S_1$, $B_{1,4}$, and $^5S_1$ states (Figure 7.15: g) as it also does in the 6 RU simulation. However, in this 1 RU structure it also populates the $^1S_5$ conformer for significant periods of time.

The β-D-Glc*p*NAc residue from the 1 RU modelled with GLYCAM06 spends approximately 80% of the simulation time in the $^1C_4$ pucker state (Figure 7.15: c, h). It also makes constant transitions into boat/skew but these are short-lived. This is markedly different not only to the CHARMM36 *Sf* 7a 1RU but also to the GLYCAM06 simulations, where significantly more time is spent in boat/skew states.

The results from these 1RU simulations are similar to the trisaccharide simulations. They show that the GLYCAM06 modelled structure has a more dynamic β-D-Glc*p*NAc ring. Making more frequent transitions between different pucker states. This indicates that the energy barrier between the different pucker conformers is smaller in GLYCAM06 than the CHARMM36 force field.

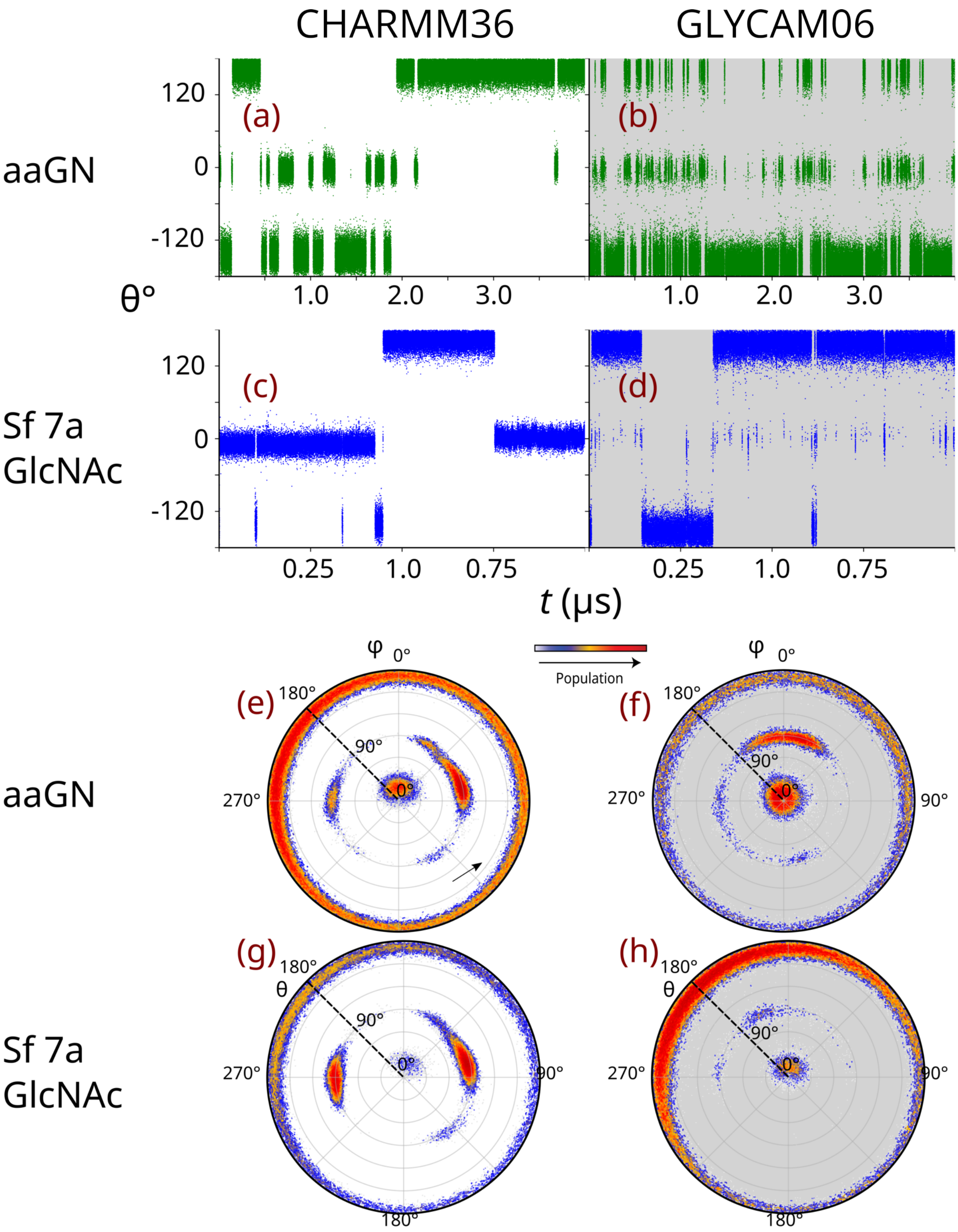


Figure 7.15: Time series of the Cremer-Pople ϕ pucker parameter for β-D-Glc*p*NAc in aaGN (a, b) and 1RU of *Sf* 7a (c, d). Polar projections of the Cremer-Pople puckering sphere for β-D-Glc*p*NAc in aaGN (e, f) and the 1RU of *Sf* 7a (g, h). Simulations with CHARMM36 have a white background and with GLYCAM06 a grey background.

7.3 Discussion

The β-D-Glc*p*NAc ring puckers into boat/skew in all the trisaccharides modelled regardless of the anomeric configuration or orientation of the glycosidic linkages. However, the extent of ring pucker is highly dependant on linkage configuration with α-disubstitution correlated to the most pucker followed, to a lesser extent, by β-disubstitution. When the anomeric configurations at the 3- and 4- positions are not the same (i.e. one is α- and the other β-) we observe the least amount of β-D-Glc*p*NAc ring pucker.

3,4-disubstituted β-D-Glc*p*NAc puckering occurs in both force fields. All GLYCAM06 simulated structures had faster ring pucker transition rates than those modelled with CHARMM36, which is consistent with previous simulations of monosaccharides.[217] These results suggest that the energy required to transition between pucker states is highly sensitive to anomeric configuration of the 3- and 4- linked residues as well as the force field employed.

It is commonly assumed that the orientations of glycosidic linkages in oligosaccharides or polysaccharides are similar to those in their individual disaccharide components. In general, information obtained from disaccharide units can be extended to oligo- and polysaccharides.[8,10,196] However, modelling of 3,4-disubstituted trisaccharides show that this assumption on additivity does not always hold; the orientations of both the 3- and 4- glycosidic linkages, as well as the β-D-Glc*p*NAc ring conformations, are different to those of related disaccharides. This could be due to stereoelectronic interactions between the substituting sugars, which cause the glycosidic linkage torsion angles in the trisaccharides to differ from those in the corresponding disaccharides — as is the case for 3,4-disubstituted β-D-Glc*p* and β-D-Glc*p*Nac residues.

For example, Soderman et al. studied glucose trisaccharides with a 3,4-disubstitution pattern and varying anomeric configurations using Metropolis Monte Carlo simulations in conjunction with the HSEA force field.[54] They showed that the trisaccharide α-D-Glcp^{I}-(1→3)-[α-D-Glcp^{I} (1→4)]-β-D-Glc*p*OMe, had “larger conformational changes with multiple energy minima” at the α-(1→4)-linkage compared to the α-(1→3)-linkage and with a correlated dependence between the glycosidic torsion angles ϕ and ψ.[54] They also note that the α-linked glucosyl groups exhibit larger flexibility and greater glycosidic linkage deviations compared to the β-linked groups, demonstrating the effect of anomeric configuration on glycosidic linkage orientation and trisaccharide conformation, which we also observe in our study.

QM and MD simulations of β-D-Glc*p*NAc monosaccharide and its derivatives all indicate that increased substitution of the β-D-Glc*p*NAc ring increases ring stability.[63,64] The $^4C_1$ state is favoured as the amount of energy required to transition to boat/skew increases with ring substitution. Consequently, in microsecond MD simulations greater substitution of β-D-Glc*p*NAc leads to the ring spending more time in $^4C_1$ and making fewer transition to other pucker states. This is not observed in aaG and aaGN where 3,4-disubstitution appears to increase the amount of ring puckering. Puckering also occurrs in baGN, abGN, and bbGN but to a lesser degree.

In 2011 Satelle and Almond reviewed the ring conformations of high-resolution ($\leq$ 2.0 Å) crystallographic Protein Data Bank (PDB) and Cambridge Structural Database (CSD) Glc*p*NAc monosaccharides.[63] They discovered that approximately 97% of the over 3100 occurrences adopted a $^4C_1$ conformer, while 0.5% were $^1C_4$, and 2.6% adopted boat or skew conformations. The PDB structures were observed in non-enzyme structures such as viral hemagglutinin, SNA-II, CD2, and interleukin-19.[282–285] In total, two free Glc*p*NAc monosaccharides were analysed and these were retrieved from the CSD.

In non-$^4C_1$ states ring substituents of β-D-Glc*p*NAc have increased axial characteristics — or are fully axial as in $^1C_4$. Bulky groups in an axial configuration are typically disfavoured in these rings as it can lead to increased 1,3-diaxial steric interactions. In the equatorial configuration these strains do no occur and therefore the $^4C_1$ pucker state is preferred. However, bulky groups substituted at the 3- and 4- positions of β-D-Glc*p*NAc will lead to increased steric crowding around the ring as there is a large group at positions 2, 3, 4 and 5.

This strain can be minimised in two general ways. The first is rotation of the glycosidic linkages at the 3- and 4- positions as well as rotation of the N-Ac and CH_2OH groups. The second is by puckering of the β-D-Glc*p*NAc ring. These changes will likely increase torsional and ring strain but may also decrease steric strain between the ring substituents. This effect is possibly the cause of β-D-Glc*p*NAc pucker in the trisaccharide simulations. The correlation of the glycosidic linkage orientations and β-D-Glc*p*NAc ring pucker further supports this hypothesis. When β-D-Glc*p*NAc adopts a boat/skew conformation, the glycosidic linkages become more flexible and closely resemble the typical disaccharide torsion angles. Puckering into pseudoequatorial conformations, as occurs in boat/skew states, potentially allows for increased flexibility of the glycosidic linkages and reduces the steric interactions of the substituents. The 3- and 4-substitution likely produces the necessary torque to pucker the ring even though this may result in increased 1,3-diaxial interactions and ring strain. These steric and torsional forces are in a constant state of interaction and change as the molecule moves towards lower energy states and are potential contributors to the ring puckering observed.

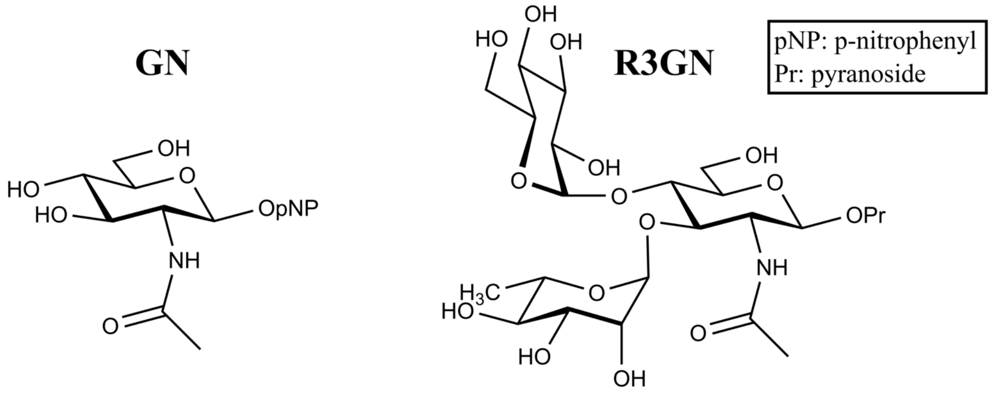
The ring pucker free energies measured can be closely correlated to the structures observed in the trisaccharide MD simulations for both the CHARMM36 and GLYCAM06 force fields. This indicates that the ring puckering behaviour is a consequence of force field parameters, at least in part, and not solely due to the nature of the simulated system. A consequence of this is that either these force fields do not accurately parameterise 3,4-disubstituted β-D-Glc*p*NAc structures, or such puckering is an actual chemical characteristic, or a combination of both. Experimental measurements may be able to provide further insights into these assumptions.

Chapter 8. NMR Conformational Studies of *Shigella flexneri* 7a and 7b O-Ags

In Chapter 6 we modelled the *Sf* 7a and 7b O-Ag and identified unusual puckering of the β-D-Glc*p*NAc ring. The 3,4-disubstitution pattern of the β-D-Glc*p*NAc residue, with the disaccharide glucose attached to the 4-position, likely provides the necessary torque to pucker the ring. The trisaccharide simulations in Chapter 7 also highlight the effects of 2-N-acetylation and anomeric configuration on β-D-Glc*p*NAc ring pucker conformation and dynamics. However, based on these observations alone we are unable to determine whether the simulations reveal a real phenomenon or if the unexpected ring conformations are a result of the force fields.

In this chapter, we analyse the conformations of *Sf* Y, 7a and 7b with NMR spectroscopy experiments. We aim to provide empirical evidence for the conformations observed in the O-Ag simulations and determine whether the unusual puckering observed can be corroborated experimentally. Previous experimental and modelling studies of *S. flexneri* Y, 1, 3, 5 and 6 O-Ags[15,16,22–25] do not reveal puckering of the β-D-Glc*p*NAc ring. As such, experimental measurements of *Sf* 7a and 7b can provide insight into the conformations of the O-Ags and potentially highlight unusual ring conformations.

To aid in the conformational analysis we perform a series of ^{1}H NMR experiments on *Sf* Y, 7a and 7b. We also compare the measurements against those from previous studies of β-D-Glc*p*NAc monosaccharide (GN) and a trisaccharide (R3GN)[125] as these structures, based on their substitution patterns, represents a model for β-D-Glc*p*NAc in *Sf* Y and the branching region of *Sf* 7a and 7b (Figure 8.1).



β-D-Glc*p*NAc-OpNP   α-L-Rha*p*-(1→3)-[α-D-Glc*p*-(1→4)]-β-D-Glc*p*NAc-OPr

Figure 8.1: Structures of GN and R3GN in chemical representation (top) and standard nomenclature (bottom).

8.1 Methods

We investigate O-Ag conformation and dynamics by NMR spectroscopy using several different approaches. We measure the change in anomeric ^{1}H chemical shift with temperature and calculate $^3J_{H,H}$ couplings for the ring protons of β-D-Glc*p*NAc in *Sf* Y and 7a by 1D ^{1}H, ^{1}H-TOCSY. We perform NOESY experiments on *Sf* 7a and 7b to identify through space coupling between proton pairs. Finally, we compare NOESY derived ^{1}H-^{1}H distances with distances calculated from the MD simulations of *Sf* 7a and 7b using a root mean square difference analysis.[*]

8.1.1 Sample preparation

The O-Ag from *S. flexneri* serotype Y (10 mg), 7a (8 mg) and 7b (10 mg) were separately deuterium-exchanged by dissolving the samples in excess of 99.9% D_2O followed by freeze-drying. The freeze-dried samples were then dissolved in 0.5 mL 99.9% D_2O added to a 5 mm NMR tube, freeze-dried, and re-dissolved in 0.6 mL of 99.99% D_2O; sodium 3-trimethylsilyl-(2,2,3,3-2H4)-propanoate (TSP) was added as internal reference (δ^1H 0.00 ppm).

8.1.2 ^{1}H NMR and ^{13}C NMR measurements

1D ^{1}H, ^{1}H TOCSY experiments with mixing times of 30, 60, 90 and 120 ms were recorded on a 700 MHz Bruker AVANCE III spectrometer equipped with a TCI Z-Gradient Cryoprobe.

---

[*] All NMR experiments were performed in collaboration with the Widmalm Research Group at the Department of Organic Chemistry, University of Stockholm.[286]

Selective irradiation was achieved using a R-SNOB pulse with a 100 ms excitation time centred on the H1-Glc*p*NAc resonance. The $^3J_{H,H}$ measurements of the Glc*p*NAc-residue of *Sf* Y and *Sf* 7a from ^{1}H-NMR and 1D ^{1}H, ^{1}H TOCSY were all processed using a Lorentz-Gaussian (lb= –3 and gb=0.15). The TOCSY spectra of *Sf* Y and 7a were made through irradiation of H1-Glc*p*NAc. Relevant $^3J_{H,H}$ couplings were calculated from position H1 as well as the total width of the peaks from positions H2, H3 and H4.

8.1.3 ^{1}H NMR temperature studies

^{1}H NMR temperature studies were performed of *S. flexneri* Y and 7a O-Ag at 5 K increments between 278.15 K and 318.15 K. An additional measurement for *Sf* 7a was made at 343.15 K. An external TSP reference was used. Methanol-d4 was employed for temperature calibration. Temperature studies on *Sf* 7b were not be performed due to significant peak overlap in the ^{1}H NMR spectra.

8.1.4 ^{1}H NMR nuclear Overhauser effect spectroscopy

^{1}H NMR chemical shifts assignments of the serotypes 7a O-Ag have been previously reported.[52] Partial assignments for *Sf* 7b have also been reported.[53] Proton – proton cross-relaxation rates were measured using a 2D ^{1}H, ^{1}H-NOESY experiment with a zero-quantum suppression filter[113] on a Bruker Avance 500 MHz, equipped with a 5 mm TCI (^{1}H/^{1}C/^{1}N) Z-Gradient (53.0 Gcm^{-1}) CryoProbe. The experiments were recorded over a spectral width of 7.0 ppm using non-uniform sampling (NUS) at a 20% level of coverage, with 16k×256 data points using 18 scans per T1 increment and a total recycle time between scans corresponding to five times the longest T1. For all experiments, a 40 kHz broad and 20 ms long adiabatic smoothed CHIRP[287] pulse was used for the elimination of zero-quantum coherence.

NOESY and HSQC experiments were recorded at 42 °C for *Sf* 7a to avoid spectral overlap between the HDO-peak and H1-Glc*p*NAc. Experiments at this temperature allows for better resolution of the anomeric protons compared to the published assignment recorded at 70 °C.[52,53] To confirm assignments at the new temperature, ^{1}H, ^{13}C-HSQC and ^{1}H, ^{13}C-HSQC-TOCSY were recorded for *Sf* 7a.

The 2D ^{1}H, ^{1}H-NOESY experiments of serotypes 7a O-Ag were run at two different temperatures. At 290.15 K eleven different cross-relaxation delays (mixing times) of 30, 40, 50, 60, 70, 80, 90, 110, 120, 150 and 200 ms were used. At 315.15 K ten different mixing times of 30, 40, 50, 60, 70, 80, 90, 110, 150 and 200 ms were used. Only the experiments at 315.15 K were used for determining cross-relaxation rates. The 2D ^{1}H, ^{1}H-NOESY experiments of serotypes 7b O-Ag were run at 325.15 K with four mixing times of 30, 50, 60, and 90 ms. Relatively short mixing times were recorded for both O-Ags to minimize the contribution from spin diffusion.

We employed non-uniforming sampling as opposed to uniform sampling for the nOe measurements. In a study comparing the two approaches, NUS was found to have slightly lower cross-relaxation rates, which may be related to the high dynamic range of nOe measurements.[139] However, when calculating ^{1}H-^{1}H distances from cross-relaxation rates we employ a similar approach, which results in the "effective inter-nuclear distances derived from US or NUS ^{1}H, ^{1}H-NOESY data to become highly similar under equal measuring time conditions".[139]

Integral volumes were calculated using the TOPSIN v4.2 software package. The F1 and F2 radii of each integrated peak (or set of peaks) were iteratively selected to provide optimal curve fitting. Cross-relaxation values were calculated using the Classical, Normalised Classical, PANIC (Peak Amplitude Normalization for Improved Cross-Relaxation)[116,117] and Dixon[118] approaches (as described in Chapter 3). The measured nOe cross-peak volumes at different mixing times were fitted by linear regression to maximise the r^2 value. AutoPeak volumes at different mixing times were fitted with a standard exponential function.

^{1}H-^{1}H cross-relaxation rates (σ) for the observed proton pairs were calculated for *Sf* 7a utilizing the classical and PANIC approach to generate nOe build-up curves. The Dixon approach was used to obtain the cross-relaxation rates for the resolved proton pairs of *Sf* 7b. We derive the cross-relaxation rates for the classical and normalised classical plots from the first order term of the second order polynomial. For the PANIC the plots the cross-relaxation is equal to the gradient of the line of best fit and for the Dixon plots the cross-relaxation is obtained from the y-intercept. The cross-relaxation rates of the resolved correlations were obtained from polynomial fittings with $r^2 > 0.97$ while ensuring that the y-intercept was within 5% of the value at the longest mixing time. Visual inspection of the plots provided further assurance of the appropriateness of the fitting procedure.[138,139]

Standard errors were calculated for all cross-relaxation rates and experimentally determined proton-proton distances. Errors in σ were calculated using the bootstrap resampling procedure and are reported as one standard deviation. Errors were subsequently estimated for the experimentally determined proton-proton distances based on a ±1 standard deviation.[138,139]

^{1}H-^{1}H distances

The approach herein to measure ^{1}H-^{1}H distances is similar to those used in studies of oligo- and polysaccharides.[110,135,139,242] The cross-relaxation rates obtained for the different proton pairs were used to calculate their corresponding distances, r_{ij}, in accordance with the isolated spin pair approximation[288] by:

$$r_{ij} = r_{ref}(\sigma_{ref}/\sigma_{ij})^{1/6} \tag{8.1}$$

where r_{ref}, is the reference ^{1}H-^{1}H distance, σ_{ref} is the reference cross-relaxation rate and σ_{ij} is the proton pair cross-relaxation rate. Effective ^{1}H-^{1}H distances, $r_{ij,eff}$, from the *Sf* 7a and 7b MD simulations were calculated by:

$$r_{ij,eff} = \langle r_{ij}^6 \rangle^{-1/6} \tag{8.2}$$

The reference distances, r_{ref}, are also calculated from the MD simulations by:

$$r_{ref} = \langle r_{ij}^6 \rangle^{-1/6} \tag{8.3}$$

All ^{1}H-^{1}H distances from MD simulations of *Sf* 7a and 7b were calculated by averaging over all saved time frames, excluding the first 100 ns, which is taken as equilibration. Only proton pairs from the four central repeating units were used in the measurements. The H1-Glc$^{\mathrm{I}}$ – H2-Glc$^{\mathrm{I}}$ (2.47 Å) distance was used as reference for methine-methine correlations. Distances from methyl-methine correlations were obtained using Me-Rha$^{\mathrm{II}}$ – H5-Rha$^{\mathrm{II}}$ (2.60 Å) as the reference. Reference distances were calculated using the effective ^{1}H-^{1}H distance from the MD simulation of the respective O-Ag. Additional reference proton pairs cross-peaks could not be clearly and unambiguously identified from the nOe spectra.

We compare ^{1}H-^{1}H distances calculated from the MD simulations of *Sf* 7a and 7b against distances derived from nOe spectroscopy of the O-Ags using two approaches. In the first approach we directly compare distances calculated from the four central repeating units of the entire MD simulation trajectory (excluding the first 100 ns) against the NOESY derived distances. In the second approach we combine repeating units that have *β*-D-Glc*p*NAc in either $^4C_1$ or boat/skew states across a range of different ratios — from 97.5:2.5 to 70:30 $^4C_1$:boat/skew in 2.5% increments. We perform these measurements in a multi-step process:

1. Extract all repeating units from the MD simulation trajectory that only contain *β*-D-Glc*p*NAc in $^4C_1$.
2. Randomly extract the required number of repeating units from the MD simulation trajectory that only contain *β*-D-Glc*p*NAc in boat/the skew states.

3. Combine the randomly selected repeating units from 1) and 2) to obtain the required ratio of $^4C_1$:boat/skew β-D-Glc*p*NAc.
4. Calculate ^{1}H-^{1}H distances from this combined group of repeating units using Equation 8.2.

Finally, we calculate and compare the RMSD between ^{1}H-^{1}H distances calculated from these measurements against the NOESY derived distances for both O-Ags. The RMSD calculations allows for determining which distances change significantly with molecular conformation which distances do not.

8.2 Results

The NMR conformational studies of *Sf* 7a and 7b are divided into several parts. We begin with ^{1}H NMR temperature measurements of *Sf* Y and 7a followed by an analysis of $^3J_{H,H}$ coupling on the β-D-Glc*p*NAc residue of the two O-Ags. *Sf* 7b is not included in these studies as the ^{1}H NMR peaks of interest are not sufficiently defined for accurate measurements. Finally, we interpret the NOESY spectra of *Sf* 7a and 7b, calculate ^{1}H-^{1}H derived from the nOe correlations, and compare these experimental ^{1}H-^{1}H distances against those determined from the respective MD simulations.

8.2.1 ^{1}H NMR temperature studies of *S. flexneri* Y and 7a O-Ags

Standard 1D ^{1}H NMR measurements, to calibrate the spectrometer and minimise peak overlap, showed significant change of chemical shift for several anomeric protons as a function of temperature. This behaviour was not observed for *Sf* Y or other *S. flexneri* O-Ags. We therefore performed ^{1}H NMR temperature studies of *Sf* Y and 7a to measure the temperature dependence of the anomeric proton chemical shift. However, the anomeric proton chemical shift peaks for *Sf* 7b are not well defined across the temperature range and the O-Ag is not included in this analysis.

Sf 7a shows a greater change in chemical shift as a function of temperature ($\Delta\delta/\Delta T$) for the anomeric protons compared to *Sf* Y (Figure 8.2). H1 of β-D-Glc*p*NAc has a $+0.3$ ppb/°C for *Sf* Y and $+1.3$ ppb/°C for *Sf* 7a (Figure 8.2: b, d), a large difference of 1 ppb/°C. Similarly, there is a 1 ppb/°C difference for H1-Rha*p*$^{\mathrm{I}}$ at $+0.7$ ppb/°C for *Sf* Y vs -0.3 ppb/°C for *Sf* 7a (Figure 8.2: b, c). The change in chemical shift for H1-Rha*p*$^{\mathrm{II}}$ in the two O-Ags is less than $+0.2$ ppb/°C for both O-Ags. However, H1-Rha*p*$^{\mathrm{III}}$ also shows a significant difference with -0.4 ppb/°C for *Sf* Y and -1.1 ppb/°C for Sf 7a (Figure 8.2: a, c).

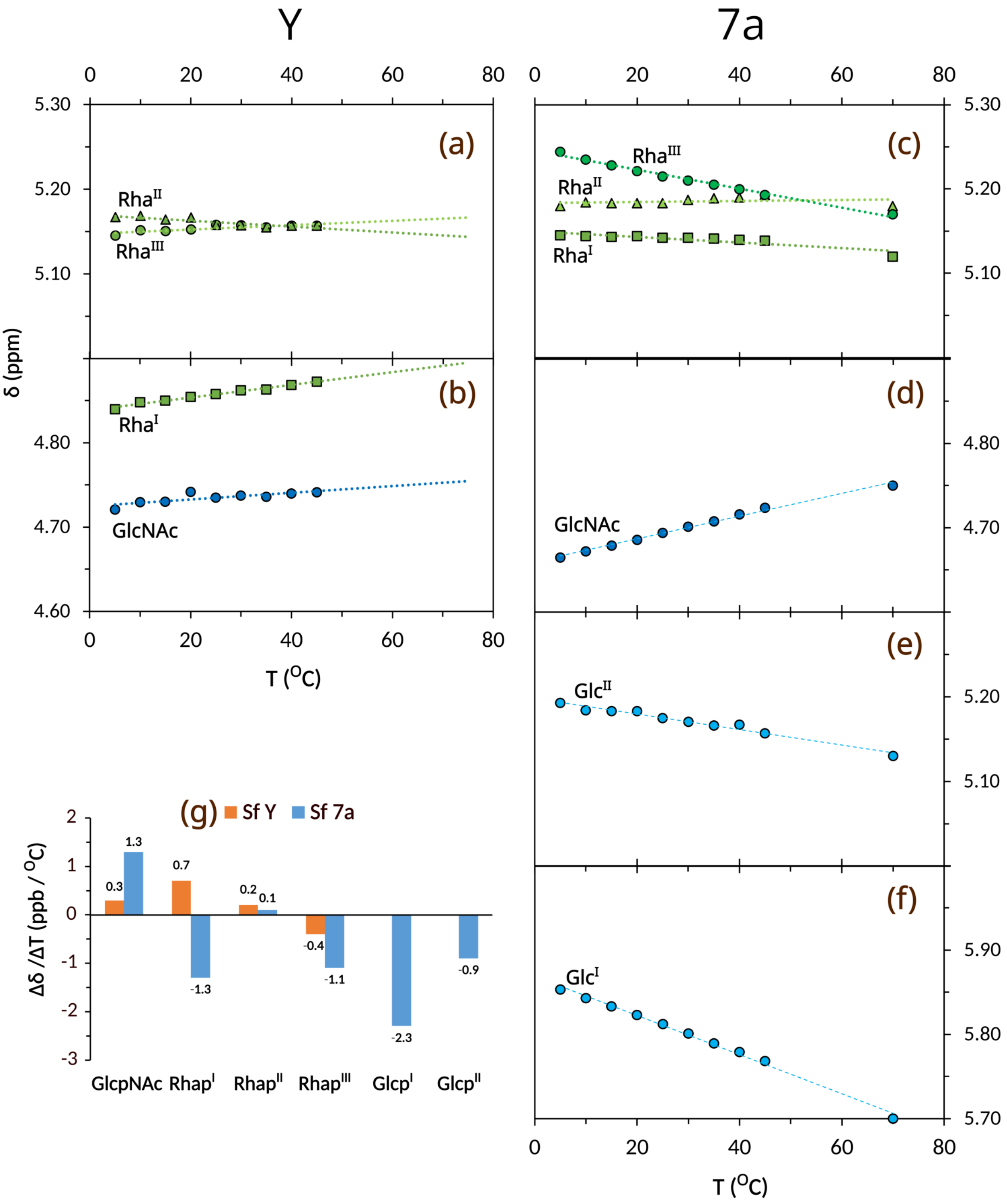


Figure 8.2: (a - f) Change in ^{1}H NMR chemical shift between the measured temperature range for the anomeric protons in *S. flexneri* Y and 7a O-Ags. (g) Change in ^{1}H NMR chemical shift as a function of temperature ($\Delta\delta/\Delta T$, ppb/°C) between the measured temperature range for the anomeric protons in *S. flexneri* 7a and Y O-Ags. For all anomeric protons, with the exception of H1-Rhap^{II}, the magnitude of $\Delta\delta/\Delta T$, ppb/°C is greater in *Sf* 7a compared to *Sf* Y.

A significant temperature dependence was also observed for the branched H1-Glc*p*$^{\mathrm{I}}$ and H1-Glc*p*$^{\mathrm{II}}$ residues in *Sf* 7a. H1-Glc*p*$^{\mathrm{I}}$ (Figure 8.2-e) has a change of −2.3 ppb/°C while H1-Glc*p*$^{\mathrm{II}}$ has a lesser but notable correlation to temperature (Figure 8.2-f) at −0.9 ppb/°C.

These results demonstrate a correlation between the change in anomeric proton chemical shift with distance from the O-Ag branch point (Figure 8.2-g) — the further the anomeric proton from the branch point on *β*-D-Glc*p*NAc, the smaller the difference in chemical shift when compared to *Sf* Y. The large change in chemical shift of H1-Glc*p*$^{\mathrm{I}}$ and H1-Glc*p*NAc may indicate conformational flexibility at the branch point of the O-Ag.[110]

8.2.2 $^3J_{H,H}$ coupling in *β*-D-Glc*p*NAc

The magnitude of the $^3J_{H,H}$ coupling is a function of the torsion angle between the vicinal proton pairs, as described by the Karplus equation (Equation 3.1),[107] and can provide an indirect measure of significant distortions in the *β*-D-Glc*p*NAc ring.[66,289] We measure $^3J_{H,H}$ couplings for *β*-D-Glc*p*NAc in *Sf* Y and 7a (Figure 8.3) and perform a comparative analysis with *β*-D-Glc*p*NAc in R3GN and GN (Figure 8.4).[125]

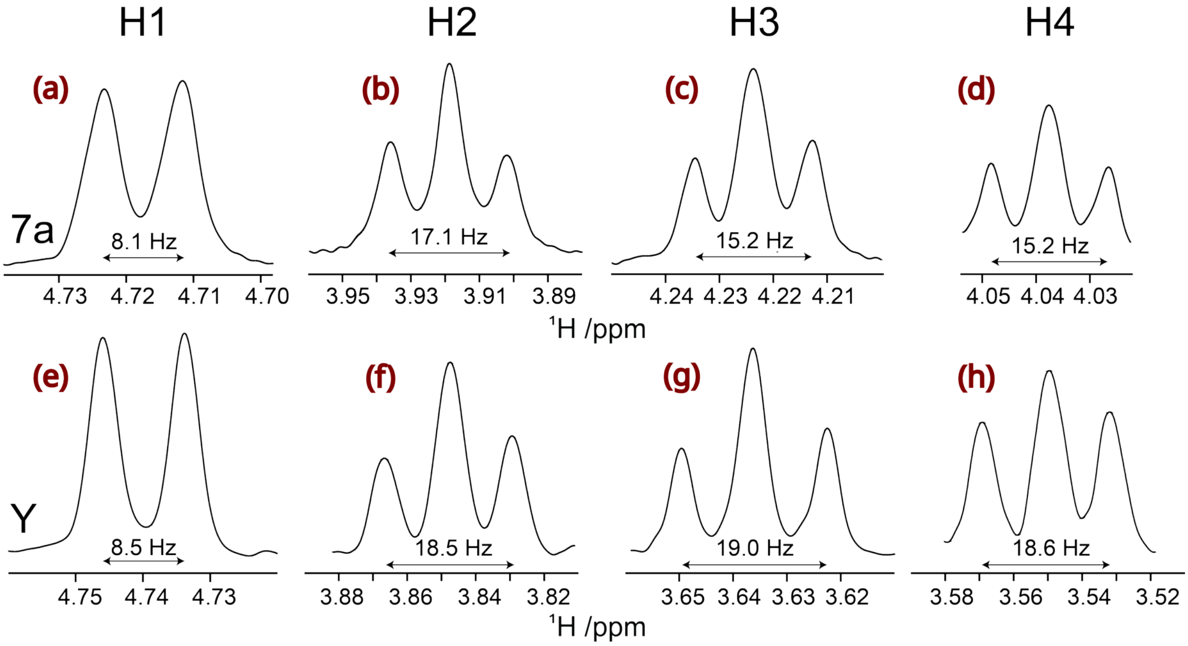


Figure 8.3: Selected regions of the 1D ^{1}H, ^{1}H-TOCSY spectra of the H1 to H4 resonances from N-acetyl-D-glucosamine residue in *Sf* 7a (top) and Y (bottom). Double arrows indicate the value of (a and e) ^{3}J H1, H2; the total width of the dd corresponding to (b and f) ^{3}J H2, H1 + ^{3} H2, H3; (c and g) ^{3} H3, H2 + ^{3} H3, H4; and (d and h) ^{3} H4, H3 + ^{3} H4, H5.[286]

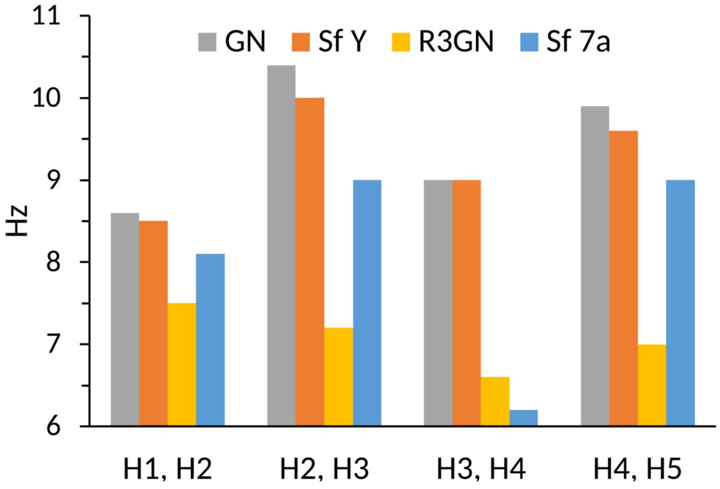


Figure 8.4: Scalar $^{3}J_{H,H}$ coupling constants (Hz) for selected protons in the Glc*p*NAc residue of *β*-D-Glc*p*NAc-OpNP (GN), *α*-L-Rha*p*-(1→3)-[*α*-D-Glc*p*-(1→4)]-*β*-D-Glc*p*NAc-OPr (R3GN), *S. flexneri* 7a, and Y.[54,125]

The *Sf* Y and 7a H1 doublet $^{3}J_{H,H}$ couplings (Figure 8.3: a, e) are similar at 8.5 Hz and 8.1 Hz respectively. As are the H2 doublet-of-doublet $^{3}J_{H,H}$ couplings (Figure 8.3: b, f): *Sf* Y = 17.1 Hz and *Sf* 7a = 8.1 Hz, indicating that the H1 – H2 and H2 – H3 torsion angles are similar between the two O-Ags. In contrast, the *Sf* Y and 7a H3 (Figure 8.3: c, g) and H4 (Figure 8.3: d, h) doublet-of-doublet $^{3}J_{H,H}$ couplings are significantly different: H3 *Sf* Y = 19.0 Hz, *Sf* 7a = 15.2 Hz; H4 *Sf* Y = 18.6, *Sf* 7a = 15.2 Hz. These measurements suggest notable differences in the stereoelectronic environments at H3 and H4 between the two O-Ags.

The largest difference in $^{3}J_{H,H}$ coupling is for the H3-H4 proton pair. In *Sf* Y and GN the value is 9.0 Hz, indicating that substitution at the 3-position by Rha^{I} has no discernible effect on the coupling magnitude. In comparison, the $^{3}J_{H,H}$ H3-H4 coupling in *Sf* 7a and R3GN are smaller at 6.2 Hz and 6.6 Hz respectively. The difference in the *β*-D-Glc*p*NAc H3-H4 (and to an extent the H2-H3) $^{3}J_{H,H}$ coupling between *Sf* Y and 7a can be rationalised by differences in ring distortion between these proton pairs. In the $^{4}C_{1}$ conformation H3-H4 are antiperiplanar with a torsion angle of approximately 170° (Figure 8.5-a), while in boat/skew conformations the angle may be smaller (Figure 8.5-b). Stereoelectronic changes caused by substitution at the 4-position of *β*-D-Glc*p*NAc in *Sf* 7a may not solely account for the difference in values. Instead, changes in the ring proton torsion angles, such as through distortions of the *β*-D-Glc*p*NAc ring, may significantly contribute to the coupling observed.

The *β*-D-Glc*p*NAc H1-H2 $^{3}J_{H,H}$ couplings for GN, *Sf* Y and 7a have similar values: 8.6 Hz, 8.5 Hz and 8.1 Hz respectively; but smaller for R3GN at 7.5 Hz. The magnitude of the H2-H3 $^{3}J_{H,H}$ coupling in GN (10.4 Hz) and *Sf* Y (10 Hz) are similar. However, the H2-H3 $^{3}J_{H,H}$

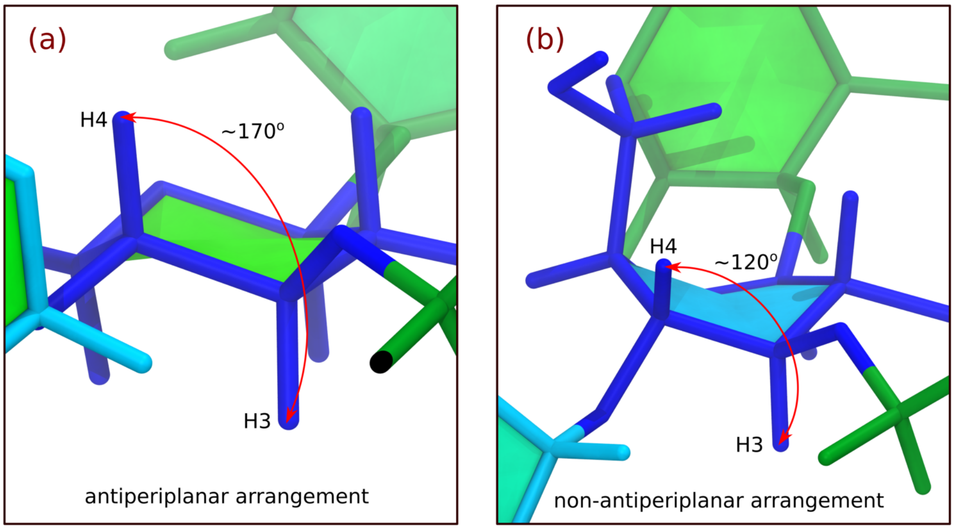


Figure 8.5: The H3-H4 torsion angle from two of the possible conformations of the β-D-Glc*p*NAc ring in *Sf* 7a. The anti-periplanar arrangement (a) will provide a larger $^3J_{H,H}$ coupling compared to the non-antiperiplanar arrangement (b).

coupling in *Sf* 7a is 9.0 Hz; as is shown in the H2 doublet-of-doublet (Figure 8.3). The H2-H3 $^3J_{H,H}$ coupling in R3GN is even smaller at 7.2 Hz. These measurements indicate that the torsion angle between H2-H3 is smaller in *Sf* 7a and R3GN, suggesting potential ring distortions between H2-H3, compared to *Sf* Y and GN. ^{1}H, ^{1}H-NOESY experiments can provide further insight into β-D-Glc*p*NAc ring conformation through detection of intra- and inter-residue correlations, and ^{1}H-^{1}H distances.

8.2.3 ^{1}H, ^{1}H-NOESY spectra of *S. flexneri* 7a and 7b O-Ags

We analyse and compare selected regions of the *Sf* 7a and 7b ^{1}H, ^{1}H nOe spectra as shown in Figure 8.6. In *Sf* 7a all trans-glycosidic linkage correlations† apart from H1-Rha*p*^{III} – H2-Rha*p*^{II} are resolved, allowing for analysis of cross-peak intensities. Three trans-glycosidic proton pairs are of similar intensity: H1-Rha*p*^{II} – H2-Rha*p*^{I}, H1-Rha*p*^{I} – H3-Glc*p*NAc, and H1-Glc*p*^{II} – H2-Glc*p*^{I} (Figure 8.6-d), which suggests that the trans-glycosidic distances may be similar. In contrast, the H1-Glc*p*NAc – H2-Rha^{III} cross-peak is stronger (Figure 8.6-c), which points to a shorter distance between the protons. H1-Glc*p*^{I} – H4-GlcNAc, however, has a noticeably weaker cross-peak and may be further apart (Figure 8.6-e).

Cross-peaks for three proton pairs on the *Sf* 7a β-D-Glc*p*NAc ring were identified: H1-Glc*p*NAc-H3-Glc*p*NAc, H1-Glc*p*NAc-H5-Glc*p*NAc, and H3-Glc*p*NAc-H5-Glc*p*NAc, for all

---

† Correlations of glycosidic protons directly across a glycosidic bond.

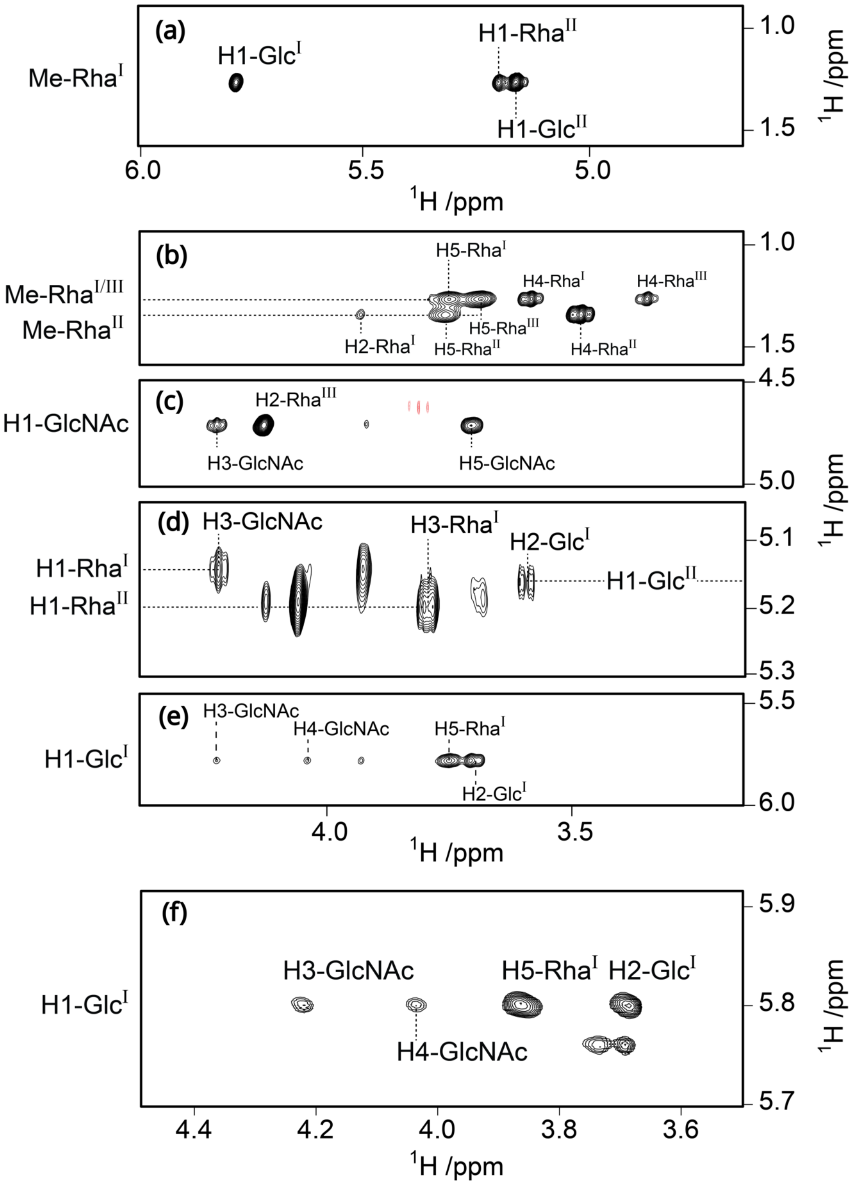


Figure 8.6: Selected regions of ^{1}H, ^{1}H-NOESY spectra at 500 MHz from the O-Ags of *S. flexneri* 7a (a – e) along the F2 dimension and 7b (f) along the F1 dimension for mixing times of 110 and 150 ms respectively with annotations showing relevant intra-residue cross-peaks.[286]

mixing times. The presence of these cross-peaks, together with the $^3J_{H,H}$ coupling measurements, indicate that the β-D-Glc*p*NAc ring does not solely adopt the unusual $^1C_4$ conformation. If $^1C_4$ was present for significant periods of time the H1, H3, and H5 ring protons would be equatorial and the nOe correlations would be notably weaker.[63,66,125]

Inter-residue methyl-methine correlations for Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{I}}$ and Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{II}}$ indicate a syn-orientation of the H1-Rha*p*$^{\mathrm{I}}$ – H3-Glc*p*NAc linkage (Figure 8.6: a, b), as only in the syn-orientation is the Me-Rha*p*$^{\mathrm{I}}$ in sufficiently close proximity to the branched Glc*p* residues. Further, these methyl-methine correlations also suggest that the H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc may be in a syn-orientation — if H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc adopts an anti-orientation for extended periods of time the Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{I}}$ and Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{II}}$ cross-peaks would either be significantly smaller or not present at all.[290,291] Therefore, based on the nOe cross-peaks observed, it is more likely the H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc linkage adopts a syn-orientation as opposed to an anti-orientation.

Additionally, the inter-residue H1-Glc*p*$^{\mathrm{I}}$ – H5-Rha*p*$^{\mathrm{I}}$ cross-peak intensity is similar to the intra-residue H1-Glc*p*$^{\mathrm{II}}$ – H2-Glc*p*$^{\mathrm{I}}$ correlation (Figure 8.6: c, d). Similarly, cross-peak intensities of H1-Glc*p*$^{\mathrm{I}}$ – H3-Glc*p*NAc and H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc are of comparable size (Figure 8.6-e). These observations suggest that each set of proton pairs are equidistant and that the measurements correspond to one or more distinct conformations of *Sf* 7a.

The ^{1}H and ^{13}C NMR spectra of the *Sf* 7b contains extensive peak overlap due to partial O-acetylation of Rha*p*$^{\mathrm{I}}$2Ac.[53] O-acetylation does not significantly change the chemical shifts of the anomeric protons, with the exception of H1-Glc*p*$^{\mathrm{I}}$, which changes from 5.76 ppm to 5.80 ppm (Figure 8.6-f). The same inter-residue cross-peaks are observed in the ^{1}H, ^{1}H NOE spectra of *Sf* 7b as in *Sf* 7a. The peak intensities of H1-Glc*p*$^{\mathrm{I}}$ – H3-Glc*p*NAc and H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc are of similar size suggesting the distances between the protons in each pair are similar between the two O-Ags. Additionally, no N-acetyl nOe contacts were identified in both *Sf* 7a and 7b. Similarly, no Rha*p*$^{\mathrm{I}}$ 2-OAc contacts in *Sf* 7b were present.

8.2.4 *S. flexneri* 7a O-Ag ^{1}H-^{1}H distances

We calculate O-Ag ^{1}H-^{1}H distances from the *Sf* 7a MD simulations and NOESY experiments. ^{1}H-^{1}H distance measurements provide a means for comparing the O-Ag conformations modelled by MD simulation against the O-Ag studied during the NOESY experiments.[135,138,139] Intra- and inter-residue ^{1}H-^{1}H cross-peaks within and between all residues in the NOESY spectra of *Sf* 7a allow for the quantitative analysis of interproton distances by auto- and cross-peak

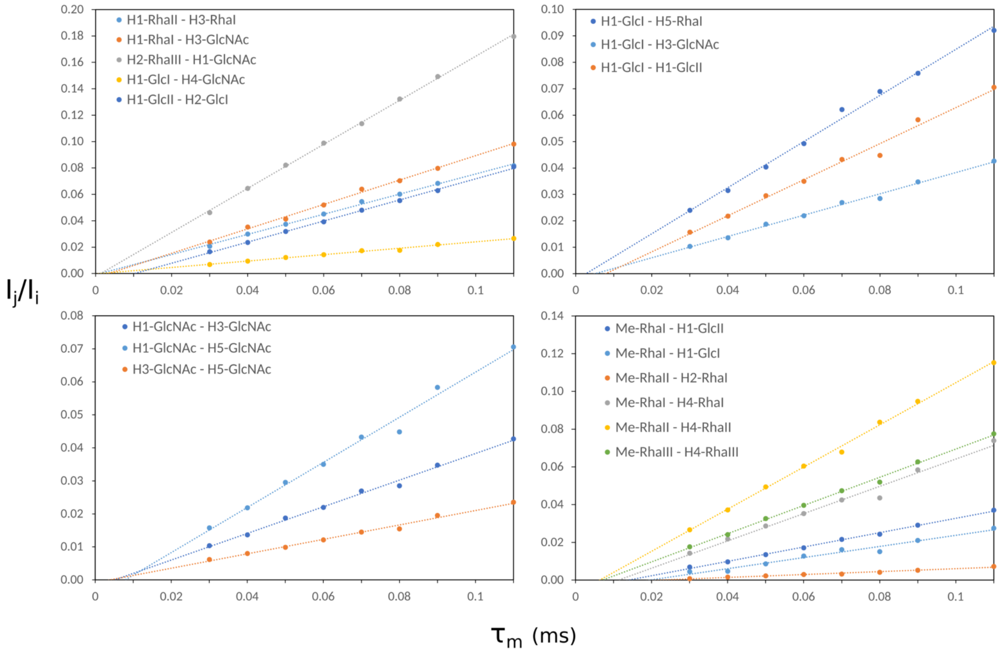


Figure 8.7: ^{1}H, ^{1}H NOESY buildup curves from the *S. flexneri* 7a O-Ag employing the PANIC approach — trans-glycosidic (top-left), inter-residue (top-right), intra-residue (bottom-left), and methyl-methine (bottom-right). Cross-relaxation rates obtained from the slopes of the fitted data.

integration.[135,136,138,139] We calculate cross-relaxation rates (Table B.1) between proton pairs using nOe buildup curves[292] derived from the four approaches (Classical, Normalised Classical, PANIC and Dixon; Equations 3.2 - 3.5).

We expect the classical, normalised classical and PANIC nOe buildup curves to be straight lines that intersect the origin. All three methods provide straight line plots for all cross-peaks but only the normalised classical and PANIC intercept near the origin. Additionally, the Dixon buildup curves have an unusual positive gradient. Typically Dixon curves have a negative gradient[118,293–295] indicating that, for increasing mixing times, the cross-peak intensity (η_{IS}) generally rises slower than the product of the auto-peak intensity (η^{ap}) and mixing time (τ_m): $\eta^{ap}(\tau_m)\tau_m$ (Equation 3.5). These results indicate that the normalised classical and PANIC approaches provide a more accurate measure of cross-relaxation rates. However, the PANIC method gives the best fit of the data points (Figure 8.7) and provides the most consistent distance measurements across all measured cross-peaks. It is therefore solely used in the further analysis of this section and from herein referred to as the experimental measure.

Figure 8.8 depicts selected ^{1}H-^{1}H distances and the conformation of residues at the β-D-Glc*p*NAc branch-point in *Sf* 7a; showing the large number of intra- and inter-resdiue nOe

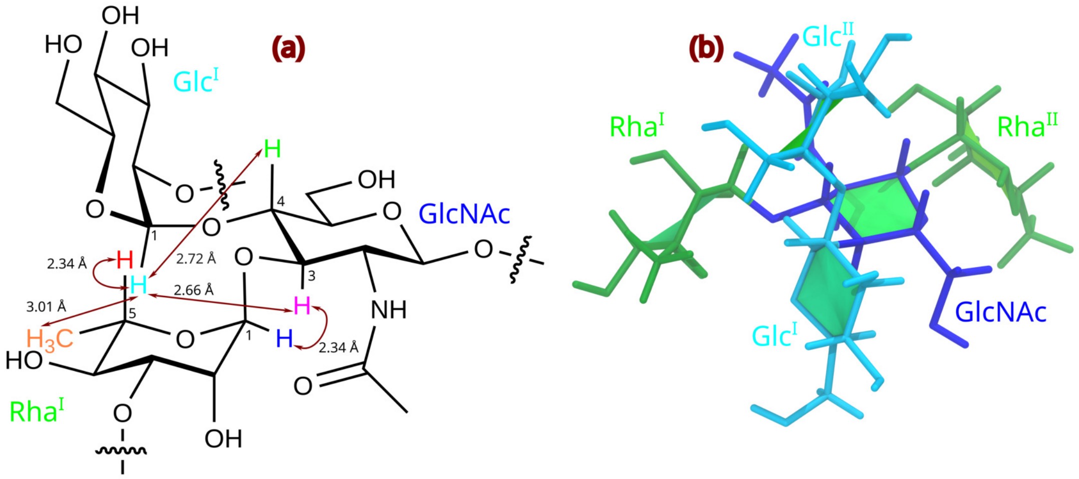


Figure 8.8: (a) Selected ^{1}H-^{1}H distances calculated from the NOESY spectra of *Sf* 7a. The are a large number of intra- and inter- residue nOe interactions, highlighting the short distances between protons on the neighbouring α-L-Rha^{I}, β-D-Glc*p*NAc, and α-D-Glc*p*^{I} residues. (b) Conformations of residues at the β-D-Glc*p*NAc branch-point in *Sf* 7a showing the orientation of the α-D-Glc*p*^{I} and α-D-Glc*p*^{II} side chains towards α-L-Rha^{I} and away from C6 of β-D-Glc*p*NAc, as indicated by the ^{1}H-^{1}H distance measurements.

interactions and the orientations of the side chain. The experimentally derived backbone trans-glycosdic distances, H1-Rha^{II} – H3-Rha^{I} (2.39 Å), H1-Rha^{I} – H3-Glc*p*NAc (2.34 Å), and H1-Glc*p*NAc – H2-Rha^{III} (2.10 Å); fall in a range that indicate the linkages are predominantly or entirely in syn-orientation (Figure 8.8-a).[15,77] The side chain H1-Glc*p*^{II} – H2-Glc*p*^{I} distance (2.37 Å) also correspond to a syn-orientation. These distance align with trans-glycosidic distances of *S. flexneri* serotypes 1, 2, 3, and 5.[15,77]

In contrast, we cannot readily determine whether the H1-Glc*p*^{I} – H4-Glc*p*NAc linkage is in solely the syn- or anti-orientation based on the experimentally measured distances alone. A distance of 2.72 Å indicates a linkage that is too long for one that is in the syn-orientation, which would be nearer to 2.2 Å,[15] and too short for one that is entirely in the anti-orientation, which would be nearer to 3.5 Å.[296] The glycosidic linkage may either alternate between syn- and anti- states (and we measure the average distance from these two states) or remain at a position between the two orientations, although the latter would be unusual.

Three inter-residue distances were measurable: H1-Glc*p*^{I} – H5-Rha^{I} (2.34 Å), H1-Glc*p*^{I} – H3-Glc*p*NAc (2.66 AA), and H1-Glc*p*^{I} — H1-Glc*p*^{II} (2.44 Å); and provide insight into the orientation of the glycosidic linkages and conformation around the β-D-Glc*p*NAc branch-point. All these measurements disfavour an anti-orientated H1-Glc*p*^{I} – H4-Glc*p*NAc linkage as the distances would otherwise be much longer. The three intra-residue β-D-Glc*p*NAc distances are

Table 8.1: ^{1}H, ^{1}H distances calculated from the MD simulations (r_{MD} in Å) and NMR nOe experiments (r_{exp} in Å) from the *S. flexneri* 7a O-Ag. Calculated from ^{1}H,^{1}H-NOESY NMR buildup curves using the PANIC approach. Computed distances from the MD simulation are based on the four inner RUs of the O-Ag and are subdivided into trans-glycosidic, inter-residual, intra-residual, methyl-methine, proton pairs.

Proton Pair		r_{exp}	r_{MD}	% diff.	**Proton Pair**		r_{exp}	r_{MD}	% diff.
Trans-glycosidic					**Intra-residue**				
H1-Rha$^{\mathrm{II}}$	H3-Rha$^{\mathrm{I}}$	2.39 ± 0.029	2.32 ± 0.003	3%	H1-Glc*p*NAc	H3-Glc*p*NAc	2.44 ± 0.030	2.97 ± 0.001	20%
H1-Rha$^{\mathrm{I}}$	H3-Glc*p*NAc	2.34 ± 0.029	2.34 ± 0.003	0%	H1-Glc*p*NAc	H5-Glc*p*NAc	2.63 ± 0.032	2.81 ± 0.003	6%
H1-Glc*p*NAc	H2-Rha$^{\mathrm{III}}$	2.10 ± 0.029	2.30 ± 0.003	9%	H3-Glc*p*NAc	H5-Glc*p*NAc	2.95 ± 0.038	3.53 ± 0.001	18%
H1-Glc*p*$^{\mathrm{I}}$	H4-Glc*p*NAc	2.72 ± 0.035	2.29 ± 0.004	17%	**Inter-residue**				
H1-Glc*p*$^{\mathrm{II}}$	H2-Glc*p*$^{\mathrm{I}}$	2.37 ± 0.030	2.38 ± 0.004	2%	H1-Glc*p*$^{\mathrm{I}}$	H5-Rha$^{\mathrm{I}}$	2.34 ± 0.028	2.79 ± 0.003	18%
Methyl-methine					H1-Glc*p*$^{\mathrm{I}}$	H3-Glc*p*NAc	2.66 ± 0.034	2.68 ± 0.003	1%
Me-Rha$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{I}}$	3.14 ± 0.023	3.29 ± 0.002	5%	H1-Glc*p*$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{II}}$	2.44 ± 0.031	2.28 ± 0.004	7%
Me-Rha$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{II}}$	3.01 ± 0.021	3.23	7%	**Reference**				
Me-Rha$^{\mathrm{II}}$	H2-Rha$^{\mathrm{I}}$	3.93 ± 0.030	3.93	3%	H1-Glc*p*$^{\mathrm{I}}$	H2-Glc*p*$^{\mathrm{I}}$	-	2.45 ± 0.001	-
Me-Rha*p*$^{\mathrm{I}}$	H4-Rha*p*$^{\mathrm{I}}$	2.70 ± 0.020	2.87	6%	H1-Glc*p*$^{\mathrm{II}}$	H2-Glc*p*$^{\mathrm{II}}$	-	2.49 ± 0.002	-
Me-Rha*p*$^{\mathrm{II}}$	H4-Rha*p*$^{\mathrm{II}}$	2.51 ± 0.017	2.86	13%	H1-Rha*p*$^{\mathrm{III}}$	H2-Rha*p*$^{\mathrm{III}}$	-	2.47 ± 0.001	-
Me-Rha*p*$^{\mathrm{III}}$	H4-Rha*p*$^{\mathrm{III}}$	2.69 ± 0.019	2.87	6%	H5-Rha*p*$^{\mathrm{II}}$	Me-Rha*p*$^{\mathrm{II}}$	-	2.60 ± 0.002	-

interesting as explained. The H1-Glc*p*NAc – H3-Glc*p*NAc (2.44 Å) and H1-Glc*p*NAc – H5-Glc*p*NAc (2.63 Å) distances are consistent with a $^4C_1$ ring. However, the H3-Glc*p*NAc – H5-Glc*p*NAc (2.95 Å) distance is longer in comparison, which would only be possible with a level of ring distortion between C3 and C5.

Six methyl-methine distances were recorded. Three of these are inter-residue and conformationally diagnostic. The Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{II}}$ (3.14 Å) and Me-Rha*p*$^{\mathrm{I}}$ – H1-Glc*p*$^{\mathrm{I}}$ (3.01 Å) distances indicate the Glc*p*$^{\mathrm{II}}$ residue is rotated toward Rha*p*$^{\mathrm{I}}$ and away from C6 of *β*-D-Glc*p*NAc (Figure 8.8-b). There is also a long range nOe contact between Me-Rha*p*$^{\mathrm{II}}$ – H2-Rha*p*$^{\mathrm{I}}$ (3.93 Å).

We now compare the experimentally measured ^{1}H-^{1}H distances in *Sf* 7a with those calculated from the MD simulations (Table 8.1, Table B.2, and Table B.6). The percentage difference for most proton pair distances are quite large with an average of 9% for methine-methine pairs and 7% for methyl-methine pairs. However, there is reasonably good agreement for almost all trans-glycosidic proton pairs, with an average difference of approximately 4%. The only exception is H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc where the difference is 17 %. These results suggest that, with the exception of H1-Glc*p*$^{\mathrm{I}}$ – H4-Glc*p*NAc, the glycosidic linkage orientations of the experimentally studied and modelled O-Ags may be similar.

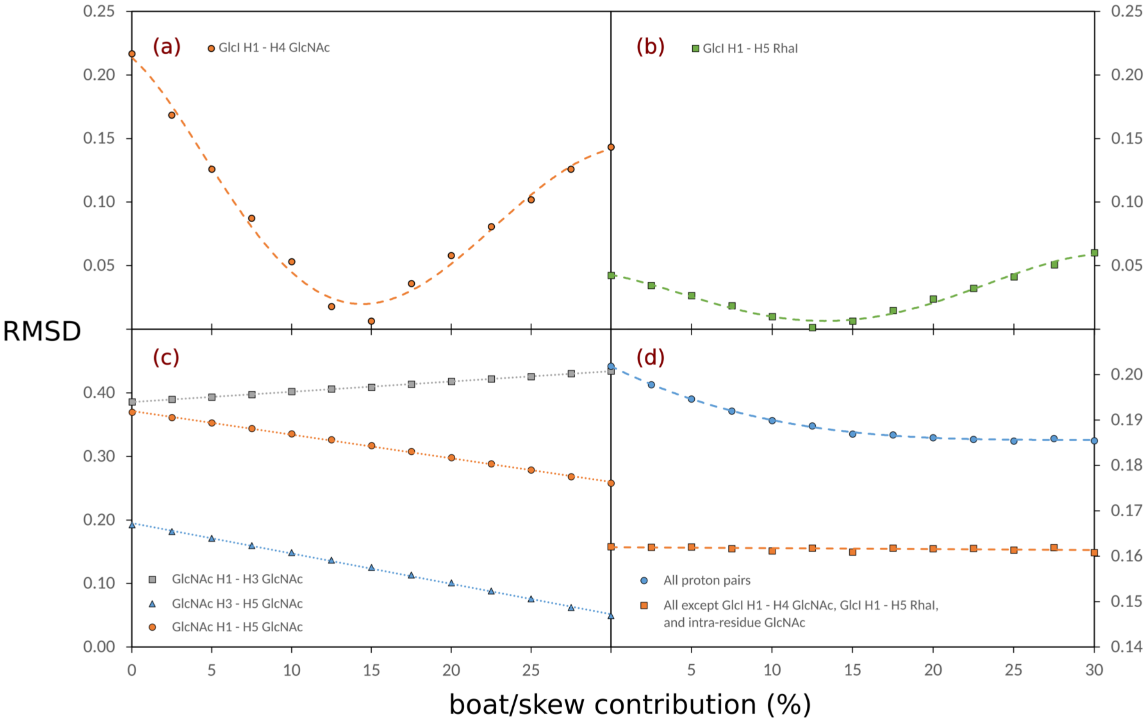


Figure 8.9: Root mean square difference for selected ^{1}H-^{1}H distances calculated by NOE experiments and MD simulations of *Sf* 7a. Measurements made at different ratios of $^4C_1$ and boat/skew conformations of β-D-Glc*p*NAc. The lower the RMSD the more closely the ^{1}H-^{1}H distances between the two methods are aligned. RMSD distance calculations made for the following proton pairs: (a) H1-Glc*p*^{I} – H4-Glc*p*NAc; (b) H1-Glc*p*^{I} – H5-Rha^I; (c) intra-residue β-D-Glc*p*NAc proton pairs; (d) All proton pairs (blue curve), and all proton pairs except H1-Glc*p*^{I} – H4-Glc*p*NAc, H1-Glc*p*^{I} – H5-Rha^I and intra-residue β-D-Glc*p*NAc (orange curve).

The intra-residue β-D-Glc*p*NAc distances as well as H1-Glc*p*^{I} — H5-Rha*p* also have large differences of more than 15%. The ^{1}H-^{1}H distances for all other methyl-methine and inter-residue methine-methine distances are within 10%. The experimentally measured distances do not suggest puckering of the β-D-Glc*p*NAc ring to the extent observed in the MD simulations and, therefore, these large differences are not unexpected.

In the *Sf* 7a MD simulation the β-D-Glc*p*NAc spends ~70% of the simulation time in boat/skew conformations. However, the experimentally measured ^{1}H-^{1}H distances do not indicate puckering of β-D-Glc*p*NAc to such an extent. The RMSD between distances from the *Sf* 7a nOe experiments and the *Sf* 7a MD simulation for different $^4C_1$ and boat/skew ratios is listed in Figure 8.9. Some distances have a large change as a function of boat/skew percentage while other distances remain relatively constant. The H1-Glc*p*^{I} – H4-Glc*p*NAc RMSD (Figure 8.9-a) changes significantly, decreasing steadily from a 0% boat/skew contribution to a minimum at ~15%, and then increasing to a 30% contribution. Similarly, the H1-Glc*p*^{I} —

H5-Rha*p* RMSD decreases between 0% and 12.5% then increases consistently up to 30% (Figure 8.9-b). The experimental and simulation distances for these two proton pairs are closest with a small percentage of boat/skew *β*-D-Glc*p*NAc residues present. If *β*-D-Glc*p*NAc is solely in $^4C_1$ or, alternatively, adopts a large percentage boat/skew conformers, the RMSD of the distances are much larger. This suggests that *β*-D-Glc*p*NAc does not remain solely in $^4C_1$ nor does it spend a significant amount of time in boat/skew conformations. Rather it remains predominantly in $^4C_1$ with short periods in boat/skew conformations.

In contrast, the intra-residue *β*-D-Glc*p*NAc RMSD values have a different pattern (Figure 8.9-c). The H1–Glc*p*NAc – H3-Glc*p*NAc RMSD increases by a small amount across the range of boat/skew contributions, indicating that the H1–Glc*p*NAc – H3-Glc*p*NAc distance does not change significantly as a function of the boat/skew percentage. However, the H1-Glc*p*NAc – H5-Glc*p*NAc and H3-*β*-D-Glc*p*NAc – H5-*β*-D-Glc*p*NAc RMSD values both fall by a large amount with increased boat/skew contribution, indicating that these distances more closely align with a boat/skew conformation, with ring distortion between C3 and C5.

The RMSD value for all other proton pairs (excluding H1-Glc*p*$^{\text{I}}$ – H4-Glc*p*NAc, H1-Glc*p*$^{\text{I}}$ – H5-Rha$^{\text{I}}$ and intra-residue *β*-D-Glc*p*NAc) remain relatively constant (Figure 8.9-d, orange curve) — their measured ^{1}H-^{1}H distances do not appear to change significantly with different *β*-D-Glc*p*NAc pucker states. For all proton distances measured, the RMSD falls from a 0% boat/skew contribution to approximately 15% after which it begins to level-off (Figure 8.9-d, blue curve).

Overall the lowest RMS, and therefore the smallest difference in distances, occurs when approximately 85% of repeating units from the MD simulation trajectory have *β*-D-Glc*p*NAc in $^4C_1$ with the remainder 15% of repeating units contain *β*-D-Glc*p*NAc in boat/skew — an 85:15 $^4C_1$:boat/skew mix. In this mix it is likely the boat/skew *β*-D-Glc*p*NAc conformations predominantly adopt $^3S_1$, $B_{1,4}$, and $^5S_1$ states. Nonetheless, in the absence of an accurate molecular model it is difficult to identify which pucker states are exactly populated and to what extent.

8.2.5 *S. flexneri* 7b O-Ag ^{1}H-^{1}H distances

Similar to the 7a O-Ag ^{1}H-^{1}H distance measurements, we calculate O-Ag ^{1}H-^{1}H distances from the *Sf* 7b MD simulations and NOESY experiments. We compare distances calculated from the simulations with those derived from the NOESY experiments. We also perform a comparison between *Sf* 7a and 7b where distances for the same ^{1}H-^{1}H pair can be measured for both O-Ags.

Table 8.2: ^{1}H,^{1}H cross-relaxation rates (σ, $s^{˘1}$) from the O-Ag of *S. flexneri* 7b derived from ^{1}H, ^{1}H-NOESY NMR experiments analysed according to the Dixon approach. Experimentally derived distances (r_{exp} in Å) and effective ^{1}H, ^{1}H distances (r_{MD} in Å). Computed distances from the MD simulation are based on the four inner RUs of the O-Ag. *reference proton-pair.

Proton Pair		σ	r_{exp}	r_{MD}
H1-Glcp^{I}	H3-Glc$p$NAc	0.08 ± 0.002	2.78 ± 0.069	2.69 ± 0.004
H1-Glcp^{I}	H4-Glc$p$NAc	0.08 ± 0.002	2.78 ± 0.069	2.31 ± 0.003
H1-Glcp^{I}	H5-Rha$^{\text{I}}$2Ac	0.26 ± 0.004	2.31 ± 0.057	2.85 ± 0.002
H1-Glcp^{I}	H2-Glcp^{I}*	0.18 ± 0.003	-	2.45 ± 0.001

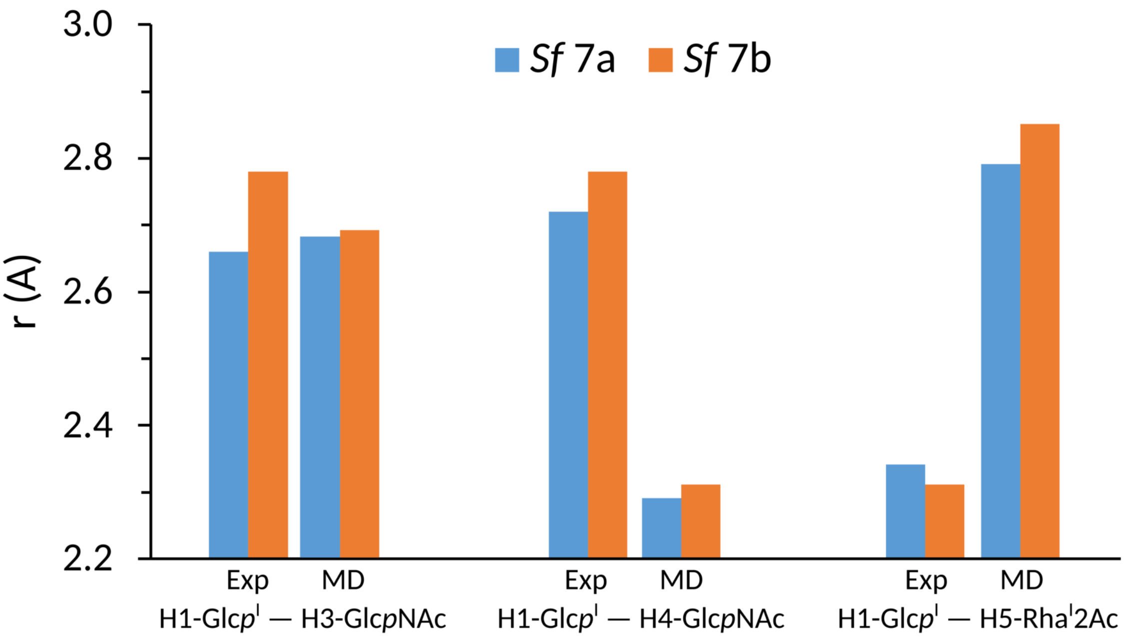


Figure 8.10: ^{1}H-^{1}H distances derived from the nOe experiments and MD simulations of *Sf* 7a and 7b illustrating the similar experimental and MD distances for H1-Glcp^{I} – H3-Glc$p$NAc, and the large differences for H1-Glcp^{I} – H4-Glc$p$NAc and H1-Glcp^{I} – H5-Rha$^{\text{I}}$2Ac. For each proton pair the experimental distances are similar for both O-Ags. Similarly, the distances calculated from the MD simulations of both O-Ags are similar for each proton pair.

The ^{1}H,^{1}H NMR NOESY spectra of *Sf* 7b has extensive peak overlap. However, a small number of peaks were sufficiently separated to allow for a quantitative analysis of three ^{1}H-^{1}H distances. Table 8.2 lists the experimentally measured cross-relaxation rates and distances for the *Sf* 7b NOESY experiments as well as the corresponding effective distances from the *Sf* 7b MD simulations.

There is poor agreement between the distances derived from experiment and modelling for two of the measured ^{1}H-^{1}H distances: H1-Glc*p*$^{\mathrm{I}}$ — H4-Glc*p*NAc ($r_{exp} = 2.78$, $r_{MD} = 2.31$), and H1-Glc*p*$^{\mathrm{I}}$ — H5-Rha$^{\mathrm{I}}$2Ac ($r_{exp} = 2.31$, $r_{MD} = 2.85$). However, the H1-Glc*p*$^{\mathrm{I}}$ — H3-Glc*p*NAc distances are more closely aligned ($r_{exp} = 2.78$, $r_{MD} = 2.69$). While there are only a few *Sf* 7b ^{1}H-^{1}H distance measurements these results suggest significant differences between the O-Ag studied experimentally and the O-Ag modelled by MD simulations, which is similar to the observations of *Sf* 7a.

Direct comparison of the experimentally measured ^{1}H-^{1}H distances between *Sf* 7a and 7b as well as distances from the MD simulation reveal interesting results (Figure 8.10-b). The experimental distances for each proton pair are similar between the two O-Ags. Likewise, the distances derived from the MD simulations of both O-Ags are comparable for each proton pair. Therefore, based on the experimental and MD simulation measurements, it appears that the *β*-D-Glc*p*NAc conformation of the two O-Ags may be similar.

8.3 Discussion

The NMR spectroscopy experiments corroborate, in part, the MD simulation results and further highlight the conformational differences between *Sf* Y, 7a and 7b. In particular, the experiments reveal unusual behaviour of the *β*-D-Glc*p*NAc residue of *Sf* 7a and provide evidence for potential puckering into boat/skew conformations.

Conformational differences between *Sf* Y and 7a are revealed by ^{1}H NMR temperature studies, confirming that the glycosidic proton chemical shifts of *Sf* Y are relatively insensitive to changes in temperature, while in *Sf* 7a they are far more sensitive. On average the glycosidic proton chemical shifts change by approximately 0.4 ppb/°C in *Sf* Y and by over 1.3 ppb/°C in *Sf* 7a. Moreover, the greatest chemical shift changes in *Sf* 7a occur for protons close to the branch-point, suggesting a greater degree of dynamic behaviour in and around the *β*-D-Glc*p*NAc ring when compared to other *Shigella* O-Ags. This could be attributed to the flexibility of the disaccharide side chain.

Potential puckering of the *Sf* 7a *β*-D-Glc*p*NAc ring is supported by both $^3J_{\mathrm{H,H}}$ coupling measurements and intra-residue *β*-D-Glc*p*NAc ^{1}H-^{1}H distances derived from NOESY experiments.

The magnitude of the $^3J_{H,H}$ couplings for the *Sf* Y *β*-D-Glc*p*NAc ring protons suggest an antiperiplanar arrangement that is consistent with *S. flexneri* 2, 3, and 5; and would suggest $^4C_1$ conformation.[15,90,125] However, there are conspicuously large deviations from these $^3J_{H,H}$ coupling constants in the *Sf* 7a *β*-D-Glc*p*NAc ring protons, suggesting that the ring protons are more axially orientated. Interestingly the largest difference is near the branch point between H3 and H4: *Sf* Y $^3J_{H3,H4}$ = 9.0 Hz and *Sf* 7a $^3J_{H3,H4}$ = 6.2 Hz, which indicate potential distortions between C3 and C5.

This assumption is further supported by NOESY experiments of *Sf* 7a. Intra-residue *β*-D-Glc*p*NAc ^{1}H-^{1}H distance measurements indicate that the H1-Glc*p*NAc – H3-Glc*p*NAc (2.44 Å) and H1-Glc*p*NAc – H5-Glc*p*NAc (2.63 Å) distances are consistent with a $^4C_1$ ring.[297] However, the H3-Glc*p*NAc – H5-Glc*p*NAc (2.95 Å) distance is far longer and, similar to the coupling measurements, also suggests a degree of ring distortion between C3 and C5. Additionally, the NOESY derived ^{1}H-^{1}H distance show that all but one glycosidic linkage predominantly adopts a syn-orientation, which is consistent with the glycosidic linkage orientations of other *S. flexneri* O-Ags.[15,16,22–25] The exception, H1-Glc*p*^{I} — H4-Glc*p*NAc, has a measured distance of 2.72 Å indicating that the linkage is between a syn- and anti-orientation. This unusual orientation is also supported by the inter-residue ^{1}H-^{1}H distances measured.

The experimentally derived distances do not suggest puckering of the *β*-D-GlcNAc ring to the extent observed in the MD simulations of *Sf* 7a. However, better agreement is obtained when considering a smaller boat/skew contribution from the MD simulations. A smaller difference in the ^{1}H-^{1}H distances is achieved when the *β*-D-Glc*p*NAc ring is approximately 85% of the time in $^4C_1$ and the remaining 15% in boat/skew. This is less than half of the observed boat/skew population from the MD simulation and with no $^1C_4$ conformers present. These results suggest that the *β*-D-Glc*p*NAc may pucker into boat/skew conformers but that the MD simulations do not accurately reflect the extent to which this occurs and are biased towards boat/skew and $^1C_4$ conformations.

Chapter 9. Conclusions

We have demonstrated that *S. flexneri* 7a and 7b O-Ags exhibit unusual conformational behaviour caused by puckering of the β-D-Glc*p*NAc residue. A systematic approach of molecular modelling confirms that the results are consistent across MD simulations with different force fields. The conformations predicted by modelling are supported by NMR spectroscopy analyses, which suggests that puckering of β-D-Glc*p*NAc does occur but that it is over represented by the simulation force fields. These unusual conformations have not been previously observed in simulations or experiments of *Shigella* or other bacterial O-Ags.

MD simulations of *Sf* 7a and 7b reveal that the side chain affects backbone conformation and dynamics, restricting its flexibility compared to *Sf* Y, which adopts a range of conformational families. Additionally, α-L-Rha$^{\mathrm{I}}$ 2-O-acetylation in *Sf* 7b causes the O-Ag to adopt more compact conformations compared to *Sf* 7a. These results are consistent with simulations of *Sf* 2, 3, and 5, which also demonstrate that side-chains and O-acetylation significantly affects O-Ag flexibility and conformation.

However, simulations of *Sf* 7a and 7b also show puckering of the β-D-Glc*p*NAc residue. The ring adopts the $^4C_1$ conformation only for short periods and spends the majority of the simulation time in either boat/skew or $^1C_4$ states. All *S. flexneri* O-Ags previously modelled do not pucker to the same extent: the β-D-Glc*p*NAc ring in all repeating units maintain a $^4C_1$ conformer.

This unusual puckering of β-D-Glc*p*NAc is not restricted to *Sf* 7a and 7b but occurs in 3,4-disubstituted β-D-Glc*p* and β-D-Glc*p*NAc trisaccharides as well. Our simulations demonstrate that 2-N-acetylation, anomeric configuration, and force field all play a role in puckering of 3,4-disubstituted β-D-Glc*p*NAc. 2-N-acetylation of 3,4-α-linked β-D-Glc*p* trisaccharides increases both the degree and frequency of β-D-Glc*p* ring puckering. In 3,4-α-linked β-D-Glc*p*NAc trisaccharides, β-D-Glc*p*NAc puckers to a similar extent as in *Sf* 7a and 7b; for both force fields (CHARMM36 and GLYCAM06) the β-D-Glc*p*NAc ring fluctuates between boat/skew and $^1C_4$

conformers. However, there is substantially more frequent puckering in the GLYCAM06 simulations compared to the CHARMM36 simulations. In contrast, puckering is significantly reduced in 3,4-β-linked trisaccharides compared to 3,4-α-linked trisaccharides. In the case of 3-β,4-α-linked and 3-α,4-β-linked β-D-Glc*p*NAc, ring puckering occurs for even shorter periods, underscoring the strong correlation between anomeric configuration and ring pucker.

PMF calculations demonstrate that 2-N-acetylation of 3,4-disubstituted β-D-Glc*p* generally decreases the energy required for the ring to convert between different pucker states. Similarly, lower energies are required for ring interconversion in 3,4-α-linked and 3,4-β-linked β-D-Glc*p*NAc compared to 3-β,4-α-linked and 3-α,4-β-linked β-D-Glc*p*NAc trisaccharides. Observations from our MD simulation and PMF calculations are novel but are also supported by early molecular modelling studies of 3,4-disubstituted β-D-Glc*p* trisaccharides.

The modelling predictions of *Sf* 7a and 7b were corroborated by NMR spectroscopy experiments. ^{1}H NMR temperature studies indicate dynamic behaviour around the branch point of β-D-Glc*p*NAc for *Sf* 7a when compared to *Sf* Y. Additional $^3J_{H,H}$ measurements of *Sf* 7a reveal much smaller coupling between the β-D-Glc*p*NAc ring protons than in *Sf* Y. The coupling constants for *Sf* 7a are also significantly different to those in monosaccharide β-D-Glc*p*NAc but align with those measured from α-L-Rha*p*-(1→3)-[α-D-Glc*p*-(1→4)]-β-D-Glc*p*NAc-OPr, which is a similarly branched 3,4-α-linked β-D-Glc*p*NAc trisaccharide. These results partially support the presence of boat/skew conformations.

NOESY derived ^{1}H-^{1}H distances of *Sf* 7a mostly match those calculated from the MD simulations. The combined experimental and simulation measurements indicate that β-D-Glc*p*NAc does pucker, with the ring is approximately 85% of the time in $^4C_1$ and the remaining 15% in boat/skew — likely corresponding to $^3S_1$, $B_{1,4}$, and $^5S_1$ states. While we could only measure a few ^{1}H-^{1}H distances for *Sf* 7b, due to peak overlap, the results do not differ significantly to those from *Sf* 7a.

With the exception of idopyranose (which prefers $^1C_4$), all pyranose sugars adopt the predominantly low-energy $^4C_1$ conformation as bulky substituents generally favour an equatorial orientation. The β-D-Glc*p*NAc monosaccharide puckers into non-$^4C_1$ states for short periods of time and increased substitution typically decreases the rate of puckering. β-D-Glc*p*NAc residues in di-, oligo-, and polysaccharides favour a $^4C_1$ conformation with little evidence of boat/skew or $^1C_4$ states. Our results suggest that 3,4-disubstitution of β-D-Glc*p*NAc with bulky groups contributes to the unusual ring pucker and conformational behaviour of *Sf* 7a, 7b, and the trisaccharides studied.

We demonstrate that the α-D-Glcp^{II}-(1→2)-α-D-Glcp^{I} side-chain in the O-Ag repeating unit of *Sf* 7a and 7b significantly impacts the backbone conformation. We also identified unusual puckering of the β-D-Glc*p*NAc ring in both *Sf* 7a and 7b during molecular simulations. ^{1}H NMR experiments support these results but show that β-D-Glc*p*NAc puckers to a lesser degree than is predicted by the simulations. These observations are important as β-D-Glc*p*NAc is central component of the *Sf* 7a and 7b O-factor — a key antigenic structure responsible for eliciting a protective immune response. These findings contribute to the understanding of *S. flexneri* O-Ag conformation and provide valuable insights for the development of polysaccharide-based vaccines against *Shigella*.

There is substantial scope for future work based on this study. The *Sf* 7a and 7b conformations can be resolved with greater accuracy. However, this will likely require force fields that more appropriately model the 3,4-disubstituted β-D-Glc*p*NAc ring pucker, which may necessitate updates to force field parameters as they currently favour boat/skew and $^1C_4$ conformations. Modelling of *S. flexneri* serotype 1 O-Ags, which have a similar structure to serotype 7 O-Ags, may also determine if β-D-Glc*p*NAc puckers to the same extent. The energies, conformations, dynamics of a wide range of N-acetylated 3,4-disubstituted pyranose sugars can also be calculated to determine if they exhibit similar conformational behaviour to 3,4-disubstituted β-D-Glc*p*NAc trisaccharides.

References

[1] I. Roberts, *Microbiology*, 1995, **141**, 2023–2031.

[2] I. Lerouge and J. Vanderleyden, *FEMS Microbiology Reviews*, 2002, **26**, 17–47.

[3] S. L. Flitsch and R. V. Ulijn, *Nature*, 2003, **421**, 219–220.

[4] B. Kuberan and R. Linhardt, *Current Organic Chemistry*, 2000, **4**, 653–677.

[5] R. D. Astronomo and D. R. Burton, *Nature Reviews Drug Discovery*, 2010, **9**, 308–324.

[6] S. F. Slovin, S. J. Keding and G. Ragupathi, *Immunology and Cell Biology*, 2005, **83**, 418–428.

[7] Y. A. Knirel and M.-R. V. Calsteren, in *Bacterial Exopolysaccharides*, ed. J. J. Barchi, Elsevier, Oxford, 2nd edn, 2021, pp. 21–95.

[8] M. M. Kuttel and N. Ravenscroft, in *The Role of Molecular Modeling in Predicting Carbohydrate Antigen Conformation and Understanding Vaccine Immunogenicity*, ed. A. Prasad, American Chemical Society, 2018, pp. 139–173.

[9] C. Erridge, E. Bennett-Guerrero and I. R. Poxton, *Microbes and Infection*, 2002, **4**, 837–851.

[10] R. J. Woods, *Chemical Reviews*, 2018, **118**, 8005–8024.

[11] M. Gomez-Redondo, A. Ardá, A. Gimeno and J. Jiménez-Barbero, *Drug Discovery Today: Technologies*, 2020, **35-36**, 1–11.

[12] J. O. Duus, C. H. Gotfredsen and K. Bock, *Chemical Reviews*, 2000, **100**, 4589–4614.

[13] G. Widmalm, *Carbohydrate Research*, 2013, **378**, 123–132.

[14] S. Perez, in *X-Ray Diffraction and Crystallography of Oligosaccharides and Polysaccharides*, Springer Berlin Heidelberg, 2013, pp. 2767–2777.

[15] F. X. Theillet, C. Simenel, C. Guerreiro, A. Phalipon, L. A. Mulard and M. Delepierre, *Glycobiology*, 2011, **21**, 109–121.

[16] Y. Kang, S. Barbirz, R. Lipowsky and M. Santer, *Journal of Physical Chemistry B*, 2014, **118**, 2523–2534.

[17] M. M. Kuttel, G. E. Jackson, M. Mafata and N. Ravenscroft, *Carbohydrate Research*, 2015, **406**, 27–33.

[18] T. Galochkina, D. Zlenko, A. Nesterenko, I. Kovalenko, M. Strakhovskaya, A. Averyanov and A. Rubin, *ChemPhysChem*, 2016, 2839–2853.

[19] M. M. Kuttel, Z. Timol and N. Ravenscroft, *Carbohydrate Research*, 2017, **446-447**, 40–47.

[20] J. Hlozek, M. M. Kuttel and N. Ravenscroft, *Carbohydrate Research*, 2018, **465**, 44–51.

[21] J. Hlozek, N. Ravenscroft and M. M. Kuttel, *Carbohydrate Research*, 2019, **486**, 107838.

[22] J. Hlozek, N. Ravenscroft and M. M. Kuttel, *The Journal of Physical Chemistry B*, 2020, **124**, 2806–2814.

[23] J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, *Vaccines*, 2020, **8**, 1–17.

[24] N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, *Vaccines*, 2021, **9**, 1–5.

[25] J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, *Vaccines*, 2020, **8**, 643.

[26] M. M. Kuttel, A. Casadevall and S. Oscarson, *Molecules*, 2020, **25**, 2651–2655.

[27] N. I. Richardson, M. M. Kuttel and N. Ravenscroft, *Frontiers in Molecular Biosciences*, 2022, **9**, 1–5.

[28] D. Goldblatt, *Clinical and Experimental Immunology*, 2002, **119**, 1–3.

[29] G. Ada and D. Isaacs, *Clinical Microbiology and Infection*, 2003, **9**, 79–85.

[30] M. E. Pichichero, *Human vaccines and immunotherapeutics*, 2013, **9**, 2505–2523.

[31] S. Lang and X. Huang, *Frontiers in Chemistry*, 2020, **8**, 1.

[32] K. L. Kotloff, J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu, S. O. Sow, D. Sur, R. F. Breiman, A. S. Faruque, A. K. Zaidi, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B. Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F. Quadri, R. A. Adegbola, M. Antonio, M. J. Hossain, A. Akinsola, I. Mandomando, T. Nhampossa, S. Acácio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K. Muhsen, H. Sommerfelt, R. M. Robins-Browne and M. M. Levine, *The Lancet*, 2013, **382**, 209–222.

[33] M. Naghavi, A. A. Abajobir, C. Abbafati, K. M. Abbas, F. Abd-Allah, S. F. Abera, V. Aboyans, O. Adetokunboh, A. Afshin, A. Agrawal, A. Ahmadi, M. B. Ahmed, A. N. Aichour, M. T. E. Aichour, I. Aichour, S. Aiyar, F. Alahdab, Z. Al-Aly, K. Alam, N. Alam, T. Alam, K. A. Alene, A. Al-Eyadhy, S. D. Ali, R. Alizadeh-Navaei, J. M. Alkaabi, A. Alkerwi, F. Alla, P. Allebeck, C. Allen, R. Al-Raddadi, U. Alsharif, K. A. Altirkawi, N. Alvis-Guzman, A. T. Amare, E. Amini, W. Ammar, Y. A. Amoako, N. Anber, H. H. Andersen, C. L. Andrei, S. Androudi, H. Ansari, C. A. T. Antonio, P. Anwari, J. Ärnlöv, M. Arora, A. Artaman, K. K. Aryal, H. Asayesh, S. W. Asgedom, T. M. Atey, L. Avila-Burgos, E. F. G. Avokpaho, A. Awasthi, T. K. Babalola, U. Bacha, K. Balakrishnan, A. Barac, M. A. Barboza, S. L. Barker-Collo, S. Barquera, L. Barregard, L. H. Barrero, B. T. Baune, N. Bedi, E. Beghi, Y. Béjot, B. B. Bekele, M. L. Bell, J. R. Bennett, I. M. Bensenor, A. Berhane, E. Bernabé, B. D. Betsu, M. Beuran, S. Bhatt, S. Biadgilign, K. Bienhoff, B. Bikbov, D. Bisanzio, R. R. A. Bourne, N. J. K. Breitborde, L. N. B. Bulto, B. R. Bumgarner, Z. A. Butt, L. Cahuana-Hurtado, E. Cameron, J. C. Campuzano, J. Car, R. Cárdenas, J. J. Carrero, A. Carter, D. C. Casey, C. A. Castañeda-Orjuela, F. Catalá-López, F. J. Charlson, C. E. Chibueze, O. Chimed-Ochir, V. H. Chisumpa, A. A. Chitheer, D. J. Christopher, L. G. Ciobanu, M. Cirillo, A. J. Cohen, D. Colombara, C. Cooper, B. C. Cowie, M. H. Criqui, L. Dandona, R. Dandona, P. I. Dargan, J. das Neves, D. V. Davitoiu, K. Davletov, B. de Courten, B. K. Defo, L. Degenhardt, S. Deiparine, K. Deribe, A. Deribew, S. Dey, D. Dicker, E. L. Ding, S. Djalalinia, H. P. Do, D. T. Doku, D. Douwes-Schultz, T. R. Driscoll, M. Dubey, B. B. Duncan, M. Echko, Z. Z. El-Khatib, C. L. Ellingsen, A. Enayati, S. P. Ermakov, H. E. Erskine, S. Eskandarieh, A. Esteghamati, K. Estep, C. S. e Sa Farinha, A. Faro, F. Farzadfar, V. L. Feigin, S.-M. Fereshtehnejad, J. C. Fernandes, A. J. Ferrari, T. R. Feyissa, I. Filip, S. Finegold, F. Fischer, C. Fitzmaurice, A. D. Flaxman, N. Foigt, T. Frank, M. Fraser, N. Fullman, T. Fürst, J. M. Furtado, E. Gakidou, A. L. Garcia-Basteiro, T. Gebre, G. B. Gebregergs, T. T. Gebrehiwot, D. Y. Gebremichael, J. M. Geleijnse, R. Genova-Maleras, H. A. Gesesew, P. W. Gething, R. F. Gillum, A. Z. Giref, M. Giroud, G. Giussani, W. W. Godwin, A. L. Gold, E. M. Goldberg, P. N. Gona, S. V. Gopalani, H. N. Gouda, A. C. Goulart, M. Griswold, R. Gupta, T. Gupta, V. Gupta, P. C. Gupta, J. A. Haagsma, N. Hafezi-Nejad, A. D. Hailu, G. B. Hailu, R. R. Hamadeh, M. T. Hambisa, S. Hamidi, M. Hammami, J. Hancock, A. J. Handal, G. J. Hankey, Y. Hao, H. L. Harb, H. A. Hareri, M. S. Hassanvand, R. Havmoeller, S. I. Hay, F. He, M. T. Hedayati, N. J. Henry, I. B. Heredia-Pi, C. Herteliu, H. W. Hoek, M. Horino, N. Horita, H. D. Hosgood, S. Hostiuc, P. J. Hotez, D. G. Hoy, C. Huynh, K. M. Iburg, C. Ikeda, B. V. Ileanu, A. A. Irenso, C. M. S. Irvine, S. M. S. Islam, K. H. Jacobsen, N. Jahanmehr, M. B. Jakovljevic, M. Javanbakht, S. P. Jayaraman, P. Jeemon, V. Jha, D. John, C. O. Johnson, S. C.

Johnson, J. B. Jonas, M. Jürisson, Z. Kabir, R. Kadel, A. Kahsay, R. Kamal, A. Karch, S. M. Karimi, C. Karimkhani, A. Kasaeian, N. A. Kassaw, N. J. Kassebaum, S. V. Katikireddi, N. Kawakami, P. N. Keiyoro, L. Kemmer, C. N. Kesavachandran, Y. S. Khader, E. A. Khan, Y.-H. Khang, A. T. A. Khoja, M. H. Khosravi, A. Khosravi, J. Khubchandani, A. A. Kiadaliri, C. Kieling, D. Kievlan, Y. J. Kim, D. Kim, R. W. Kimokoti, Y. Kinfu, N. Kissoon, M. Kivimaki, A. K. Knudsen, J. A. Kopec, S. Kosen, P. A. Koul, A. Koyanagi, X. R. Kulikoff, G. A. Kumar, P. Kumar, M. Kutz, H. H. Kyu, D. K. Lal, R. Lalloo, T. L. N. Lambert, Q. Lan, V. C. Lansingh, A. Larsson, P. H. Lee, J. Leigh, J. Leung, M. Levi, Y. Li, D. L. Kappe, X. Liang, M. L. Liben, S. S. Lim, P. Y. Liu, A. Liu, Y. Liu, R. Lodha, G. Logroscino, S. Lorkowski, P. A. Lotufo, R. Lozano, T. C. D. Lucas, S. Ma, E. R. K. Macarayan, E. R. Maddison, M. M. A. E. Razek, M. Majdan, R. Majdzadeh, A. Majeed, R. Malekzadeh, R. Malhotra, D. C. Malta, H. Manguerra, T. Manyazewal, C. C. Mapoma, L. B. Marczak, D. Markos, J. Martinez-Raga, F. R. Martins-Melo, I. Martopullo, C. McAlinden, M. McGaughey, J. J. McGrath, S. Mehata, T. Meier, K. G. Meles, P. Memiah, Z. A. Memish, M. M. Mengesha, D. T. Mengistu, B. G. Menota, G. A. Mensah, T. J. Meretoja, A. Meretoja, A. Millear, T. R. Miller, S. Minnig, M. Mirarefin, E. M. Mirrakhimov, A. Misganaw, S. R. Mishra, I. A. Mohamed, K. A. Mohammad, A. Mohammadi, S. Mohammed, A. H. Mokdad, G. L. D. Mola, S. K. Mollenkopf, M. Molokhia, L. Monasta, J. C. Montañez, M. Montico, M. D. Mooney, M. Moradi-Lakeh, P. Moraga, L. Morawska, C. Morozoff, S. D. Morrison, C. Mountjoy-Venning, K. B. Mruts, K. Muller, G. V. S. Murthy, K. I. Musa, J. B. Nachega, A. Naheed, L. Naldi, V. Nangia, B. R. Nascimento, J. T. Nasher, G. Natarajan, I. Negoi, J. W. Ngunjiri, C. T. Nguyen, Q. L. Nguyen, T. H. Nguyen, G. Nguyen, M. Nguyen, E. Nichols, D. N. A. Ningrum, V. M. Nong, J. J. N. Noubiap, F. A. Ogbo, I.-H. Oh, A. Okoro, A. T. Olagunju, H. E. Olsen, B. O. Olusanya, J. O. Olusanya, K. Ong, J. N. Opio, E. Oren, A. Ortiz, M. Osman, E. Ota, M. PA, R. E. Pacella, S. Pakhale, A. Pana, B. K. Panda, S. Panda-Jonas, C. Papachristou, E.-K. Park, S. B. Patten, G. C. Patton, D. Paudel, K. Paulson, D. M. Pereira, F. Perez-Ruiz, N. Perico, A. Pervaiz, M. Petzold, M. R. Phillips, D. M. Pigott, C. Pinho, D. Plass, M. A. Pletcher, S. Polinder, M. J. Postma, F. Pourmalek, C. Purcell, M. Qorbani, B. P. A. Quintanilla, A. Radfar, A. Rafay, V. Rahimi-Movaghar, M. H. U. Rahman, M. Rahman, R. K. Rai, C. L. Ranabhat, Z. Rankin, P. C. Rao, G. K. Rath, S. Rawaf, S. E. Ray, J. Rehm, R. C. Reiner, M. B. Reitsma, G. Remuzzi, S. Rezaei, M. S. Rezai, M. B. Rokni, L. Ronfani, G. Roshandel, G. A. Roth, D. Rothenbacher, G. M. Ruhago, R. SA, S. Saadat, P. S. Sachdev, N. Sadat, M. Safdarian, S. Safi, S. Safiri, R. Sagar, R. Sahathevan, J. Salama, P. Salamati, J. A. Salomon, A. M. Samy, J. R. Sanabria, M. D. Sanchez-Niño, D. Santomauro, I. S. Santos, M. M. S. Milicevic, B. Sartorius, M. Satpathy, M. I. Schmidt, I. J. C. Schneider, S. Schulhofer-Wohl, A. E. Schutte, D. C. Schwebel,

F. Schwendicke, S. G. Sepanlou, E. E. Servan-Mori, K. A. Shackelford, S. Shahraz, M. A. Shaikh, M. Shamsipour, M. Shamsizadeh, J. Sharma, R. Sharma, J. She, S. Sheikhbahaei, M. Shey, P. Shi, C. Shields, M. Shigematsu, R. Shiri, S. Shirude, I. Shiue, H. Shoman, M. G. Shrime, I. D. Sigfusdottir, N. Silpakit, J. P. Silva, J. A. Singh, A. Singh, E. Skiadaresi, A. Sligar, D. L. Smith, A. Smith, M. Smith, B. H. A. Sobaih, S. Soneji, R. J. D. Sorensen, J. B. Soriano, C. T. Sreeramareddy, V. Srinivasan, J. D. Stanaway, V. Stathopoulou, N. Steel, D. J. Stein, C. Steiner, S. Steinke, M. A. Stokes, M. Strong, B. Strub, M. Subart, M. B. Sufiyan, B. F. Sunguya, P. J. Sur, S. Swaminathan, B. L. Sykes, R. Tabarés-Seisdedos, S. K. Tadakamadla, K. Takahashi, J. S. Takala, R. T. Talongwa, M. R. Tarawneh, M. Tavakkoli, N. Taveira, T. K. Tegegne, A. Tehrani-Banihashemi, M.-H. Temsah, A. S. Terkawi, J. S. Thakur, O. Thamsuwan, K. R. Thankappan, K. E. Thomas, A. H. Thompson, A. J. Thomson, A. G. Thrift, R. Tobe-Gai, R. Topor-Madry, A. Torre, M. Tortajada, J. A. Towbin, B. X. Tran, C. Troeger, T. Truelsen, D. Tsoi, E. M. Tuzcu, S. Tyrovolas, K. N. Ukwaja, E. A. Undurraga, R. Updike, O. A. Uthman, B. S. C. Uzochukwu, J. F. M. van Boven, T. Vasankari, N. Venketasubramanian, F. S. Violante, V. V. Vlassov, S. E. Vollset, T. Vos, T. Wakayo, M. T. Wallin, Y.-P. Wang, E. Weiderpass, R. G. Weintraub, D. J. Weiss, A. Werdecker, R. Westerman, B. Whetter, H. A. Whiteford, T. Wijeratne, C. S. Wiysonge, B. G. Woldeyes, C. D. A. Wolfe, R. Woodbrook, A. Workicho, D. Xavier, Q. Xiao, G. Xu, M. Yaghoubi, B. Yakob, Y. Yano, M. Yaseri, H. H. Yimam, N. Yonemoto, S.-J. Yoon, M. Yotebieng, M. Z. Younis, Z. Zaidi, M. E. S. Zaki, E. A. Zegeye, Z. M. Zenebe, T. A. Zerfu, A. L. Zhang, X. Zhang, B. Zipkin, S. Zodpey, A. D. Lopez and C. J. L. Murray, *The Lancet*, 2017, **390**, 1151–1210.

[34] K. A. Lampel, S. B. Formal† and A. T. Maurelli, *EcoSal Plus*, 2018, **8**, 1–5.

[35] M. L. Bennish and S. Ahmed, in *Hunter's Tropical Medicine and Emerging Infectious Diseases*, Elsevier, 2020, pp. 492–499.

[36] H. Wang, A. A. Abajobir, K. H. Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, S. F. Abera, H. N. Abraha, L. J. Abu-Raddad, N. M. E. Abu-Rmeileh, I. A. Adedeji, R. A. Adedoyin, I. M. O. Adetifa, O. Adetokunboh, A. Afshin, R. Aggarwal, A. Agrawal, S. Agrawal, A. A. Kiadaliri, M. B. Ahmed, M. T. E. Aichour, A. N. Aichour, I. Aichour, S. Aiyar, A. S. Akanda, T. F. Akinyemiju, N. Akseer, F. H. A. Lami, S. Alabed, F. Alahdab, Z. Al-Aly, K. Alam, N. Alam, D. Alasfoor, R. W. Aldridge, K. A. Alene, A. Al-Eyadhy, S. Alhabib, R. Ali, R. Alizadeh-Navaei, S. M. Aljunid, J. M. Alkaabi, A. Alkerwi, F. Alla, S. D. Allam, P. Allebeck, R. Al-Raddadi, U. Alsharif, K. A. Altirkawi, N. Alvis-Guzman, A. T. Amare, E. A. Ameh, E. Amini, W. Ammar, Y. A. Amoako, N. Anber, C. L. Andrei, S. Androudi, H. Ansari, M. G. Ansha, C. A. T. Antonio, P. Anwari, J. Ärnlöv, M. Arora, A. Artaman, K. K. Aryal, H. Asayesh, S. W. Asgedom, R. J. Asghar, R. Assadi, A. M. Assaye,

T. M. Atey, S. R. Atre, L. Avila-Burgos, E. F. G. A. Avokpaho, A. Awasthi, T. K. Babalola, U. Bacha, A. Badawi, K. Balakrishnan, S. Balalla, A. Barac, R. M. Barber, M. A. Barboza, S. L. Barker-Collo, T. Bärnighausen, S. Barquera, L. Barregard, L. H. Barrero, B. T. Baune, S. Bazargan-Hejazi, N. Bedi, E. Beghi, Y. Béjot, B. B. Bekele, M. L. Bell, A. K. Bello, D. A. Bennett, J. R. Bennett, I. M. Bensenor, J. Benson, A. Berhane, D. F. Berhe, E. Bernabé, M. Beuran, A. S. Beyene, N. Bhala, A. Bhansali, S. Bhaumik, Z. A. Bhutta, B. K. Bicer, H. H. Bidgoli, B. Bikbov, C. Birungi, S. Biryukov, D. Bisanzio, H. M. Bizuayehu, P. Bjerregaard, C. D. Blosser, D. J. Boneya, S. Boufous, R. R. A. Bourne, A. Brazinova, N. J. K. Breitborde, H. Brenner, T. S. Brugha, G. Bukhman, L. N. B. Bulto, B. R. Bumgarner, M. Burch, Z. A. Butt, L. E. Cahill, L. Cahuana-Hurtado, I. R. Campos-Nonato, J. Car, M. Car, R. Cárdenas, D. O. Carpenter, J. J. Carrero, A. Carter, C. A. Castañeda-Orjuela, F. F. Castro, R. E. Castro, F. Catalá-López, H. Chen, P. P.-C. Chiang, M. Chibalabala, V. H. Chisumpa, A. A. Chitheer, J.-Y. J. Choi, H. Christensen, D. J. Christopher, L. G. Ciobanu, M. Cirillo, A. J. Cohen, S. M. Colquhoun, J. Coresh, M. H. Criqui, E. A. Cromwell, J. A. Crump, L. Dandona, R. Dandona, P. I. Dargan, J. das Neves, G. Davey, D. V. Davitoiu, K. Davletov, B. de Courten, D. D. Leo, L. Degenhardt, S. Deiparine, R. P. Dellavalle, K. Deribe, A. Deribew, D. C. D. Jarlais, S. Dey, S. D. Dharmaratne, M. K. Dherani, C. Diaz-Torné, E. L. Ding, P. Dixit, S. Djalalinia, H. P. Do, D. T. Doku, C. A. Donnelly, K. P. B. dos Santos, D. Douwes-Schultz, T. R. Driscoll, L. Duan, M. Dubey, B. B. Duncan, L. K. Dwivedi, H. Ebrahimi, C. E. Bcheraoui, C. L. Ellingsen, A. Enayati, A. Y. Endries, S. P. Ermakov, S. Eshetie, B. Eshrati, S. Eskandarieh, A. Esteghamati, K. Estep, F. B. B. Fanuel, A. Faro, M. S. Farvid, F. Farzadfar, V. L. Feigin, S.-M. Fereshtehnejad, J. G. Fernandes, J. C. Fernandes, T. R. Feyissa, I. Filip, F. Fischer, N. Foigt, K. J. Foreman, T. Frank, R. C. Franklin, M. Fraser, J. Friedman, J. J. Frostad, N. Fullman, T. Fürst, J. M. Furtado, N. D. Futran, E. Gakidou, K. Gambashidze, A. Gamkrelidze, F. G. Gankpé, A. L. Garcia-Basteiro, G. B. Gebregergs, T. T. Gebrehiwot, K. G. Gebrekidan, M. W. Gebremichael, A. A. Gelaye, J. M. Geleijnse, B. L. Gemechu, K. S. Gemechu, R. Genova-Maleras, H. A. Gesesew, P. W. Gething, K. B. Gibney, P. S. Gill, R. F. Gillum, A. Z. Giref, B. W. Girma, G. Giussani, S. Goenka, B. Gomez, P. N. Gona, S. V. Gopalani, A. C. Goulart, N. Graetz, H. C. Gugnani, P. C. Gupta, R. Gupta, R. Gupta, T. Gupta, V. Gupta, J. A. Haagsma, N. Hafezi-Nejad, A. Hakuzimana, Y. A. Halasa, R. R. Hamadeh, M. T. Hambisa, S. Hamidi, M. Hammami, J. Hancock, A. J. Handal, G. J. Hankey, Y. Hao, H. L. Harb, H. A. Hareri, S. Harikrishnan, J. M. Haro, M. S. Hassanvand, R. Havmoeller, R. J. Hay, S. I. Hay, F. He, I. B. Heredia-Pi, C. Herteliu, E. H. Hilawe, H. W. Hoek, N. Horita, H. D. Hosgood, S. Hostiuc, P. J. Hotez, D. G. Hoy, M. Hsairi, A. S. Htet, G. Hu, J. J. Huang, H. Huang, K. M. Iburg, E. U. Igumbor, B. V. Ileanu, M. Inoue, A. A. Irenso, C. M. S. Irvine,

S. M. S. Islam, N. Islam, K. H. Jacobsen, T. Jaenisch, N. Jahanmehr, M. B. Jakovljevic, M. Javanbakht, A. U. Jayatilleke, P. Jeemon, P. N. Jensen, V. Jha, Y. Jin, D. John, O. John, S. C. Johnson, J. B. Jonas, M. Jürisson, Z. Kabir, R. Kadel, A. Kahsay, Y. Kalkonde, R. Kamal, H. Kan, A. Karch, C. K. Karema, S. M. Karimi, G. Karthikeyan, A. Kasaeian, N. A. Kassaw, N. J. Kassebaum, A. Kastor, S. V. Katikireddi, A. Kaul, N. Kawakami, K. Kazanjan, P. N. Keiyoro, S. G. Kelbore, A. H. Kemp, A. P. Kengne, A. Keren, M. Kereselidze, C. N. Kesavachandran, E. B. Ketema, Y. S. Khader, I. A. Khalil, E. A. Khan, G. Khan, Y.-H. Khang, S. Khera, A. T. A. Khoja, M. H. Khosravi, G. D. Kibret, C. Kieling, Y. J. Kim, C. il Kim, D. Kim, P. Kim, S. Kim, R. W. Kimokoti, Y. Kinfu, S. Kishawi, N. Kissoon, M. Kivimaki, A. K. Knudsen, Y. Kokubo, J. A. Kopec, S. Kosen, P. A. Koul, A. Koyanagi, M. Kravchenko, K. J. Krohn, B. K. Defo, E. J. Kuipers, X. R. Kulikoff, V. S. Kulkarni, G. A. Kumar, P. Kumar, F. A. Kumsa, M. Kutz, C. Lachat, A. K. Lagat, A. C. J. Lager, D. K. Lal, R. Lalloo, N. Lambert, Q. Lan, V. C. Lansingh, H. J. Larson, A. Larsson, D. O. Laryea, P. M. Lavados, A. Laxmaiah, P. H. Lee, J. Leigh, J. Leung, R. Leung, M. Levi, Y. Li, Y. Liao, M. L. Liben, S. S. Lim, S. Linn, S. E. Lipshultz, S. Liu, R. Lodha, G. Logroscino, S. A. Lorch, S. Lorkowski, P. A. Lotufo, R. Lozano, R. Lunevicius, R. A. Lyons, S. Ma, E. R. Macarayan, I. E. Machado, M. T. Mackay, M. M. A. E. Razek, C. Magis-Rodriguez, M. Mahdavi, M. Majdan, R. Majdzadeh, A. Majeed, R. Malekzadeh, R. Malhotra, D. C. Malta, L. G. Mantovani, T. Manyazewal, C. C. Mapoma, L. B. Marczak, G. B. Marks, E. A. Martin, J. Martinez-Raga, F. R. Martins-Melo, J. Massano, P. K. Maulik, B. M. Mayosi, M. Mazidi, C. McAlinden, S. T. McGarvey, J. J. McGrath, M. McKee, S. Mehata, M. M. Mehndiratta, K. M. Mehta, T. Meier, T. C. Mekonnen, K. G. Meles, P. Memiah, Z. A. Memish, W. Mendoza, M. M. Mengesha, M. A. Mengistie, D. T. Mengistu, G. R. Menon, B. G. Menota, G. A. Mensah, T. J. Meretoja, A. Meretoja, H. B. Mezgebe, R. Micha, J. Mikesell, T. R. Miller, E. J. Mills, S. Minnig, M. Mirarefin, E. M. Mirrakhimov, A. Misganaw, S. R. Mishra, K. A. Mohammad, A. Mohammadi, K. E. Mohammed, S. Mohammed, M. B. V. Mohan, S. K. Mohanty, A. H. Mokdad, S. K. Mollenkopf, M. Molokhia, L. Monasta, J. C. M. Hernandez, M. Montico, M. D. Mooney, A. R. Moore, M. Moradi-Lakeh, P. Moraga, L. Morawska, R. Mori, S. D. Morrison, K. B. Mruts, U. O. Mueller, E. Mullany, K. Muller, G. V. S. Murthy, S. Murthy, K. I. Musa, J. B. Nachega, C. Nagata, G. Nagel, M. Naghavi, K. S. Naidoo, L. Nanda, V. Nangia, B. R. Nascimento, G. Natarajan, I. Negoi, C. T. Nguyen, Q. L. Nguyen, T. H. Nguyen, G. Nguyen, D. N. A. Ningrum, M. I. Nisar, M. Nomura, V. M. Nong, O. F. Norheim, B. Norrving, J. J. N. Noubiap, L. Nyakarahuka, M. J. O'Donnell, C. M. Obermeyer, F. A. Ogbo, I.-H. Oh, A. Okoro, O. Oladimeji, A. T. Olagunju, B. O. Olusanya, J. O. Olusanya, E. Oren, A. Ortiz, A. Osgood-Zimmerman, E. Ota, M. O. Owolabi, A. S. Oyekale, M. PA, R. E. Pacella, S. Pakhale, A. Pana, B. K.

Panda, S. Panda-Jonas, E.-K. Park, M. Parsaeian, T. Patel, S. B. Patten, G. C. Patton, D. Paudel, D. M. Pereira, R. Perez-Padilla, F. Perez-Ruiz, N. Perico, A. Pervaiz, K. Pesudovs, C. B. Peterson, W. A. Petri, M. Petzold, M. R. Phillips, F. B. Piel, D. M. Pigott, F. Pishgar, D. Plass, S. Polinder, S. Popova, M. J. Postma, R. G. Poulton, F. Pourmalek, N. Prasad, M. Purwar, M. Qorbani, B. P. A. Quintanilla, R. H. S. Rabiee, A. Radfar, A. Rafay, A. Rahimi-Movaghar, V. Rahimi-Movaghar, M. H. U. Rahman, S. U. Rahman, M. Rahman, R. K. Rai, S. Rajsic, U. Ram, S. M. Rana, C. L. Ranabhat, P. V. Rao, S. Rawaf, S. E. Ray, M. A. S. Rego, J. Rehm, R. C. Reiner, G. Remuzzi, A. M. N. Renzaho, S. Resnikoff, S. Rezaei, M. S. Rezai, A. L. Ribeiro, J. C. Rivas, M. B. Rokni, L. Ronfani, G. Roshandel, G. A. Roth, D. Rothenbacher, A. Roy, E. Rubagotti, G. M. Ruhago, S. Saadat, Y. D. Sabde, P. S. Sachdev, N. Sadat, M. Safdarian, S. Safi, S. Safiri, R. Sagar, R. Sahathevan, A. Sahebkar, M. A. Sahraian, J. Salama, P. Salamati, J. A. Salomon, S. S. Salvi, A. M. Samy, J. R. Sanabria, M. D. Sanchez-Niño, I. S. Santos, M. M. S. Milicevic, R. Sarmiento-Suarez, B. Sartorius, M. Satpathy, M. Sawhney, S. Saxena, M. I. Saylan, M. I. Schmidt, I. J. C. Schneider, S. Schulhofer-Wohl, A. E. Schutte, D. C. Schwebel, F. Schwendicke, S. Seedat, A. M. Seid, S. G. Sepanlou, E. E. Servan-Mori, K. A. Shackelford, A. Shaheen, S. Shahraz, M. A. Shaikh, M. Shamsipour, M. Shamsizadeh, J. Sharma, R. Sharma, J. She, J. Shen, B. P. Shetty, P. Shi, K. Shibuya, G. T. Shifa, M. Shigematsu, R. Shiri, I. Shiue, M. G. Shrime, I. D. Sigfusdottir, D. H. Silberberg, N. Silpakit, D. A. S. Silva, J. P. Silva, D. G. A. Silveira, S. Sindi, J. A. Singh, P. K. Singh, A. Singh, V. Singh, D. N. Sinha, K. A. Kissimova-Skarbek, E. Skiadaresi, A. Sligar, D. L. Smith, B. H. A. Sobaih, E. Sobngwi, S. Soneji, J. B. Soriano, C. T. Sreeramareddy, V. Srinivasan, V. Stathopoulou, N. Steel, D. J. Stein, C. Steiner, H. Stöckl, M. A. Stokes, M. Strong, M. B. Sufiyan, R. A. Suliankatchi, B. F. Sunguya, P. J. Sur, S. Swaminathan, B. L. Sykes, C. E. I. Szoeke, R. Tabarés-Seisdedos, S. K. Tadakamadla, F. Tadese, N. Tandon, D. Tanne, M. Tarajia, M. Tavakkoli, N. Taveira, A. Tehrani-Banihashemi, T. Tekelab, D. Y. Tekle, M.-H. Temsah, A. S. Terkawi, C. L. Tesema, B. Tesssema, A. Theis, N. Thomas, A. H. Thompson, A. J. Thomson, A. G. Thrift, T. Y. Tiruye, R. Tobe-Gai, M. Tonelli, R. Topor-Madry, F. Topouzis, M. Tortajada, B. X. Tran, T. Truelsen, U. Trujillo, N. Tsilimparis, K. B. Tuem, E. M. Tuzcu, S. Tyrovolas, K. N. Ukwaja, E. A. Undurraga, O. A. Uthman, B. S. C. Uzochukwu, J. F. M. van Boven, Y. Y. Varakin, S. Varughese, T. Vasankari, A. M. N. Vasconcelos, I. M. Velasquez, N. Venketasubramanian, R. Vidavalur, F. S. Violante, A. Vishnu, S. K. Vladimirov, V. V. Vlassov, S. E. Vollset, T. Vos, J. L. Waid, T. Wakayo, Y.-P. Wang, S. Weichenthal, E. Weiderpass, R. G. Weintraub, A. Werdecker, J. Wesana, T. Wijeratne, J. D. Wilkinson, C. S. Wiysonge, B. G. Woldeyes, C. D. A. Wolfe, A. Workicho, S. B. Workie, D. Xavier, G. Xu, M. Yaghoubi, B. Yakob, A. Z. Yalew, L. L. Yan, Y. Yano, M. Yaseri, P. Ye, H. H. Yimam, P. Yip, B. D.

Yirsaw, N. Yonemoto, S.-J. Yoon, M. Yotebieng, M. Z. Younis, Z. Zaidi, M. E. S. Zaki, H. Zeeb, Z. M. Zenebe, T. A. Zerfu, A. L. Zhang, X. Zhang, S. Zodpey, L. J. Zuhlke, A. D. Lopez and C. J. L. Murray, *The Lancet*, 2017, **390**, 1084–1150.

[37] B. A. Muzembo, K. Kitahara, D. Mitra, A. Ohno, J. Khatiwada, S. Dutta and S.-I. Miyoshi, *Journal of Travel Medicine*, 2022, 1–7.

[38] I. Nisa, M. Qasim, N. Yasin, R. Ullah and A. Ali, *Folia Microbiologica*, 2020, **65**, 275–291.

[39] S. A. Plotkin, *Nature Medicine*, 2005, **11**, S5–S11.

[40] I. Delany, R. Rappuoli and E. D. Gregorio, *EMBO Molecular Medicine*, 2014, **6**, 708–720.

[41] C. A. MacLennan, S. Grow, L. fu Ma and A. D. Steele, *Vaccines*, 2022, **10**, 1376.

[42] S. Mani, T. Wierzba and R. I. Walker, *Vaccine*, 2016, **34**, 2887–2894.

[43] M. M. Levine, K. L. Kotloff, E. M. Barry, M. F. Pasetti and M. B. Sztein, *Nature Reviews Microbiology*, 2007, **5**, 540–553.

[44] K. L. Kotloff, J. K. Simon, M. F. Pasetti, M. B. Sztein, S. L. Wooden, S. Livio, J. P. Nataro, W. C. Blackwelder, E. M. Barry, W. Picking and M. M. Levine, *Human Vaccines*, 2007, **3**, 268–275.

[45] J. H. Passwell, E. Harlev, S. Ashkenazi, C. Chu, D. Miron, R. Ramon, N. Farzan, J. Shiloach, D. A. Bryla, F. Majadly, R. Roberson, J. B. Robbins and R. Schneerson, *Infection and Immunity*, 2001, **69**, 1351–1357.

[46] F. Bélot, C. Guerreiro, F. Baleux and L. A. Mulard, *Chemistry - A European Journal*, 2005, **11**, 1625–1635.

[47] L.-A. Barel and L. A. Mulard, *Human Vaccines and Immunotherapeutics*, 2019, **15**, 1338–1356.

[48] A. Ardèvol, X. Biarnés, A. Planas and C. Rovira, *Journal of the American Chemical Society*, 2010, **132**, 16058–16065.

[49] K. L. Kotloff, J. A. Platts-Mills, D. Nasrin, A. Roose, W. C. Blackwelder and M. M. Levine, *Vaccine*, 2017, **35**, 6783–6789.

[50] L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen and E. Romanowska, *European Journal of Biochemistry*, 1978, **91**, 279–284.

[51] Y. A. Knirel, Q. Sun, S. N. Senchenkova, A. V. Perepelov, A. S. Shashkov and J. Xu, *Biochemistry (Moscow)*, 2015, **80**, 901–914.

[52] T. Wehler and N. I. A. Carlin, *Eur. J. Biochem*, 1988, **176**, 471–476.

[53] R. A. Foster, N. I. Carlin, M. Majcher, H. Tabor, L. K. Ng and G. Widmalm, *Carbohydrate Research*, 2011, **346**, 872–876.

[54] P. Söderman, P.-E. Jansson and G. Widmalm, *Journal of the Chemical Society, Perkin Transactions 2*, 1998, **2**, 639–648.

[55] O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2009, **5**, 2353–2370.

[56] E. Hatcher, O. Guvench and A. D. MacKerell, *Journal of Physical Chemistry B*, 2009, **113**, 12466–12476.

[57] E. R. Hatcher, O. Guvench and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2009, **5**, 1315–1327.

[58] E. P. Raman, O. Guvench and A. D. MacKerell, *Journal of Physical Chemistry B*, 2010, **114**, 12981–12994.

[59] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2011, **7**, 3162–3180.

[60] S. S. Mallajosyula, O. Guvench, E. Hatcher and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 759–776.

[61] R. J. Woods, R. A. Dwek, C. J. Edge and B. Fraser-Reid, *The Journal of Physical Chemistry*, 1995, **99**, 3832–3846.

[62] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *Journal of Computational Chemistry*, 2008, **29**, 622–655.

[63] B. M. Sattelle and A. Almond, *Glycobiology*, 2011, **21**, 1651–1662.

[64] H. B. Mayes, L. J. Broadbelt and G. T. Beckham, *Journal of the American Chemical Society*, 2014, **136**, 1008–1022.

[65] V. S. R. Rao, P. K. Qasba, P. V. Balaji and R. Chandrasekaran, *Conformation of Carbohydrates*, CRC Press, 1998.

[66] B. M. Sattelle, B. Bose-Basu, M. Tessier, R. J. Woods, A. S. Serianni and A. Almond, *The Journal of Physical Chemistry B*, 2012, **116**, 6380–6386.

[67] O. Guvench, D. Martin and M. Greene, *International Journal of Molecular Sciences*, 2021, **23**, 473.

[68] J. Agirre, *Acta Crystallographica Section D: Structural Biology*, 2017, **73**, 171–186.

[69] G. J. Davies, A. Planas and C. Rovira, *Accounts of chemical research*, 2012, **45**, 308–316.

[70] J. C. P. Schwarz, *Journal of the Chemical Society, Chemical Communications*, 1973, 505–508.

[71] D. Cremer and J. A. Pople, *Journal of the American Chemical Society*, 1975, **97**, 1354–1358.

[72] A. D. Mcnaught, H. B. F. Dixon, F. Cornish-Bowden, M. A. Chester, A. J. Barrett, J. C. Rigg, D. Horton, L. Anderson, D. C. Baker, H. H. Baer, J. N. Bemiller, B. Bossenbroek, R. W. Jeanloz, K. L. Loening, W. A. Szarek, R. S. Tipson, W. J. Whelan and R. L. Whistler, *Pure and Appl. Chem*, 1996, **68**, 1919–2008.

[73] B. M. Sattelle and A. Almond, *Carbohydrate Research*, 2014, **383**, 34–42.

[74] B. M. Sattelle and A. Almond, *Physical Chemistry Chemical Physics*, 2014, **16**, 8119–8126.

[75] A. Geyer, M. Muller and R. R. Schmidt, *Journal of the American Chemical Society*, 1999, **121**, 6312–6313.

[76] C. Höög, C. Landersjö and G. Widmalm, *Chemistry - A European Journal*, 2001, **7**, 3069–3077.

[77] M. J. Clement, A. Imberty, A. Phalipon, S. Pérez, C. Simenel, L. A. Mulard and M. Delepierre, *Journal of Biological Chemistry*, 2003, **278**, 47928–47936.

[78] M. Kuttel, M. Gordon and N. Ravenscroft, *Carbohydrate Research*, 2014, **390**, 20–27.

[79] B. M. Sattelle and A. Almond, *Biomacromolecules*, 2013, **14**, 1149–1159.

[80] P. Reeves, *Trends in Microbiology*, 1995, **3**, 381–386.

[81] C. Whitfield, D. M. Williams and S. D. Kelly, *Journal of Biological Chemistry*, 2020, **295**, 10593–10609.

[82] J. Dahl, *Gram negative cell wall*, 2021, `https://commons.wikimedia.org/wiki/File:Gram_negative_cell_wall.svg`.

[83] M. Jones, *Lipopolysaccharide*, 2021, `https://commons.wikimedia.org/w/index.php?curid=10422301`.

[84] I. Lerouge, *FEMS Microbiology Reviews*, 2001, **25**, 1.

[85] S. Pelkonen, J. Hayrinen and J. Finne, *Polyacrylamide Gel Electrophoresis of the Capsular Polysaccharides of Escherichia coli Ki and Other Bacteria*, 1988.

[86] K. D. Kröncke, G. Boulnois, I. Roberts, D. Bitter-Suermann, J. R. Golecki, B. Jann and K. Jann, *Journal of Bacteriology*, 1990, **172**, 1085–1091.

[87] K. Bazaka, R. J. Crawford, E. L. Nazarenko and E. P. Ivanova, *Advances in Experimental Medicine and Biology*, 2011, **715**, 213–226.

[88] B. Liu, Y. A. Knirel, L. Feng, A. V. Perepelov, S. N. Senchenkova, Q. Wang, P. R. Reeves and L. Wang, *FEMS Microbiology Reviews*, 2008, **32**, 627–653.

[89] S. Livio, N. A. Strockbine, S. Panchalingam, S. M. Tennant, E. M. Barry, M. E. Marohn, M. Antonio, A. Hossain, I. Mandomando, J. B. Ochieng, J. O. Oundo, S. Qureshi, T. Ramamurthy, B. Tamboura, R. A. Adegbola, M. J. Hossain, D. Saha, S. Sen, A. S. G. Faruque, P. L. Alonso, R. F. Breiman, A. K. Zaidi, D. Sur, S. O. Sow, L. Y. Berkeley, C. E. O'Reilly, E. D. Mintz, K. Biswas, D. Cohen, T. H. Farag, D. Nasrin, Y. Wu, W. C. Blackwelder, K. L. Kotloff, J. P. Nataro and M. M. Levine, *Clinical Infectious Diseases*, 2014, **59**, 933–941.

[90] A. V. Perepelov, M. E. Shekht, B. Liu, S. D. Shevelev, V. A. Ledov, S. N. Senchenkova, V. L. L'vov, A. S. Shashkov, L. Feng, P. G. Aparin, L. Wang and Y. A. Knirel, *FEMS Immunology and Medical Microbiology*, 2012, **66**, 201–210.

[91] A. N. Kondakova, E. V. Vinogradov, M. E. Shekht, A. A. Markina, B. Lindner, V. L. Vov, P. G. Aparin and Y. A. Knirel, *Russian Journal of Bioorganic Chemistry*, 2010, **36**, 396–399.

[92] L. L. V. D. Verg, N. O. Bendiuk, K. Kotloff, M. M. Marsh, J. L. Ruckert, J. L. Puryear, D. N. Taylor and A. B. Hartman, *Vaccine*, 1996, **14**, 1062–1068.

[93] F. R. Noriega, F. M. Liao, D. R. Maneval, S. Ren, S. B. Formal and M. M. Levine, *Strategy for Cross-Protection among Shigella flexneri Serotypes*, 1999.

[94] N. P. West, P. Sansonetti, J. Mounier, R. M. Exley, C. Parsot, S. Guadagnini, M.-C. Prevost, A. Prochnicka-Chalufour, M. Delepierre, M. Tanguy and C. M. Tang, *Science*, 2005, **307**, 1313–1317.

[95] R. T. Trevejo, S. L. Abbott, M. I. Wolfe, J. Meshulam, D. Yong and G. R. Flores, *Journal of Clinical Microbiology*, 1999, **37**, 2352–2353.

[96] A. El-Gendy, N. El-Ghorab, E. M. Lane, R. A. Elyazeed, N. I. A. Carlin, M. M. Mitry, B. A. Kay, S. J. Savarino and L. F. Peruski, *Journal of Clinical Microbiology*, 1999, **37**, 873–874.

[97] K. A. Talukder, M. A. Islam, B. K. Khajanchi, D. K. Dutta, Z. Islam, A. Safa, K. Alam, A. Hossain, G. B. Nair and D. A. Sack, *Journal of Clinical Microbiology*, 2003, **41**, 5053–5058.

[98] S. F. Ahmed, M. S. Riddle, T. F. Wierzba, I. A. Messih, M. R. Monteville, J. W. Sanders and J. D. Klena, *Epidemiology and Infection*, 2006, **134**, 1237–1248.

[99] L. von Seidlein, D. R. Kim, M. Ali, H. Lee, X. Wang, V. D. Thiem, D. G. Canh, W. Chaicumpa, M. D. Agtini, A. Hossain, Z. A. Bhutta, C. Mason, O. Sethabutr, K. Talukder, G. B. Nair, J. L. Deen, K. Kotloff and J. Clemens, *PLoS Medicine*, 2006, **3**, e353.

[100] R. Stagg, P. Cam and N. Verma, *Epidemiology and Infection*, 2008, **136**, 1134–1140.

[101] R. M. Stagg, S.-S. Tang, N. I. A. Carlin, K. A. Talukder, P. D. Cam and N. K. Verma, *Journal of Bacteriology*, 2009, **191**, 6612–6617.

[102] T. R. Connor, C. R. Barker, K. S. Baker, F.-X. Weill, K. A. Talukder, A. M. Smith, S. Baker, M. Gouali, D. P. Thanh, I. J. Azmi, W. D. da Silveira, T. Semmler, L. H. Wieler, C. Jenkins, A. Cravioto, S. M. Faruque, J. Parkhill, D. W. Kim, K. H. Keddy and N. R. Thomson, *eLife*, 2015, **4**, 1001–1006.

[103] K. L. Kotloff, D. Nasrin, W. C. Blackwelder, Y. Wu, T. Farag, S. Panchalingham, S. O. Sow, D. Sur, A. K. Zaidi, A. S. Faruque, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, S. Ahmed, S. Qureshi, F. Quadri, A. Hossain, S. K. Das, M. Antonio, M. J. Hossain, I. Mandomando, S. Acácio, K. Biswas, S. M. Tennant, J. J. Verweij, H. Sommerfelt, J. P. Nataro, R. M. Robins-Browne and M. M. Levine, *The Lancet Global Health*, 2019, **7**, e568–e584.

[104] C. A. Bush, M. Martin-Pastor and A. Imbery, *Annual Review of Biophysics and Biomolecular Structure*, 1999, **28**, 269–293.

[105] A. T. Brünger, *Nature structural biology*, 1997, **4**, 862–865.

[106] J. Keeler, *Understanding NMR spectroscopy*, Wiley, 2005, p. 459.

[107] M. Karplus, *The Journal of Chemical Physics*, 1959, **30**, 11–15.

[108] F. Cloran, I. Carmichael and A. S. Serianni, *Journal of the American Chemical Society*, 1999, **121**, 9843–9851.

[109] K. Trainor, J. A. Palumbo, D. W. S. MacKenzie and E. M. Meiering, *Protein Science*, 2020, **29**, 306–314.

[110] E. Säwén, F. Hinterholzinger, C. Landersjö and G. Widmalm, *Organic and Biomolecular Chemistry*, 2012, **10**, 4577–4585.

[111] J.-R. Brisson, S. Uhrinova, R. J. Woods, M. van der Zwan, H. C. Jarrell, L. C. Paoletti, D. L. Kasper and H. J. Jennings, *Biochemistry*, 1997, **36**, 3278–3292.

[112] C. Jones, F. Currie and M. J. Forster, *Carbohydrate Research*, 1991, **221**, 95–121.

[113] M. J. Thrippleton and J. Keeler, *Angewandte Chemie - International Edition*, 2003, **42**, 3938–3941.

[114] M. P. Williamson, in *Applications of the NOE in Molecular Biology*, Elsevier, 2009, vol. 65, pp. 77–109.

[115] B. Vögeli, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2014, **78**, 1–46.

[116] S. Macura, B. T. F. Ii and L. R. Brown, *Journal of Magnetic Resonance*, 1986, **70**, 493–499.

[117] H. Hu and K. Krishnamurthy, *Journal of Magnetic Resonance*, 2006, **182**, 173–177.

[118] A. M. Dixon, G. Widmalm and T. Bull, *Journal of Magnetic Resonance*, 2000, **147**, 266–272.

[119] M. Linnerborg, A. Weintraub and G. Widmalm, *European Journal of Biochemistry*, 1999, **266**, 246–251.

[120] K. H. M. Jonsson, A. Weintraub and G. Widmalm, *Carbohydrate Research*, 2006, **341**, 2986–2989.

[121] E. Säwén, *PhD thesis*, University of Stockholm, 2011.

[122] D. I. Freedberg and J. Kwon, *Israel Journal of Chemistry*, 2019, **59**, 1039–1058.

[123] P.-E. Jansson, R. Stenutz and G. Widmalm, *Carbohydrate Research*, 2006, **341**, 1003–1010.

[124] M. Lundborg and G. Widmalm, *Analytical Chemistry*, 2011, **83**, 1514–1517.

[125] A. Furevi, A. Ruda, T. A. d'Ortoli, H. Mobarak, J. Ståhle, C. Hamark, C. Fontana, O. Engström, P. Apostolica and G. Widmalm, *Carbohydrate Research*, 2022, **513**, 1–5.

[126] B. Liu, A. Furevi, A. V. Perepelov, X. Guo, H. Cao, Q. Wang, P. R. Reeves, Y. A. Knirel, L. Wang and G. Widmalm, *FEMS Microbiology Reviews*, 2020, **44**, 655–683.

[127] J. C. Paton and C. Trappetti, *Microbiology Spectrum*, 2019, **7**, 1–15.

[128] W. Plazinski, A. Lonardi and P. H. Hünenberger, *Journal of Computational Chemistry*, 2016, **37**, 354–365.

[129] H. Masoud, M. B. Perry, J.-R. Brisson, D. Uhrin and J. C. Richards, *Canadian Journal of Chemistry*, 1994, **72**, 1466–1477.

[130] T. Holme, M. Rahman, P.-E. Jansson and G. Widmalm, *European Journal of Biochemistry*, 1999, **265**, 524–529.

[131] S. Kunstmann, O. Engström, M. Wehle, G. Widmalm, M. Santer and S. Barbirz, *Chemistry - A European Journal*, 2020, **26**, 7263–7273.

[132] B. P. Hills, C. Cano and P. S. Belton, *Macromolecules*, 1991, **24**, 2944–2950.

[133] H.-Y.-Y. Yao, J.-Q. Wang, J.-Y. Yin, S.-P. Nie and M.-Y. Xie, *Food Research International*, 2021, **143**, 110290.

[134] K. Lycknert, P. Edebrink and G. Widmalm, *Angewandte Chemie International Edition*, 2004, **43**, 2288–2290.

[135] A. Sarkar, C. Fontana, A. Imberty, S. Pérez and G. Widmalm, *Biomacromolecules*, 2013, **14**, 2215–2224.

[136] E. Wu, O. Engström, S. Jo, D. Stuhlsatz, M. Yeom, J. Klauda, G. Widmalm and W. Im, *Biophysical Journal*, 2013, **105**, 1444–1455.

[137] M. A. Rojas-Macias, J. Stahle, T. Lutteke and G. Widmalm, *Glycobiology*, 2015, **25**, 341–347.

[138] P. Blasco, D. S. Patel, O. Engström, W. Im and G. Widmalm, *Biochemistry*, 2017, **56**, 3826–3839.

[139] D. S. Patel, P. Blasco, G. Widmalm and W. Im, *Current Research in Structural Biology*, 2020, **2**, 79–88.

[140] R. J. Woods, *Glycoconjugate Journal volume* , 1998, **15**, 209–216.

[141] A. Laederach and P. J. Reilly, *Proteins: Structure, Function, and Bioinformatics*, 2005, **60**, 591–597.

[142] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.

[143] J. W. Ponder and D. A. Case, in *Force Fields for Protein Simulations*, ed. V. Daggett, Elsevier, 2003, pp. 27–85.

[144] A. D. Mackerell, M. Feig and C. L. Brooks, *Journal of Computational Chemistry*, 2004, **25**, 1400–1415.

[145] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 3257–3273.

[146] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, *Nature Methods*, 2017, **14**, 71–73.

[147] M. Schlenkrich, J. Brickmann, A. D. MacKerell and M. Karplus, in *An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications*, Birkhäuser Boston, 1996, pp. 31–81.

[148] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor, *The Journal of Physical Chemistry B*, 2010, **114**, 7830–7843.

[149] S. Feller, D. Yin, R. Pastor and A. MacKerell, *Biophysical Journal*, 1997, **73**, 2269–2279.

[150] N. Foloppe and A. D. M. Jr., *Journal of Computational Chemistry*, 2000, **21**, 86–104.

[151] E. J. Denning, U. D. Priyakumar, L. Nilsson and A. D. Mackerell, *Journal of Computational Chemistry*, 2011, **32**, 1929–1943.

[152] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 348–362.

[153] T. E. Cheatham and D. A. Case, *Biopolymers*, 2013, **99**, 969–77.

[154] E. Fadda and R. J. Woods, *Drug Discovery Today*, 2010, **15**, 596–609.

[155] A. Almond, *Current Opinion in Structural Biology*, 2018, **50**, 58–64.

[156] E. Fadda, *Current Opinion in Chemical Biology*, 2022, **69**, 102175.

[157] L. J. Madsen, S. N. Ha, V. H. Tran and J. W. Brady, in *Computer Modeling of Carbohydrate Molecules*, American Chemical Society, 1990, pp. 69–90.

[158] R. J. Woods, in *The Application of Molecular Modeling Techniques to the Determination of Oligosaccharide Solution Conformations*, John Wiley and Sons, Inc., 2007, vol. 9, pp. 129–165.

[159] M. L. DeMarco and R. J. Woods, *Glycobiology*, 2008, **18**, 426–440.

[160] M. W. van der Kamp and A. J. Mulholland, *Biochemistry*, 2013, **52**, 2708–2728.

[161] L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella and L. Genovese, *WIREs Computational Molecular Science*, 2017, **7**, 1–8.

[162] D. Frenkel and B. Smit, *Understanding molecular simulation*, Academic Press, San Diego, CA, 1996.

[163] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *The Journal of Chemical Physics*, 1982, **76**, 637–649.

[164] A. Laio and M. Parrinello, *Proceedings of the National Academy of Sciences*, 2002, **99**, 12562–12566.

[165] G. Bussi and A. Laio, *Nature Reviews Physics*, 2020, **2**, 200–212.

[166] L. V. Zhigilei, *Boundary Conditions*, 2021, `https://www.docsity.com/en/notes-on-boundary-conditions-introduction-to-atomistic-simulations-mse-627/6739325/`.

[167] K. M. Dyer, J. S. Perkyns, G. Stell and B. M. Pettitt, *Molecular Physics*, 2009, **107**, 423–431.

[168] H. Nada and J. P. J. M. van der Eerden, *The Journal of Chemical Physics*, 2003, **118**, 7401.

[169] J. L. F. Abascal, R. G. Fernández, C. Vega and M. A. Carignano, *The Journal of Chemical Physics*, 2006, **125**, 166101.

[170] H. Nada, *The Journal of Chemical Physics*, 2016, **145**, 244706.

[171] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926–935.

[172] B. L. Foley, M. B. Tessier and R. J. Woods, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 652–697.

[173] T. J. Rutherford, J. Partridge, C. T. Weller and S. W. Homans, *Biochemistry*, 1993, **32**, 12715–12724.

[174] E. W. Wooten, C. J. Edge, R. Bazzo, R. A. Dwek and T. W. Rademacher, *Carbohydrate Research*, 1990, **203**, 13–17.

[175] B. M. Sattelle and A. Almond, *Journal of Computational Chemistry*, 2010, **31**, 2932–2947.

[176] R. D. Lins and P. H. Hünenberger, *Journal of Computational Chemistry*, 2005, **26**, 1400–1412.

[177] E. Autieri, M. Sega, F. Pederiva and G. Guella, *The Journal of Chemical Physics*, 2010, **133**, 095104.

[178] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. van Gunsteren, *European Biophysics Journal*, 2011, **40**, 843–856.

[179] T. G. S. for (Bio)Molecular Simulation, *Volume 3: Force Field and Topology Data Set*, 2023, `https://www.gromos.net/gromos11_pdf_manuals/vol3.pdf`.

[180] W. Damm, A. Frontera, J. Tirado-Rives and W. L. Jorgensen, *Journal of Computational Chemistry*, 1997, **18**, 1955–1970.

[181] D. Kony, W. Damm, S. Stoll and W. F. V. Gunsteren, *Journal of Computational Chemistry*, 2002, **23**, 1416–1429.

[182] M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *Journal of Chemical Theory and Computation*, 2015, **11**, 3499–3509.

[183] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.

[184] A. H. Aytenfisu, M. Yang and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2018, **14**, 3132–3143.

[185] D. S. Patel, X. He and A. D. Mackerell, *Journal of Physical Chemistry B*, 2015, **119**, 637–652.

[186] M. Jana and A. D. MacKerell, *The Journal of Physical Chemistry B*, 2015, **119**, 7846–7859.

[187] M. Yang, A. H. Aytenfisu and A. D. MacKerell, *Carbohydrate Research*, 2018, **457**, 41–50.

[188] X. He, P. E. M. Lopes and A. D. MacKerell, *Biopolymers*, 2013, **99**, 724–738.

[189] A. A. Kognole, J. Lee, S. Park, S. Jo, P. Chatterjee, J. A. Lemkul, J. Huang, A. D. MacKerell and W. Im, *Journal of Computational Chemistry*, 2022, **43**, 359–375.

[190] J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, *Chemical Reviews*, 2016, **116**, 4983–5013.

[191] A. Ruda, A. H. Aytenfisu, T. A. d'Ortoli, A. D. MacKerell and G. Widmalm, *Physical Chemistry Chemical Physics*, 2023, 3042–3060.

[192] M. L. DeMarco and R. J. Woods, *Glycobiology*, 2008, **19**, 344–355.

[193] M. L. DeMarco, in *Molecular Dynamics Simulations of Membrane- and Protein-Bound Glycolipids Using GLYCAM*, Humana Press, 2015, vol. 1273, pp. 379–390.

[194] K. N. Kirschner, R. D. Lins, A. Maass and T. A. Soares, *Journal of Chemical Theory and Computation*, 2012, **8**, 4719–4731.

[195] M. Tessier, M. DeMarco, A. Yongye and R. Woods, *Molecular Simulation*, 2008, **34**, 349–364.

[196] S. Perez and O. Makshakova, *Chemical Reviews*, 2022, **122**, 15914–15970.

[197] C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *The Journal of Physical Chemistry*, 1993, **97**, 10269–10280.

[198] M. Basma, S. Sundara, D. Çalgan, T. Vernali and R. J. Woods, *Journal of Computational Chemistry*, 2001, **22**, 1125–1137.

[199] R. D. Lazar, F. B. Akher, N. Ravenscroft and M. M. Kuttel, *Journal of Chemical Theory and Computation*, 2022, **18**, 1156–1172.

[200] M. Zaccheus, R. Pendrill, T. A. Jackson, A. Wang, F.-I. Auzanneau and G. Widmalm, *European Journal of Organic Chemistry*, 2012, **2012**, 4705–4715.

[201] K. N. Kirschner and R. J. Woods, *Proceedings of the National Academy of Sciences*, 2001, **98**, 10541–10545.

[202] O. Guvench, S. N. Greenr, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell, *Journal of Computational Chemistry*, 2008, **29**, 2543–2564.

[203] O. Guvench and A. D. MacKerell, *Journal of Molecular Modeling*, 2008, **14**, 667–679.

[204] J. N. Chythra and S. S. Mallajosyula, *Journal of Chemical Information and Modeling*, 2022, 7846–7859.

[205] H. Miyamoto, U. Schnupf, M. F. Crowley and J. W. Brady, *Carbohydrate Research*, 2016, **422**, 17–23.

[206] B. Nagarajan, N. V. Sankaranarayanan and U. R. Desai, *Glycobiology*, 2020, **30**, 516–527.

[207] J. F. Matthews, G. T. Beckham, M. Bergenstråhle-Wohlert, J. W. Brady, M. E. Himmel and M. F. Crowley, *Journal of Chemical Theory and Computation*, 2012, **8**, 735–748.

[208] J. Sauter and A. Grafmüller, *Journal of Chemical Theory and Computation*, 2015, **11**, 1765–1774.

[209] J. Gebhardt, C. Kleist, S. Jakobtorweihen and N. Hansen, *The Journal of Physical Chemistry B*, 2018, **122**, 1608–1626.

[210] J. K. Koneru, X. Zhu and J. Mondal, *Journal of Chemical Theory and Computation*, 2019, **15**, 6203–6212.

[211] M. C. Owen, A. Karner, R. Šachl, J. Preiner, M. Amaro and R. Vácha, *The Journal of Physical Chemistry B*, 2019, **123**, 7504–7517.

[212] A. Plazinska and W. Plazinski, *Journal of Chemical Theory and Computation*, 2021, **17**, 2575–2585.

[213] A. R. Fourie, *PhD thesis*, University of Cape Town, 2020.

[214] M. M. Kuttel, *Mini-Reviews in Organic Chemistry*, 2011, **8**, 256–262.

[215] X. Biarnés, A. Ardèvol, A. Planas, C. Rovira, A. Laio and M. Parrinello, *Journal of the American Chemical Society*, 2007, **129**, 10686–10693.

[216] C. B. Barnett and K. J. Naidoo, *Molecular Physics*, 2009, **107**, 1243–1250.

[217] V. Spiwok, B. Králová and I. Tvaroška, *Carbohydrate Research*, 2010, **345**, 530–537.

[218] H. S. Hansen and P. H. HÃ¼nenberger, *Journal of Computational Chemistry*, 2010, **31**, 1–23.

[219] D. A. Rees, *Journal of the Chemical Society B: Physical Organic*, 1970, 877.

[220] R. Polacek, J. Stenger and U. Kaatze, *The Journal of Chemical Physics*, 2002, **116**, 2973–2982.

[221] S. Lyu, N. Beiranvand, M. Freindorf and E. Kraka, *Journal of Physical Chemistry A*, 2019, **123**, 7087–7103.

[222] H. A. Taha, P. N. Roy and T. L. Lowary, *Journal of Chemical Theory and Computation*, 2011, **7**, 420–432.

[223] A. G. Gerbst, A. A. Grachev, D. V. Yashunsky, Y. E. Tsvetkov, A. S. Shashkov and N. E. Nifantiev, *Journal of Carbohydrate Chemistry*, 2013, **32**, 205–221.

[224] J. R. Bishop, M. Schuksz and J. D. Esko, *Nature*, 2007, **446**, 1030–1037.

[225] L. Kong and R. A. Bryce, *Journal of Computational Chemistry*, 2022, **43**, 2009–2022.

[226] C. R. Raetz and W. Dowhan, *Journal of Biological Chemistry*, 1990, **265**, 1235–1238.

[227] C. R. Raetz and C. Whitfield, *Lipopolysaccharide endotoxins*, 2002.

[228] M. M. Kuttel, J. Ståhle and G. Widmalm, *Journal of Computational Chemistry*, 2016, 2098–2105.

[229] A. D. French, in *Computerized Models of Carbohydrates*, Springer International Publishing, 2014, pp. 1–38.

[230] J. Landström and G. Widmalm, *Carbohydrate Research*, 2010, **345**, 330–333.

[231] C. Fontana, S. Li, Z. Yang and G. Widmalm, *Carbohydrate Research*, 2015, **402**, 87–94.

[232] K. Heyns and G. Kiessling, *Carbohydrate Research*, 1967, **3**, 340–353.

[233] S. C. Szu and S. Bystrický, *Methods in Enzymology*, 2003, **363**, 552–567.

[234] L. Legnani, F. Compostella, G. Grazioso, F. M. Albini and L. Toma, *Organic and Biomolecular Chemistry*, 2011, **9**, 5554.

[235] K. Hitri, M. M. Kuttel, G. De Benedetto, K. Lockyer, F. Gao, P. Hansal, T. R. Rudd, E. Beamish, S. Rijpkema, N. Ravenscroft and B. Bolgiano, *Vaccine*, 2019, **37**, 3866–3875.

[236] D. H. Shah, J. R. Elder, K. L. Chiok and N. C. Paul, in *Genetic Basis of Salmonella Enteritidis Pathogenesis in Chickens*, Elsevier, 2017, pp. 187–208.

[237] D. S. Snyder, D. Gibson, C. Heiss, W. Kay and P. Azadi, *Carbohydrate Research*, 2006, **341**, 2388–2397.

[238] J. D. M. Olsson, J. Landstrom, J. Ronnols, S. Oscarson and G. Widmalm, *Organic and Biomolecular Chemistry*, 2009, **7**, 1612.

[239] A. M. Dixon, R. Venable, G. Widmalm, T. Bull and R. W. Pastor, *Biopolymers*, 2003, **69**, 448–460.

[240] J. B. Kaper, J. P. Nataro and H. L. T. Mobley, *Nature Reviews Microbiology*, 2004, **2**, 123–140.

[241] B. Jann, A. A. Shashkov, H. Kochanowski and K. Jann, *Carbohydrate Research*, 1994, **263**, 217–225.

[242] D. S. Patel, R. Pendrill, S. S. Mallajosyula, G. Widmalm and A. D. MacKerell, *Journal of Physical Chemistry B*, 2014, **118**, 2851–2871.

[243] T. J. Rutherford, C. Jones, D. B. Davies and A. C. Elliott, *Carbohydrate Research*, 1994, **265**, 79–96.

[244] R. Kadirvelraj, J. Gonzalez-Outeiriño, B. L. Foley, M. L. Beckham, H. J. Jennings, S. Foote, M. G. Ford and R. J. Woods, *Proceedings of the National Academy of Sciences*, 2006, **103**, 8149–8154.

[245] L. Legnani, S. Ronchi, S. Fallarini, G. Lombardi, F. Campo, L. Panza, L. Lay, L. Poletti, L. Toma, F. Ronchetti and F. Compostella, *Organic and Biomolecular Chemistry*, 2009, **7**, 4428.

[246] M. Kuttel and N. Ravenscroft, *Pharmaceuticals*, 2019, **12**, 28–31.

[247] L. Hao, M. M. Kuttel, N. Ravenscroft, A. Thompson, A. K. Prasad, S. Gangolli, C. Tan, D. Cooper, W. Watson, P. Liberator, M. W. Pride, K. U. Jansen, A. S. Anderson and I. L. Scully, *Vaccine*, 2022, **40**, 4872–4880.

[248] A. J. van Tonder, J. E. Bray, S. J. Quirk, G. Haraldsson, K. A. Jolley, M. C. J. Maiden, S. Hoffmann, S. D. Bentley, A. Haraldsson, H. Erlendsdottir, K. G. Kristinsson and A. B. Brueggemann, *Microbial Genomics*, 2016, **2**, 5.

[249] L. R. Grant, S. E. O'Brien, P. Burbidge, M. Haston, M. Zancolli, L. Cowell, M. Johnson, R. C. Weatherholtz, R. Reid, M. Santosham, K. L. O'Brien and D. Goldblatt, *PLoS ONE*, 2013, **8**, e74906.

[250] K. A. Geno, G. L. Gilbert, J. Y. Song, I. C. Skovsted, K. P. Klugman, C. Jones, H. B. Konradsen and M. H. Nahm, *Clinical Microbiology Reviews*, 2015, **28**, 871–899.

[251] J. M. Griffiss, B. L. Brandt, P. L. Altieri, G. B. Pier and S. L. Berman, *Infection and Immunity*, 1981, **34**, 725–732.

[252] M. M. Kuttel, F. Berti and N. Ravenscroft, *Glycoconjugate Journal*, 2021, **38**, 411–419.

[253] P. Domingo, V. Pomar, A. Mauri and N. Barquet, *The Lancet Infectious Diseases*, 2019, **19**, e284–e294.

[254] K. Bock, S. Josephson and D. R. Bundle, *Journal of the Chemical Society, Perkin Transactions 2*, 1982, 59.

[255] M. Kuttel, Y. Mao, G. Widmalm and M. Lundborg, *2011 IEEE Seventh International Conference on eScience*, 2011, 395–402.

[256] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781–1802.

[257] R. Bernardi, M. Bhandarkar, A. Bhatele, E. Bohm, R. B. F. Buelens, C. Chipot, A. Dalke, S. Dixit, G. Fiorin, P. Freddolino, H. Fu, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, C. Harrison, J. Henin, W. Humphrey, D. Hurwitz, A. Hynninen, N. Jain, N. Krawetz, S. Kumar, D. Kunzman, J. Lai, C. Lee, J. Maia, R. McGreevy, C. Mei, M. Melo, M. Nelson, J. Phillips, B. Radak, T. Rudack, O. Sarood, A. Shinozaki, D. Tanner, D. Wells, G. Zheng

and F. Zhu, *NAMD 2.13 User's Guide*, `https://www.ks.uiuc.edu/Research/namd/2.13/ug/ug.html`, 2018, [Accessed 25-07-2019].

[258] S. Nosé and M. L. Klein, *Molecular Physics*, 1983, **50**, 1055–1076.

[259] W. G. Hoover, *Physical Review A*, 1985, **31**, 1695–1697.

[260] S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *The Journal of Chemical Physics*, 1995, **103**, 4613–4621.

[261] *NAMD Configuration Files*, `https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/node27.html`.

[262] W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

[263] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.

[264] R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney and O. Beckstein, 2016, pp. 98–105.

[265] M. Kuttel, J. Gain, A. Burger and I. Eborn, *Journal of Molecular Graphics and Modelling*, 2006, **25**, 380–388.

[266] S. Cross, M. M. Kuttel, J. E. Stone and J. E. Gain, *Journal of Molecular Graphics and Modelling*, 2009, **28**, 131–139.

[267] H. Flyvbjerg and H. G. Petersen, *The Journal of Chemical Physics*, 1989, **91**, 461–466.

[268] A. Grossfield and D. M. Zuckerman, in *Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations*, Elsevier, 2009, vol. 5, pp. 23–48.

[269] G. I. Csonka, Éliás Krisztina and I. G. Csizmadia, *Chemical Physics Letters*, 1996, **257**, 49–60.

[270] H. Kooijman, *Interpretation of crystal structure determinations*, 2005, `http://www.crystal.chem.uu.nl/people/huub.html`.

[271] J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, *Journal of Chemical Theory and Computation*, 2007, **3**, 2312–2334.

[272] T. M. Abramyan, J. A. Snyder, A. A. Thyparambil, S. J. Stuart and R. A. Latour, *Journal of computational chemistry*, 2016, **37**, 1973–1982.

[273] L. J. Heyer, S. Kruglyak and S. Yooseph, *Genome Research*, 1999, **9**, 1106–1115.

[274] C. B. Barnett and K. J. Naidoo, *Journal of Physical Chemistry B*, 2010, **114**, 17142–17154.

[275] C. B. Barnett and K. J. Naidoo, *Journal of Physical Chemistry B*, 2013, **117**, 6019–6026.

[276] W. F. van Gunsteren, X. Daura, N. Hansen, A. E. Mark, C. Oostenbrink, S. Riniker and L. J. Smith, *Angewandte Chemie International Edition*, 2018, **57**, 884–902.

[277] B. Nagarajan, S. G. Holmes, N. V. Sankaranarayanan and U. R. Desai, *Current Opinion in Structural Biology*, 2022, **74**, 102356.

[278] W. Group, *GLYCAM Web*, 2020, `http://glycam.org`.

[279] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Computer Physics Communications*, 2009, **180**, 1961–1972.

[280] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604–613.

[281] D. A. Colón-Ramos, P. L. Riviere, H. Shroff and R. Oldenbourg, *Nature Methods*, 2019, **16**, 667–669.

[282] S. Ikemizu, L. M. Sparks, P. A. van der Merwe, K. Harlos, D. I. Stuart, E. Y. Jones and S. J. Davis, *Proceedings of the National Academy of Sciences*, 1999, **96**, 4289–4294.

[283] C. Chang, E. Magracheva, S. Kozlov, S. Fong, G. Tobin, S. Kotenko, A. Wlodawer and A. Zdanov, *Journal of Biological Chemistry*, 2003, **278**, 3308–3313.

[284] L. Maveyraud, H. Niwa, V. Guillet, D. I. Svergun, P. V. Konarev, R. A. Palmer, W. J. Peumans, P. Rougé, E. J. M. Van Damme, C. D. Reynolds and L. Mourey, *Proteins: Structure, Function, and Bioinformatics*, 2009, **75**, 89–103.

[285] R. Xu, R. McBride, J. C. Paulson, C. F. Basler and I. A. Wilson, *Journal of Virology*, 2010, **84**, 1715–1721.

[286] A. Furevi, *PhD thesis*, University of Stockholm, 2022.

[287] Ēriks Kupče, *Methods in Enzymology*, 2002, 82–111.

[288] P. D. Thomas, V. J. Basus and T. L. James, *Source*, 1991, **88**, 1237–1241.

[289] K. Panczyk and W. Plazinski, *Carbohydrate Research*, 2018, **455**, 62–70.

[290] R. Pendrill, *Conformational studies of carbohydrates: MD simulation and NMR spectroscopy*, 2011.

[291] R. Pendrill, E. Säwén and G. Widmalm, *The Journal of Physical Chemistry B*, 2013, **117**, 14709–14722.

[292] G. Widmalm, R. Byrd and W. Egan, *Carbohydrate Research*, 1992, **229**, 195–211.

[293] K. H. M. Jonsson, R. Pendrill and G. Widmalm, *Magnetic Resonance in Chemistry*, 2011, **49**, 117–124.

[294] E. Hatcher, E. Säwén, G. Widmalm and A. D. MacKerell, *Journal of Physical Chemistry B*, 2011, **115**, 597–608.

[295] W. Plazinski, T. Angles d'Ortoli and G. Widmalm, *Org. Biomol. Chem.*, 2023, **21**, 6979–6994.

[296] M. Yang, T. Angles, E. S. wé, M. Jana, ac Gö ran Widmalm and A. D. M. Jr, *Phys. Chem. Chem. Phys. Phys. Chem. Chem. Phys*, 2016, **18**, 18776–18794.

[297] D. Mikhailov, K. H. Mayo, I. R. Vlahov, T. Toida, A. PErvin and R. J. Linhardt, *Biochemical Journal*, 1996, **318**, 93–102.

[298] G. D. Rockwell and T. B. Grindley, *Journal of the American Chemical Society*, 1998, **120**, 10953–10963.

[299] C. Haasnoot, F. de Leeuw and C. Altona, *Tetrahedron*, 1980, **36**, 2783–2792.

Appendix A: Modelling

A.1 *S. flexneri* O-Ag O-factors

A.2 MD simulations of 3,4-disubstituted βDGlcNAc trisaccharides

A.2.1 Simulation convergence

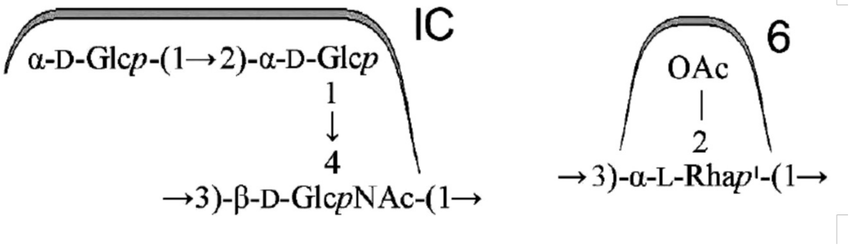


Figure A.1: O-factors of *S. flexneri* 7a and 7b O-Ags. *Sf* 7a includes O-factor IC only while *Sf* 7b includes O-factors IC and 6.[51]

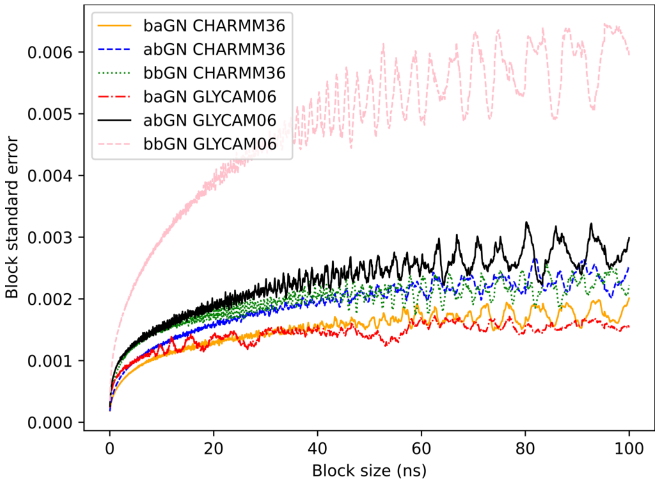


Figure A.2: Block Standard Error (BSE) of baGN, abGN, and bbGN from the CHARMM36 and GLYCAM06 simulations.

A.2.2 Glycosidic linkage torsion angles

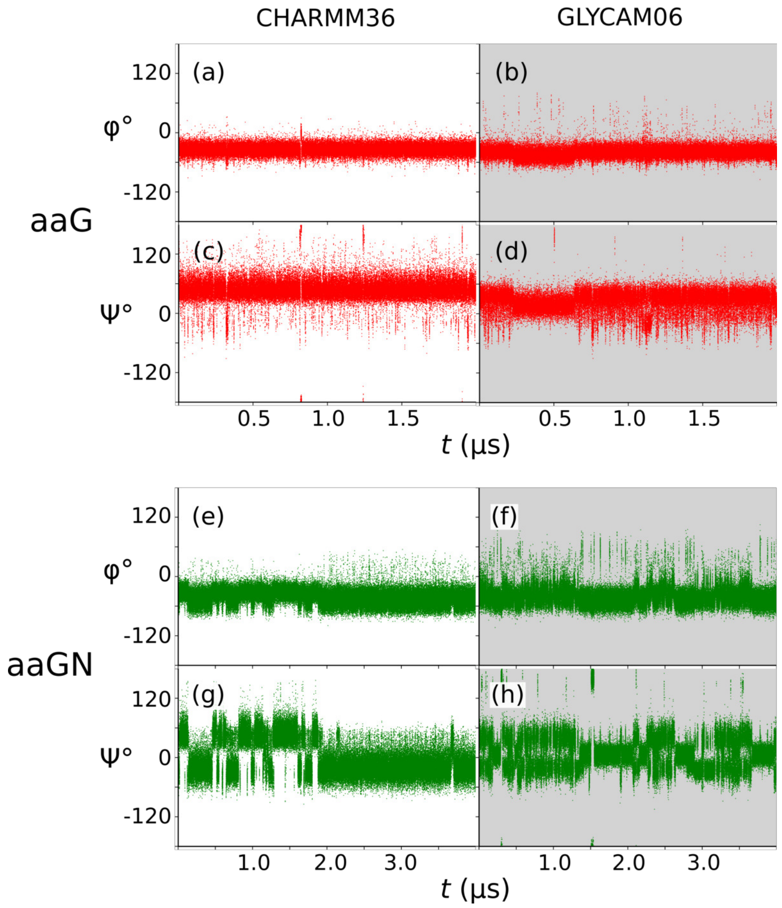


Figure A.3: Time series plots of the 1→3 glycosidic linkage in aaG and aaGN.

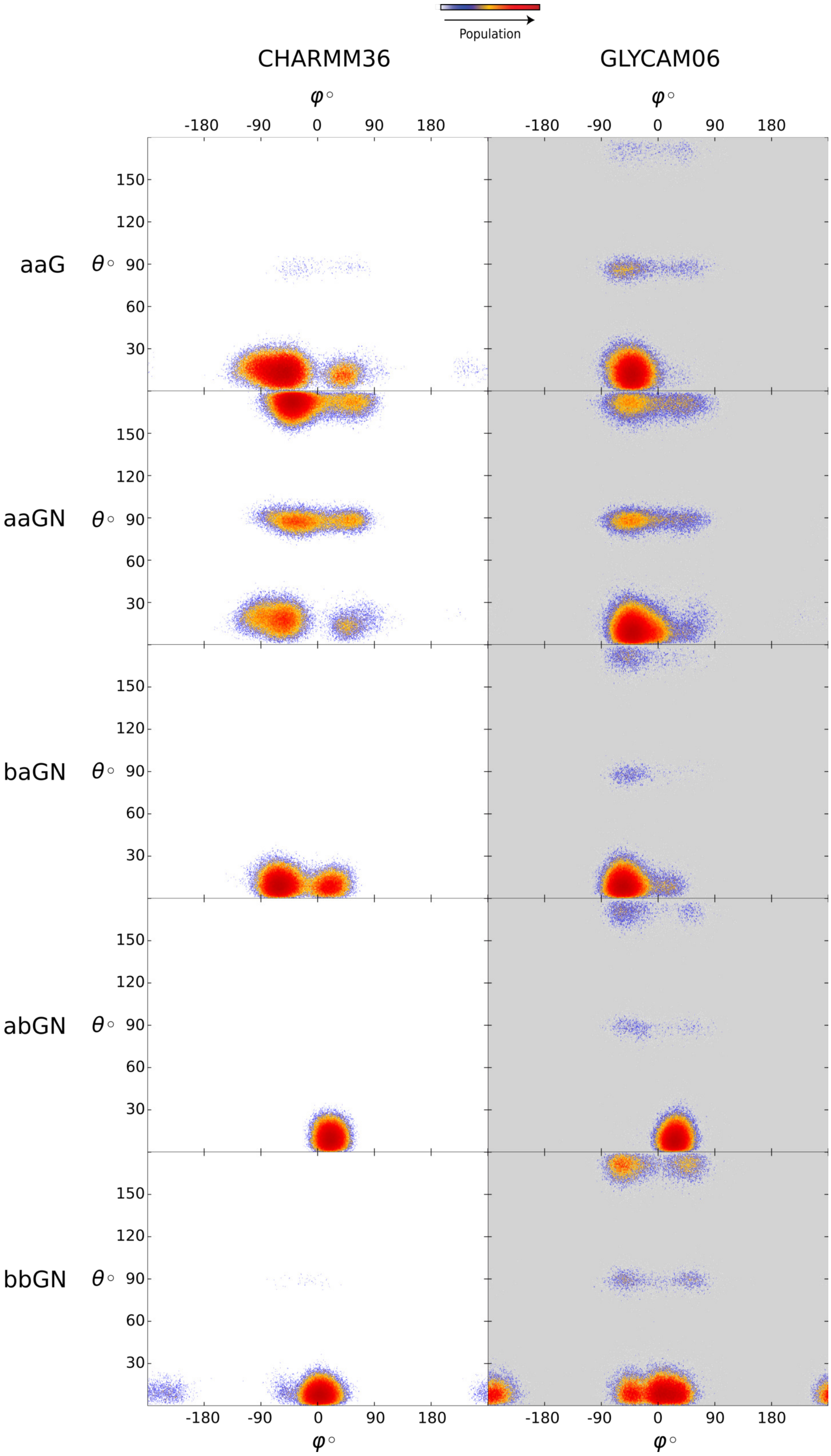


Figure A.4: Scatter plots illustrated as heat maps, with heat scale on the top, for the **Glc^{II}** glycosidic torsion angle (ψ) against Cremer-Pople azimuth (θ) parameter for GlcNAc in aaG, aaGN, baGN, abGN and bbGN. Modelled with CHARMM36 (white background) and GLYCAM06 (grey background).

A.2.3 Pucker kinetics of β-D-Glc*p*NAc in aaGN, baGN, abGN and bbGN

Table A.1: Pucker and kinetic parameters for β-D-Glc*p*NAc in α-D-Glc(1→3)[α-D-Glc(1→4)]β-D-Glc*p*NAc (aaGN). Pucker conformers in the CHARMM36 simulation show distinct differences at before and after approx. 2 μs. Parameters measured on both sides of this time are also provided.

Pucker Kinetic Parameters	**CHARMM36**			**GLYCAM06**
	0-2 μs	2-4 μs	0-4 μs	0-4 μs
Average Q	0.598	0.529	0.559	0.552
Total number of $^4C_1$ →boat/skew transitions	12 (5)	0	12	28
Average rate of $^4C_1$ →boat/skew transitions (μs^{-1})	5	0	3	7
Total number of boat/skew→^{4} C_1 transitions	10 (5)	0	10	52
Average rate of boat/skew→^{4} C_1 transitions (μs^{-1})	5	0	2.5	13
Total number of $^1C_4$ →boat/skew transitions	1 (0.5)	2	3	16
Average rate of $^1C_4$ →boat/skew transitions (μs^{-1})	0.5	1	8.5	8
Total number of boat/skew→^{1} C_4 transitions	1 (0.5)	2	3	15
Average rate of boat/skew→^{1} C_4 transitions (μs^{-1})	0.5	1	7.5	7.5
Average lifetime of $^4C_1$ chair (μs)	0.26	0	0.6	0.54
Average lifetime of $^1C_4$ chair (μs)	0.23	0.96	0.89	50
Average lifetime of boat/skew (μs)	0.3	0.02	0.08	0.42

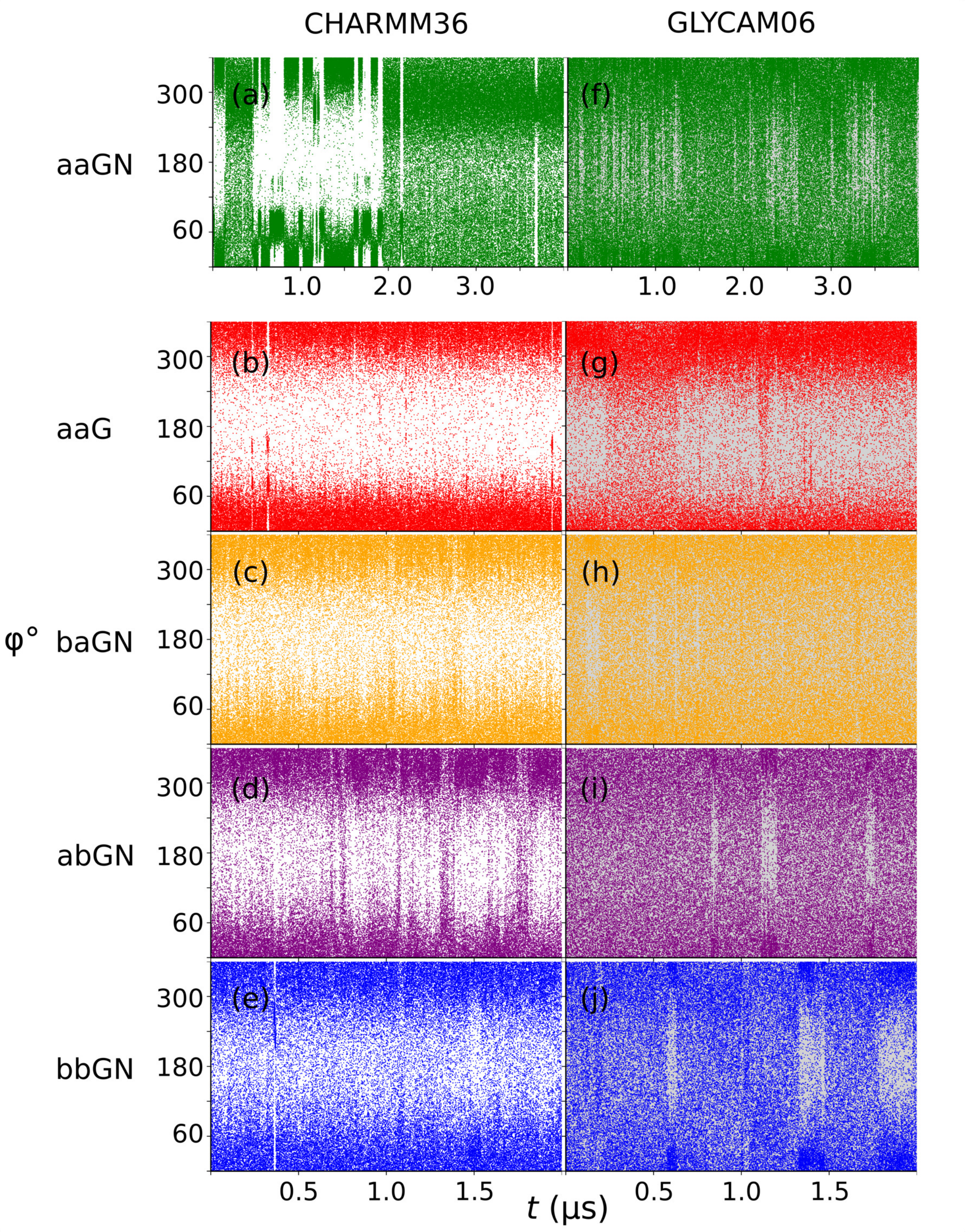


Figure A.5: Cremer-Pople pucker meridian (ϕ) parameters for β-D-Glcp^{III} in aaG and in β-D-Glcp^{III} (e-h) as simulated with CHARMM36 and GLYCAM06 force fields. Simulation time for aaG is 2 μs and 4 μs for aaGN. Note that when θ is at 0° or 180° (i.e. the poles of the Cremer-Pople sphere) there is no direct correlation between ϕ and pucker conformation.

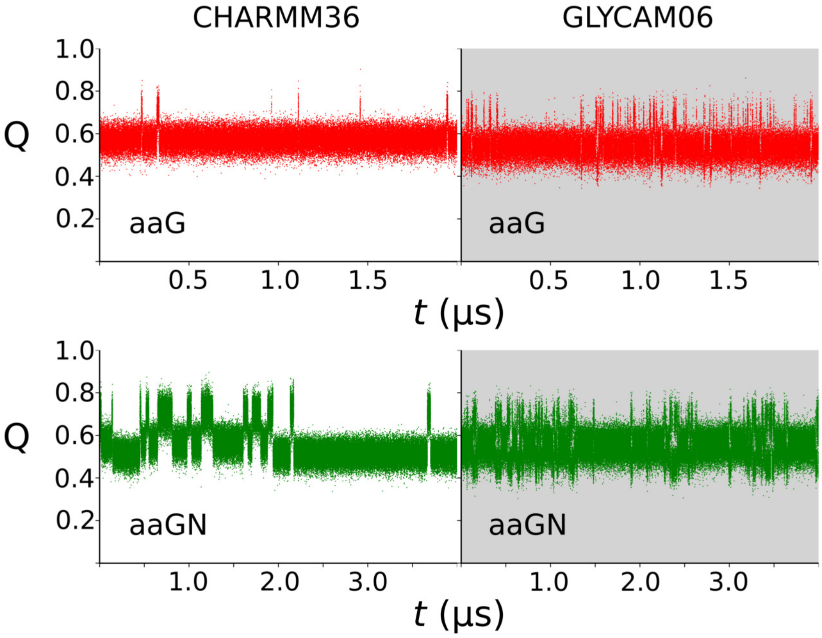


Figure A.6: Cremer-Pople pucker amplitude (*Q*) for *β*-D-Glc in *α*-D-Glc(1→3)[*α*-D-Glc(1→4)]*β*-D-Glc (aaG) and *β*-D-GlcNAC in *α*-D-Glc(1→3)[*α*-D-Glc(1→4)]*β*-D-Glc*p*NAc (aaGN) as simulated with CHARMM36 and GLYCAM06 force fields. Simulation time is 2 *µ*s for aaG and 4 *µ*s for aaGN.

A.2.4 Glycosidic linkage torsion angles and Glc*p*$^{\mathrm{III}}$/Glc*p*NAc pucker conformation

The time series plots of the Glc$^{\mathrm{I}}$ (the 1 → 3 linkage) ψ glycosidic torsion angle and the *β*-D-Glc*p*NAc Cremper-Pople azimuth (θ) angle suggest a correlation between the two parameters. Figure A.9 depicts the relationship between these two values in the form a scatter plot illustrated as heat map. It should be noted that there does not appear to be an immediate correlation between the Glc*p*$^{\mathrm{II}}$ (the 1 → 4 linkage) ψ and θ (Figure A.4).

There is no direct relationship between the Glc*p*$^{\mathrm{I}}$ ψ glycosidic linkage torsion angle and the ring pucker θ parameter for aaG (Figure A.9: a, b). When *β*-D-Glc*p*NAc is in $^{4}C_{1}$ ($\theta \approx 0°$) ψ is centred at $+50°$ for CHARMM36 and $+40°$ for the GLYCAM06 simulation. For GLYCAM06 there is a region populated when *β*-D-Glc*p*NAc is in boat/skew ($\theta \approx 90°$) and $^{1}C_{4}$ ($\theta \approx 180°$), although, these are too sparesly populated to draw a definitive correlation.

However, studying first the CHARMM36 of aaGN, when *β*-D-Glc*p*NAc is in $^{4}C_{1}$ the Glc*p*$^{\mathrm{I}}$ linkage is predominantly in $+\psi$ ($\approx +70°$) with a very small fragment in $-\psi$ ($\approx -70°$) (Figure A.9-c). When *β*-D-Glc*p*NAc is in boat/skew the Glc*p*$^{\mathrm{I}}$ linkage populates a small region

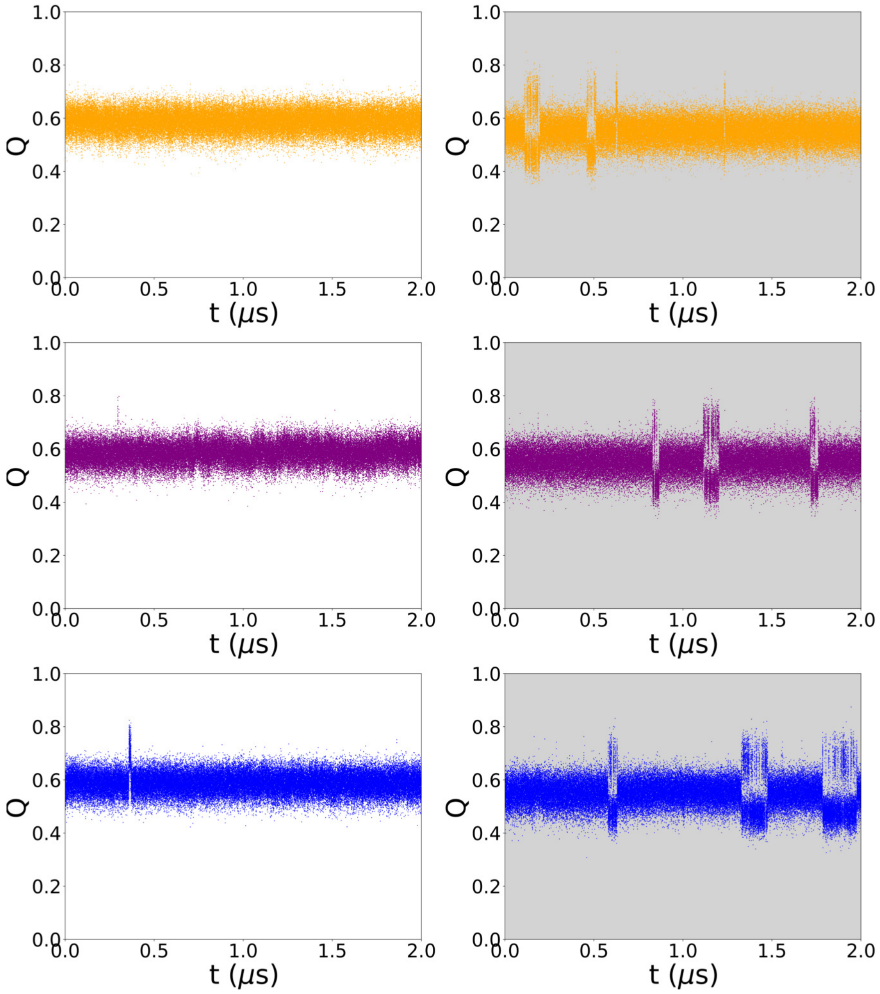


Figure A.7: Cremer-Pople pucker amplitude (Q) for β-D-GlcNAC in β-D-Glc(1→3)[α-D-Glc(1→4)]β-D-Glc*p*NAc (baGN), α-D-Glc(1→3)[β-D-Glc(1→4)]β-D-Glc*p*NAc (abGN), and β-D-Glc(1→3)[β-D-Glc(1→4)]β-D-Glc*p*NAc (bbGN) as simulated with CHARMM36 and GLYCAM06 force fields.

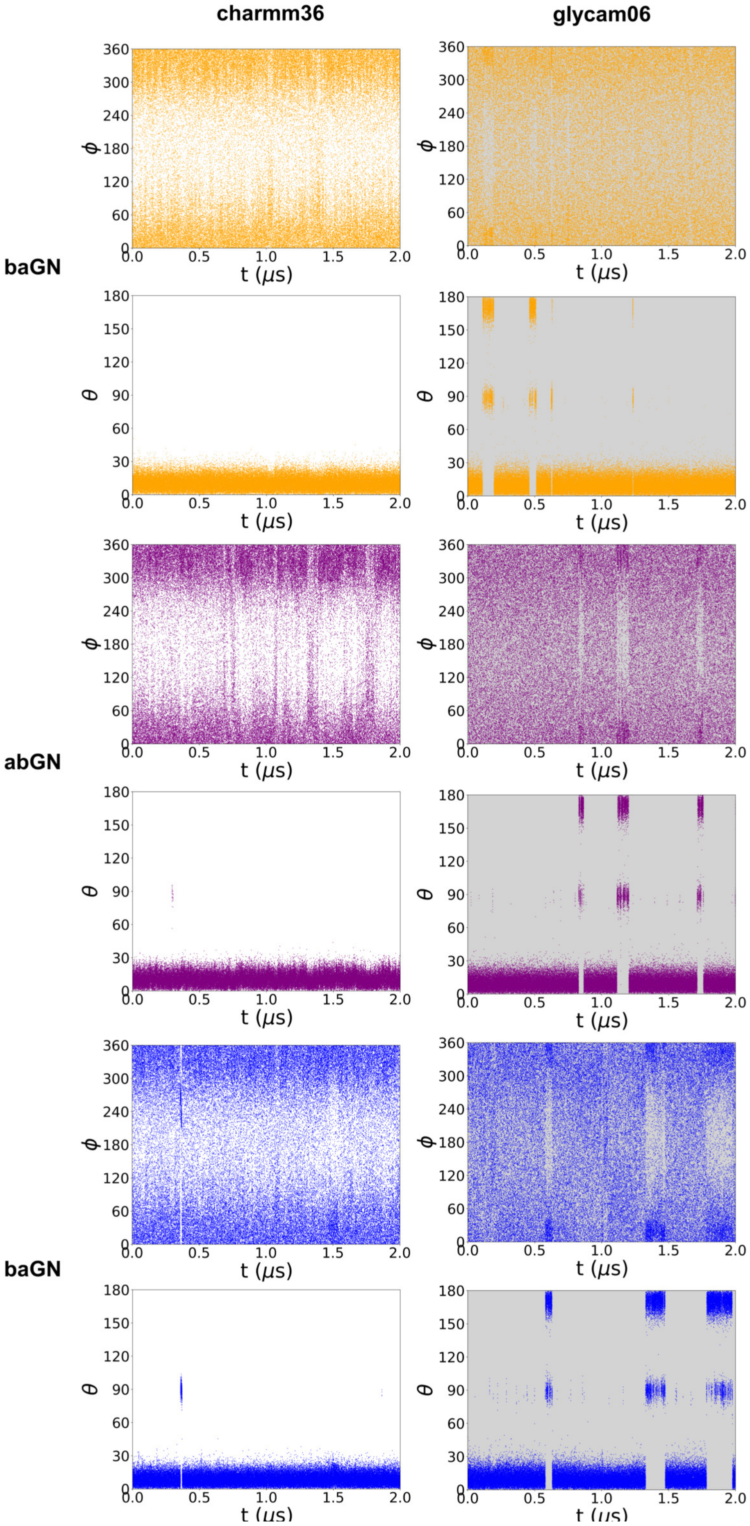


Figure A.8: Cremer-Pople pucker meridian (ϕ) and azimuth (θ) parameters for β-D-Glc*p*NAc in baGN, abGN and bbGN as simulated with the (a) CHARMM36 and (b) GLYCAM06 force fields. Note that when ϕ is at or close to 0° or 180° (i.e. the poles of the Cremer-Pople sphere) there is no direct correlation between ϕ and pucker conformation.

Table A.2: Free energies of the β-D-Glc$^{\mathrm{III}}$ (in aaG) and β-D-Glc*p*NAc (in aaGN, baGN, abGN, and bbGN) ring pucker as calculated from simulated equilibrium data of the trisaccharides for the CHARMM36 and GLYCAM06 force fields computed using the standard $\Delta G = -RTln|K_{eq}|$ relationship. The global minimum energy values corresponding to a defined pucker conformer is shown in **blue**. Energies of less than 1 kcal mol^{-1} are shown in **green**. The minimum energy ring pucker conformer for β-D-Glc$^{\mathrm{III}}$/β-D-Glc*p*NAc for all but one trisaccharide is $^4C_1$ - in aaGN it is $^5S_1$. The '-' symbol indicates the energy for the pucker state could not be directly calculated due to an insufficient number of conformational samples. Only chair, boat and skew conformers are listed.

Pucker Conformer	**ΔG CHARMM36/GLYCAM06 (kcal mol^{-1})**				
	aaG	**aaGN**	**baGN**	**abGN**	**bbGN**
$^4C_1$	**0.00** / **0.00**	+0.54 / **0.00**	**0.00** / **0.00**	**0.00** / **0.00**	**0.00** / **0.00**
$^{3,O}B$	- / **+0.95**	+2.04 / **+0.22**	- / **+0.10**	- / **+0.87**	- / **+0.74**
$^3S_1$	-/ +1.20	**+0.85** / **+0.21**	- / +1.10	- / **+0.96**	+1.70 / **+0.57**
$B_{1,4}$	+1.84 / +1.86	**+0.20** / +1.65	- / -	- / -	+1.94 / +1.63
$^5S_1$	+1.43 / -	**+0.12** / +1.89	- / -	- / -	+1.70 / -
$^{2,5}B$	+1.84 / -	+2.04 / -	- / -	- / -	- / -
^2S_O	- / -	+1.73 / +1.48	- / -	- / -	- / -
$B_{3,O}$	- / +1.86	- / +1.65	- / -	- / -	- / -
$^1S_3$	- / +1.86	+2.04 / +1.65	- / -	- / -	- / -
$^{1,4}B$	- / +1.61	+2.04 / +1.48	- / -	- / -	- / -
$^1S_5$	- / +1.31	+0.85 / +1.48	- / +1.65	- / +1.61	- / -
$B_{2,5}$	- / +1.31	+1.31 / +1.65	- / -	- / -	- / -
$^OS_2$	- / **+0.61**	+2.04 / +1.65	- / -	- / **+0.96**	- / -
$^1C_4$	- / +1.44	**+0.31**/ **+0.55**	- / -	- / **+0.87**	- / **+0.76**

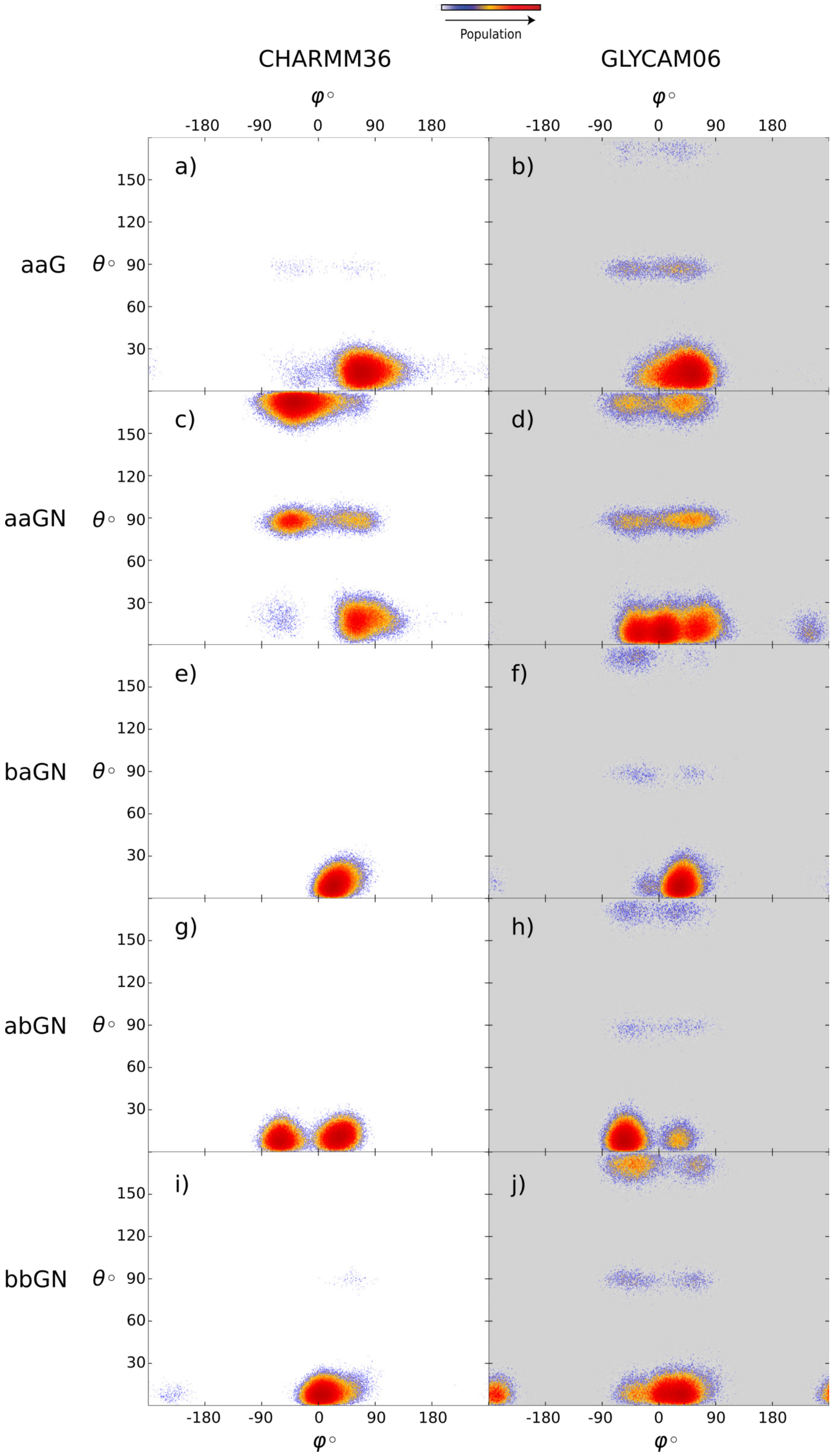


Figure A.9: Scatter plots for the **Glcp^{I}** glycosidic torsion angle (ψ) against Cremer-Pople azimuth (θ) angle for Glc*p*NAc in aaG, aaGN, baGN, abGN and bbGN. Modelled with CHARMM36 (white background) and GLYCAM06 (grey background). The colour scale on the top indicates the relative occupancy of θ and ψ

at $+\psi$ ($\approx +70°$) and a more densely populated area at $-\psi$ ($\approx -30°$). Moreover, when β-D-Glc*p*NAc is in $^1C_4$ there is a single large population centre at $-\psi$ ($\approx -30°$). This indicates a shift from $+\psi$ for $^4C_1$ for $-\psi$ to $^1C_4$ for the Glcp^{I} linkage in the CHARMM36 simulation. In the GLYCAM06 simulation both the Glcp^{I} linkage rotates between $+\psi$ and $-\psi$ when β-D-Glc*p*NAc is in $^4C_1$ and this continues to occur when it is in boat/skew as well as $^1C_4$.

The degree of puckering of β-D-Glc*p*NAc in baGN, baGN and bbGN (Figure A.9: e-j) is minimal and there is no apparent relationship between glycosidic linkage torsion angle, ψ, and the ring pucker θ parameter in these trisaccharides. However, the correlation between ψ and θ in the CHARMM36 simulation of aaGN is clear. Nonetheless, in this case it is difficult to identify whether the change in the torsion angle preceeds the change in pucker or vice-versa and whether a change in one causes, in part or in whole, a change of the other.

A.2.5 Effect of the 1→3 and 1→4 glycosidic linkage starting torsion angle on *β*-D-Glc*p*NAc ring pucker in aaGN

In the initial stages of this study we trialled several different combinations of torsion angles for the 1→3 and 1→4 glycosidic linkages in aaGN. These structures were built with CarbBuilderII[228,255] and based on known preferred torsion angles for the *α*-D-Glc(1→3)*β*-D-Glc*p*NAc and *α*-D-Glc(1→4)*β*-D-Glc*p*NAc linkages.

The purpose of this analysis was to investigate whether the starting torsion angle of the glycosidic linkages played a significant role in the conformations adopted by aaGN over the course of the simulations. In particular, we focused on the *β*-D-Glc*p*NAc ring puckering i.e was there puckering present and if so to what extent does it occur.

Furthermore, a potential correlation between the glycosidic linkage starting orientations and *β*-D-Glc*p*NAc is tentatively suggested by the *Sf* 7a and *Sf* 7b O-antigen studies (Chapter 7) where *β*-D-Glc*p*NAc consistently remains in $^4C_1$ when the *α*-L-Rha(1→3)*β*-D-Glc*p*NAc linkage adopts an anti-orientation.

Simulations were run for between 50 ns and 200 ns with the CHARMM36 force field and were stopped if there was more than 20 ns of continuous *β*-3scD-Glc*p*NAc puckering. Table A.3 provides the list of the starting glycosidic linkage torsions trialled and the simulation times for each.

In short, the initial starting orientations of the 1→3 and 1→4 glycosidic linkages had no impact on *β*-D-Glc*p*NAc ring puckering or, additionally, the torsion angles eventually adopted by the 1→3 and 1→4 linkages. In each of the simulations *β*-D-Glc*p*NAc ring would pucker into the conformations observed in the aaGN simulations and the glycosidic torsion angles would do the same.

Table A.3: Starting torsion angles for trial simulations of aaGN

αDGlc(1→3)βDGlcNAc	αDGlc(1→4)βDGlcNAc	Simulation time (ns)
$\phi, \psi = +3, +10$	$\phi, \psi = -31, +45$	100
$\phi, \psi = +20, +24$	$\phi, \psi = +44, -70$	200
$\phi, \psi = +28, +31$	$\phi, \psi = +24, +13$	200
$\phi, \psi = +40, -36$	$\phi, \psi = +32, +62$	100
$\phi, \psi = +60, +40$	$\phi, \psi = +40, -30$	150
$\phi, \psi = +2, -6$	$\phi, \psi = +40, -32$	50
$\phi, \psi = -70, -70$	$\phi, \psi = +8, +64$	150

A.2.6 Hydroxyl, hydroxymethyl and N-Acetyl rotamer populations

Orientation of the hydroxyl, hydroxymethyl and N-Acetyl groups of sugars can play an important role in their interactions with other molecules,[298] such as through hydrogen bonding with water molecules or enzyme receptor sites. We provide here a detailed breakdown of the percentage distributions for the rotamer populations of these groups in the trisaccharides simulated, for both the CHARMM36 and GLYCAM06 force fields (Table A.4).

Across all GLYCAM06 simulated structures the trans orientation of the NHAc residue is preferred. In the CHARMM36 simulations the syn orientation is more prevalent (≈75%) with the exception of aaGN, where NHAc is neither in a distinct cis or trans state for the majority of time. (Figure A.10). This again is likely due to the constant puckering of the β-D-Glc*p*NAc ring in aaGN.

There is a similar distinction for the O-1 and O-6 rotamer populations between the two force fields. For both O-1 and O-6 the $g+$ and $g-$ are preferred for CHARMM36 while t is more prevalent for GLYCAM06.

The hydroxymethyl group adopts different rotamers for aaG and aaGN compared to the other trisaccharides for the CHARMM36 simulations. The *tg* and *gt* rotamers are more prevalent in aaG and aaGN while the *gt* and *gg* are more populated in the other trisaccharides. In the GLYCAM06 simulations gg is the single dominant orientation.

Table A.4: Rotamer populations of hydroxyl, hydroxymethyl and N-Acetyl groups on β-D-Glc*p*NAc from the CHARMM36 and GLYCAM06 simulations. The GLYCAM06 values are in bold and italicised.

Rotamer		**Population (%)**				
		aaG	aaGN	baGN	abGN	bbGN
O-1	**t**	14, ***33***	17, ***35***	16, ***33***	16, ***36***	16, ***38***
	g^+	39, ***3***	21, ***4***	20, ***4***	22, ***3***	21, ***3***
	g^-	23, ***11***	34, ***10***	34, ***11***	33, ***10***	34, ***10***
O-2	**t**	1, ***8***	-	-	-	-
	g^+	3, ***2***	-	-	-	-
	g^-	43, ***57***	-	-	-	-
-NHAc	**cis**	-	32, ***0***	80, ***0***	74, ***0***	76, ***0***
	trans	-	9, ***61***	7, ***75***	8, ***67***	9, ***67***
O-6	**t**	8, ***33***	71, ***30***	7, ***37***	8, ***30***	7, ***27***
	g^+	25, ***18***	14, ***17***	27, ***19***	37, ***25***	34, ***20***
	g^-	21, ***9***	24, ***11***	21, ***6***	12, ***5***	15, ***10***
-CH$_2$OH	**tg**	48, ***1***	28, ***2***	10, ***2***	2, ***2***	2, ***5***
	gt	41, ***26***	32, ***25***	50, ***19***	31, ***15***	35, ***38***
	gg	2, ***61***	18, ***58***	40, ***70***	61, ***7***	55, ***55***

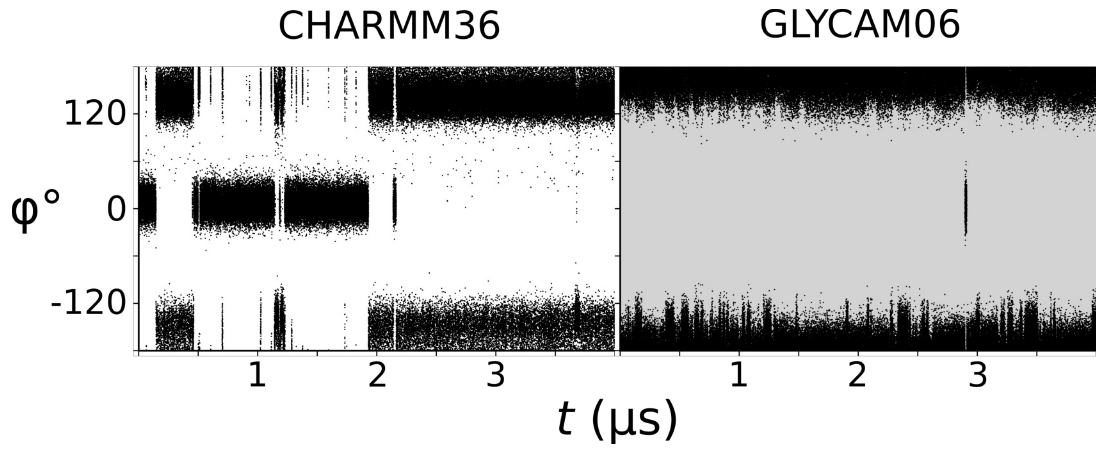


Figure A.10: NHAc torsion angle (H_2 - C_2 - N_2 - HN) in aaGN for the CHARMM36 and GLYCAM06 simulations

A.3 β-D-Glc*p*^{III}/β-D-Glc*p*NAc ring energies

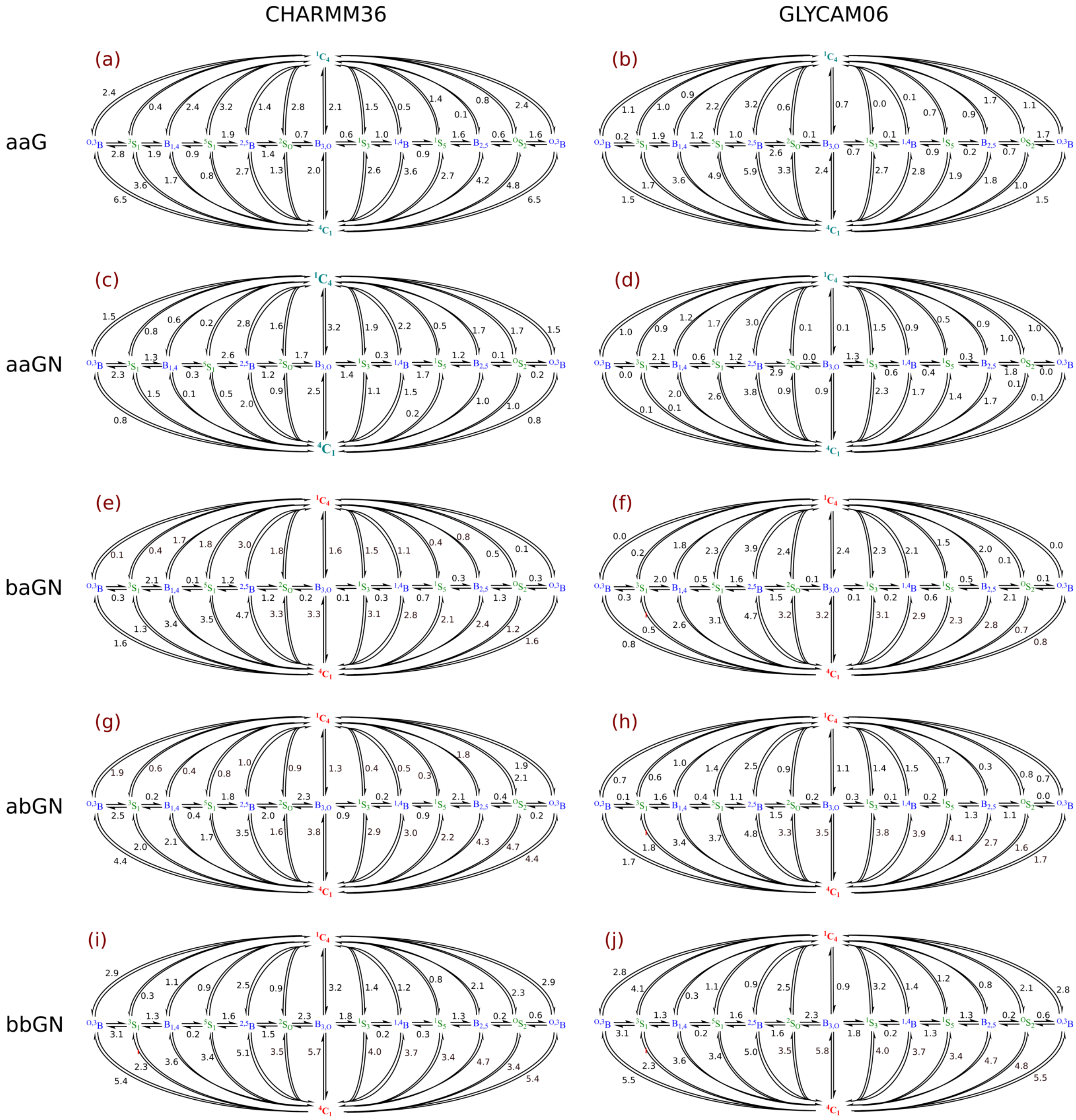


Figure A.11: Interconversion pathways as calculated from potential of mean force calculations for the ring pucker of *β*-D-Glc^{III} in aaG and *β*-D-Glc*p*NAc in aaGN, baGN, abGN, and bbGN. The energies between different pucker states are smaller for aaGN compared to baGN, abGN, and bbGN. Comparing force fields, the energy differences between different pucker states are greater in the CHARMM36 simulations than the GLYCAM06 simulations. Only chair and boat/skew conformers are shown.

A.4 Distance restrained molecular modelling of *S. flexneri* 7a and 7b O-Ags

MD simulations performed on 6RU of *Sf* 7a and 7b were performed for 1 μs length using the CHARMM26 force field, with a 1 kcal distance restraint applied to the $C1 - C3$, $C1 - C5$, and $C3 - C5$ atoms of the β-D-Glc*p*NAc residues for deviations away from $^4C_1$. All other simulation parameters are the same as the unrestrained *Sf* 7a and 7b simulations.

A.4.1 Glycosidic linkage torsion angles

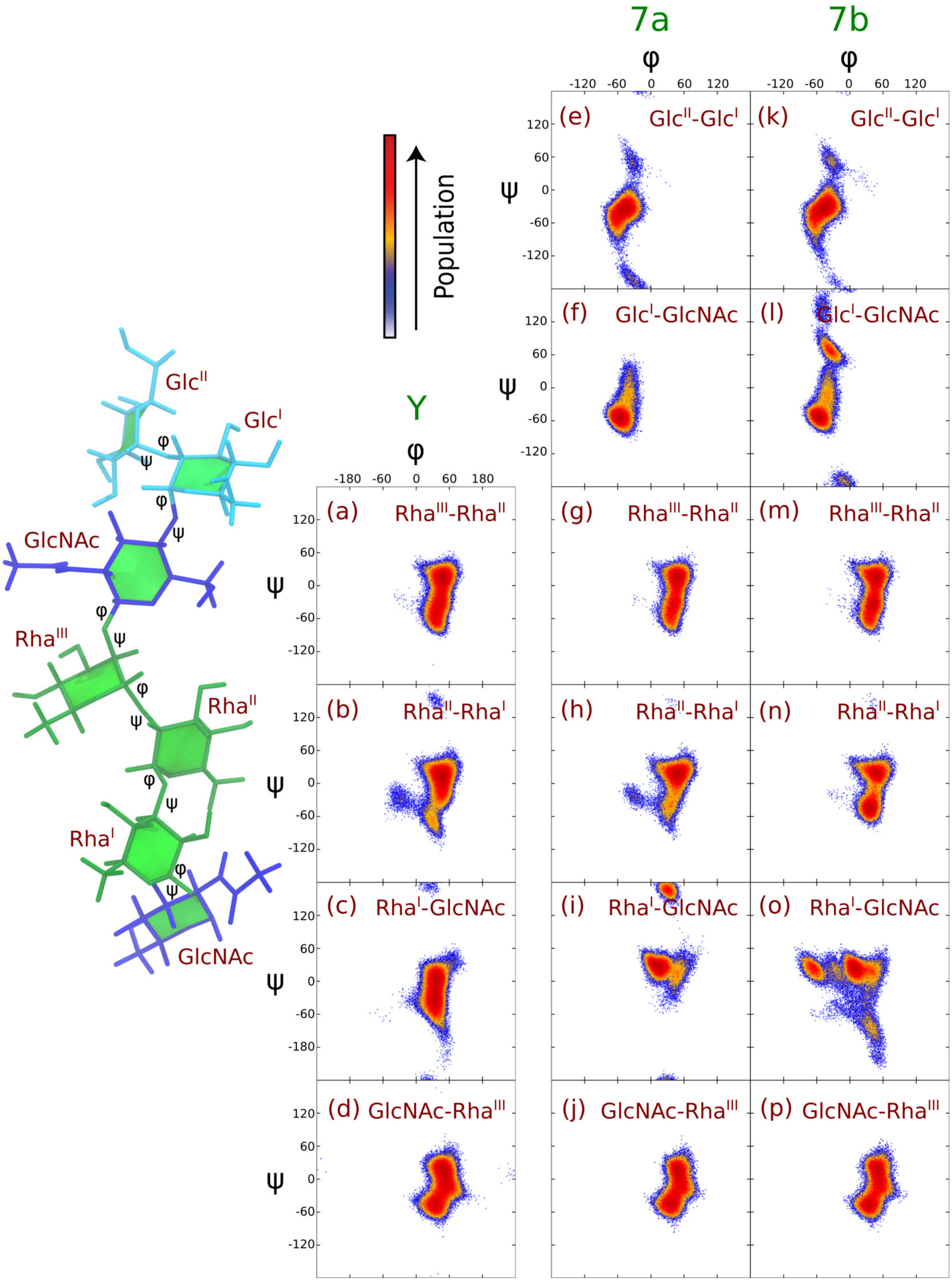


Figure A.12: Scatter plots illustrated as heat maps for the glycosidic linkage torsion angles (ϕ, ψ) in the four central repeating units (RU2-RU5) of the modelled O-Ag of *Sf* Y (a-d), 7a (e-j), and 7b (k-p). ϕ, ψ defined as H_1 - C_1 - O_1 - C'_x and C_1 - O_1 - C'_x - H'_x respectively. The heat maps combine the points from the four central repeating units to broadly sample backbone behavior. The color scale on the upper-left indicates the relative occupancy of the torsion angles during the simulations. The structure on the left depicts the O-Ag residues and glycosidic linkages. Distance restraints are applied to the β-D-Glc*p*NAc rings to *Sf* 7a and 7b only for deviations away from $^4C_1$.

A.4.2 β-D-GlcNAc ring pucker

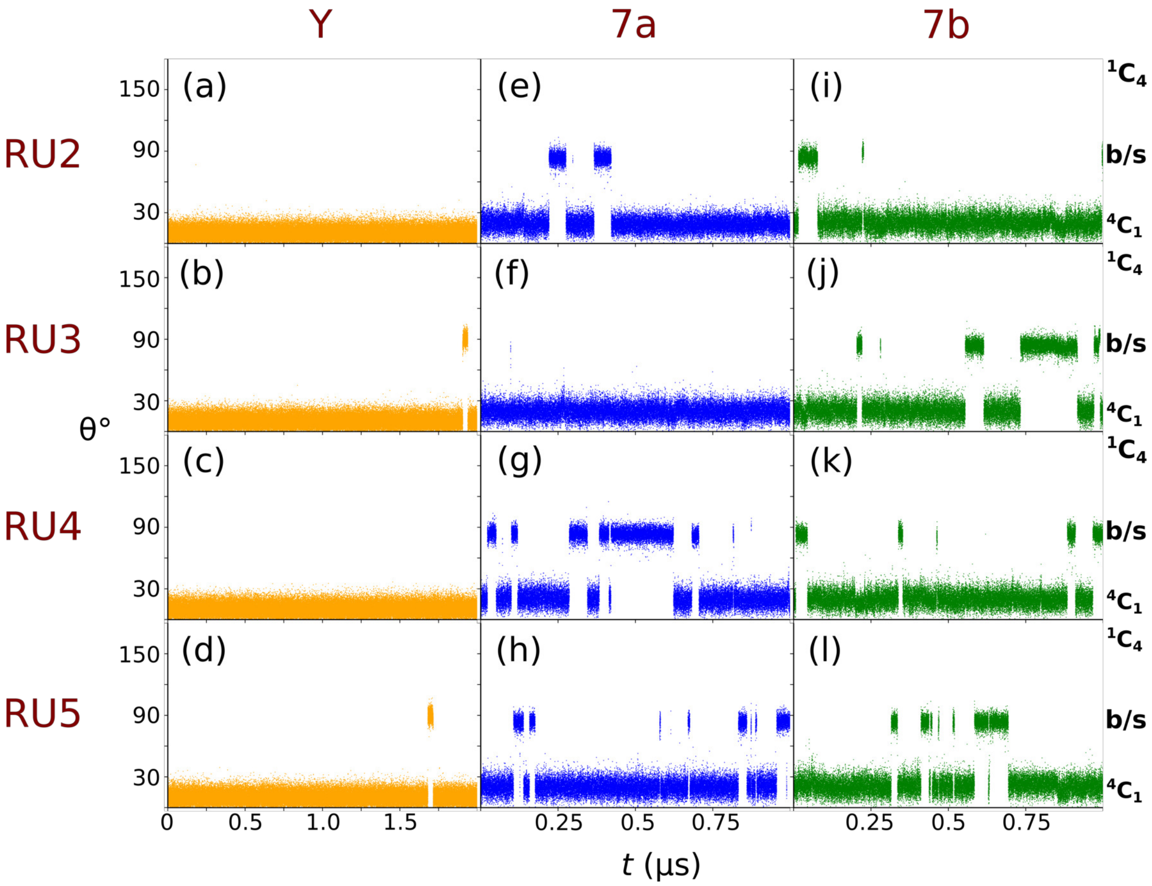


Figure A.13: Time series plots for the β-D-Glc*p*NAc Cremer-Pople ring pucker θ paramter in the four-central repeating units of the modelled O-Ags polysaccharide of *Sf* Y, 7a and 7b. Relationship between θ and ring pucker conformation: $^4C_1$ ($\theta \approx 0°$), boat/skew (b/s, $\theta \approx 90°$), $^1C_4$ ($\theta \approx 180°$). In all four central repeating units of *Sf* Y β-D-Glc*p*NAc remains entirely in $^4C_1$, with the exception of a short transition into boat/skew conformers. β-D-Glc*p*NAc in both *Sf* 7a and 7b on average adopt boat/skew states for approximately 16% of the simulation time. The $^1C_4$ state is not occupied for any period. Distance restraints are applied to the β-D-Glc*p*NAc rings to *Sf* 7a and 7b only for deviations away from $^4C_1$.

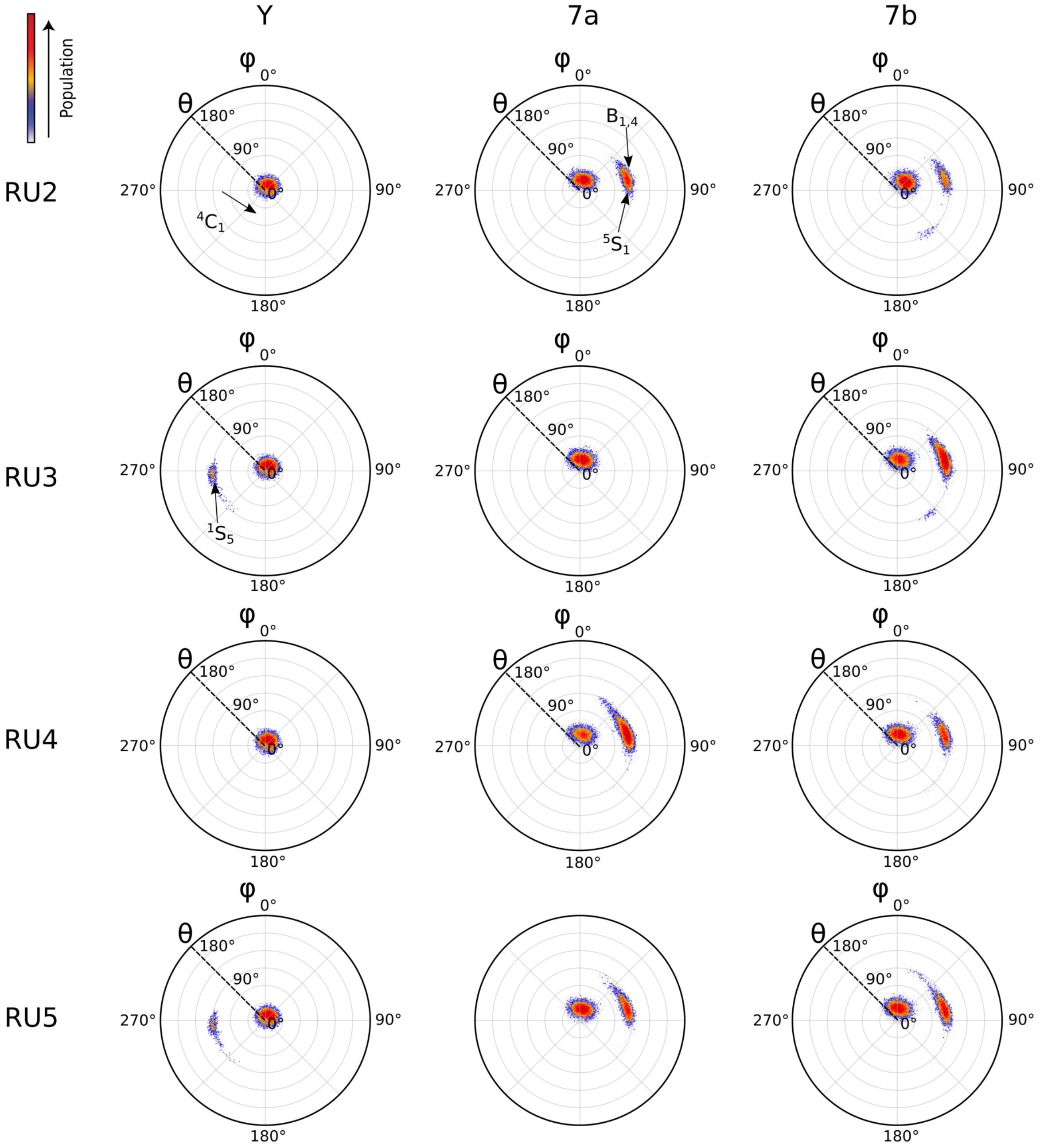


Figure A.14: Polar heatmap plots of the Cremer Pople ϕ and θ parameters for β-D-Glc*p*NAc in the four-central repeating units of *Sf* Y, 7a and 7b.

Table A.5: Calculated coupling constants for *β*-D-Glc*p* in aaG and *β*-D-Glc*p*NAc in aaGN, baGN, abGN, and bbGN. Coupling constants calculated using an approximation of the Haasnoot-Arnot equation.[299] Observed and calculated coupling constants for D-Glc*p*NAc monosacharide by Sattelle et al. also provided.[63]

Coupling	**CHARMM36**					**GLYCAM06**					**GlcNAc**	
	aaG	aaGN	baGN	abGN	bbGN	aaG	aaGN	baGN	abGN	bbGN	NMR	MD
$J_{1,2}$	7.83	4.33	8.76	8.68	8.74	6.37	6.76	7.81	7.71	6.95	3.5	3.2
$J_{2,3}$	8.54	3.99	10.11	10.02	10.23	6.87	7.62	9.01	8.99	8.15	10.7	10.1
$J_{3,4}$	7.96	3.36	8.79	8.75	8.97	7.35	7.16	7.98	8.07	7.57	9.9	9.9
$J_{4,5}$	9.51	2.87	9.73	9.75	9.76	9.18	7.77	8.97	8.91	7.84	9.3	10.0
$J_{5,6\ \mathrm{proR}}$	7.09	4.56	7.12	8.14	7.96	8.48	8.07	8.84	9.13	7.92	5.2	4.5
$J_{5,6\ \mathrm{proS}}$	2.18	5.43	2.17	2.24	2.25	2.14	2.54	2.27	2.10	2.52	2.3	3.2

Appendix B: NMR

B.1 NMR spectra

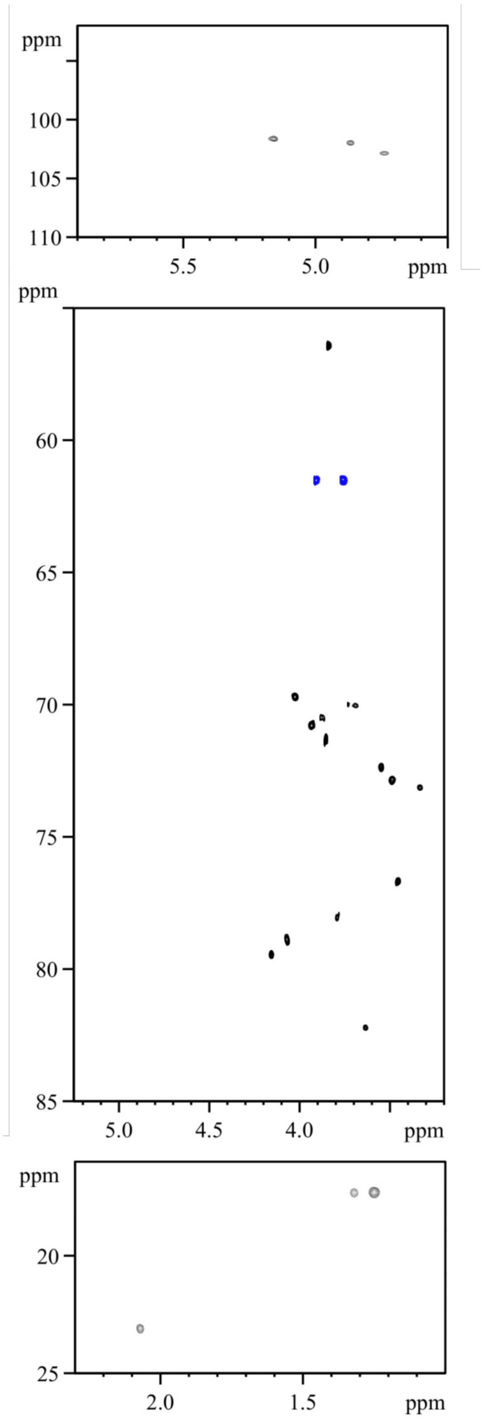
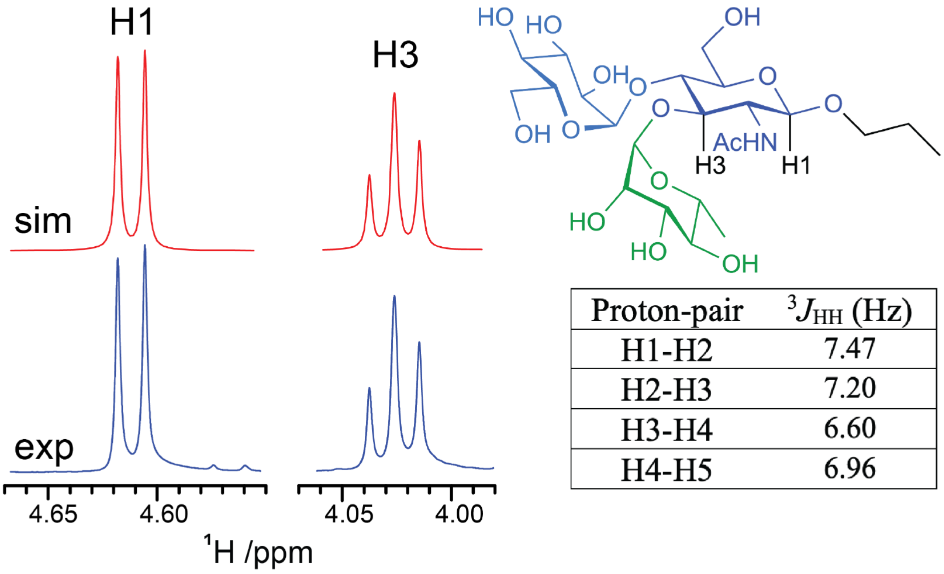


Figure B.1: HSQC of *Sf* 7a O-Ag.

B.2 ${}^{3}J_{H,H}$ Coupling measurements



Proton-pair	${}^{3}J_{HH}$ (Hz)
H1-H2	7.47
H2-H3	7.20
H3-H4	6.60
H4-H5	6.96

Figure B.2: Selected regions of a PERCH-simulated (top-left) and an experimental (bottom-left) ^{1}H NMR spectrum of the vicinal branched trisaccharide: α-L-Rha*p*-(1→3)[α-D-Glc*p*-(1→4)]-β-D-Glc*p*NAcOPr (top-right) and related ${}^{3}J_{H,H}$ couplings (table). The measurements demonstrate the differences in the ${}^{3}J_{H,H}$ couplings between R3GN and GN.[286]

B.3 nOe spectroscopy studies

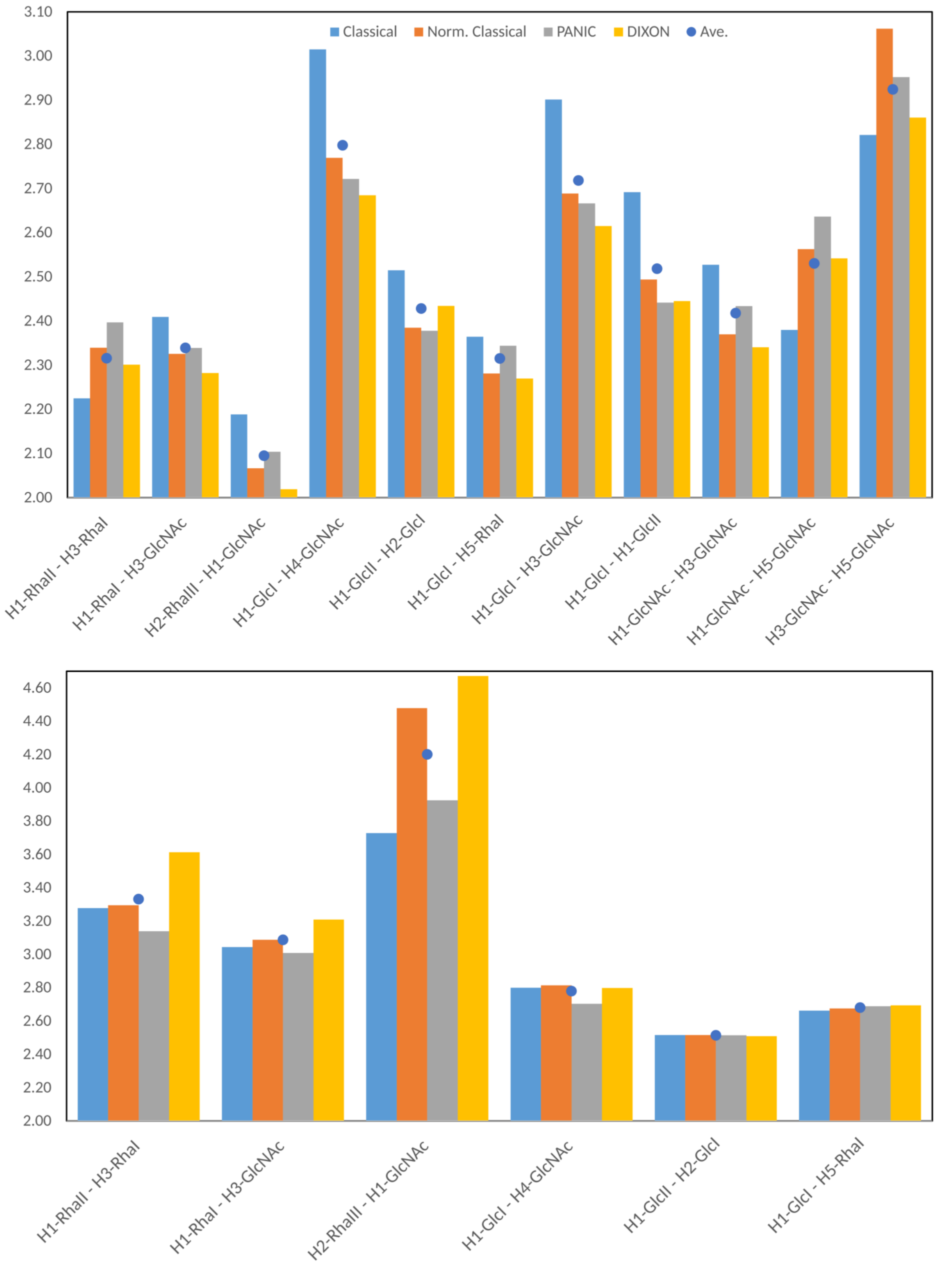


Figure B.3: Trans-glycosidic, inter-residue, and intra-residue (top), and Methyl-methine (bottom) ^{1}H-^{1}H distances derived from ^{1}H,^{1}H-NOESY NMR buildup curves analysed using the Classical, Normalised Classical, PANIC and DIXON methods for *Sf* 7a. The average of the methods for each proton pair is shown as dark blue dots on the graph.

Table B.1: ^{1}H-^{1}H cross-relaxation rates (σ in s^{-1}), respective distances (in Å), and associated statistical measures from the O-Ag of *S. flexneri* 7a derived from ^{1}H,^{1}H-NOESY NMR buildup curves and analysed by the Classical, Normalised Classical, PANIC, and DIXON approaches subdivided into trans-glycosidic, inter-residual, intra-residual, methyl-methine and reference proton pairs. Methine-methine reference distances: H1-Glcp^{I} — H2-Glcp^{I} (2.45 Å), H1-Glcp^{II} — H2-Glcp^{II} (2.49 Å), H1-Rhap^{III} — H2 Rhap^{III} (2.47 Å). Methyl-methine reference distance H5-Rhap^{II} — Me-Rhap^{II} (2.60 Å).

Proton Pair		σ_C^*	σ_{NC}	σ_P	σ_D	r_C	r_{NC}	r_P	r_D	$r_{ave.}$	r_{range}	$r_{std.dev.}$	r_{SE}	r_{cv}
Trans-glycosidic														
H1-Rhap^{II}	H3-Rhap^{I}	1.27	0.96	0.76	0.72	2.23	2.34	2.40	2.31	2.32	0.17	0.06	0.03	0.03
H1-Rhap^{I}	H3-Glc$p$NAc	0.79	0.99	0.88	0.75	2.41	2.33	2.34	2.29	2.34	0.13	0.05	0.02	0.02
H2-Rhap^{III}	H1-Glc$p$NAc	1.40	2.01	1.67	1.57	2.19	2.07	2.11	2.02	2.10	0.17	0.06	0.03	0.03
H1-Glcp^{I}	H4-Glc$p$NAc	0.21	0.35	0.36	0.28	3.02	2.78	2.73	2.69	2.80	0.33	0.13	0.06	0.05
H1-Glcp^{II}	H2-Glcp^{I}	0.61	0.85	0.80	0.51	2.52	2.39	2.38	2.44	2.43	0.14	0.05	0.03	0.02
Inter-residue														
H1-Glcp^{I}	H5-Rha$^{\mathrm{I}}$	0.88	1.11	0.87	0.78	2.37	2.29	2.35	2.27	2.32	0.09	0.04	0.02	0.02
H1-Glcp^{I}	H3-Glc$p$NAc	0.26	0.42	0.40	0.33	2.91	2.69	2.67	2.62	2.72	0.29	0.11	0.05	0.04
H1-Glcp^{I}	H1-Glcp^{II}	0.40	0.65	0.68	0.50	2.70	2.50	2.45	2.45	2.52	0.25	0.10	0.05	0.04

[a] Calculated with $r_{ij} = r_{ref} \cdot (\sigma_{ref}/\sigma_{ij})^{1/6}$ with reference proton distances derived from 1 μs, of MD simulation of the six-repeating unit polysaccharide in solution.

[b] Reference distance r_{ref} derived from MD simulations were calculated according to $r_{ref} = \langle r_{ij}^6 \rangle^{-1/6}$.

[*] in ten-billions

Table B.1 Continued:

Proton Pair		σ_C^*	σ_{NC}	σ_P	σ_D	r_C	r_{NC}	r_P	r_D	$r_{ave.}$	r_{range}	$r_{std.dev.}$	r_{SE}	r_{CV}
Intra-residue														
H1-Glc*p*NAc	H3-Glc*p*NAc	0.59	0.89	0.70	0.65	2.53	2.37	2.44	2.35	2.42	0.19	0.07	0.04	0.03
H1-Glc*p*NAc	H5-Glc*p*NAc	0.85	0.55	0.43	0.39	2.38	2.57	2.64	2.54	2.53	0.26	0.09	0.05	0.04
H3-Glc*p*NAc	H5-Glc*p*NAc	0.31	0.19	0.22	0.19	2.83	3.07	2.96	2.87	2.93	0.24	0.09	0.05	0.03
Methyl-methine														
Me-Rhap^{I}	H1-Glcp^{II}	0.28	0.40	0.38	0.20	3.28	3.30	3.14	3.61	3.33	0.47	0.17	0.09	0.05
Me-Rhap^{I}	H1-Glcp^{I}	0.18	0.27	0.29	0.10	3.04	3.09	3.01	3.21	3.09	0.20	0.08	0.04	0.02
Me-Rhap^{II}	H2-Rhap^{I}	0.82	0.04	0.08	0.02	3.73	4.48	3.93	4.67	4.20	0.94	0.39	0.19	0.09
Me-Rhap^{I}	H4-Rhap^{I}	0.46	0.70	0.72	0.45	2.80	2.82	2.70	2.80	2.78	0.11	0.04	0.02	0.02
Me-Rhap^{II}	H4-Rhap^{II}	0.87	1.38	1.12	0.86	2.52	2.52	2.51	2.51	2.51	0.01	0.00	0.00	0.00
Me-Rhap^{III}	H4-Rhap^{III}	0.62	0.96	0.75	0.56	2.66	2.68	2.69	2.69	2.68	0.03	0.01	0.01	0.00
Reference														
H1-Glcp^{I}	H2-Glcp^{I}	0.54	0.77	0.79	0.55	-	-	-	-	-	-	-	-	-
H1-Glcp^{II}	H2-Glcp^{II}	0.42	0.59	0.60	0.40	-	-	-	-	-	-	-	-	-
H1-Rhap^{III}	H2-Rhap^{III}	0.84	0.85	0.65	0.54	-	-	-	-	-	-	-	-	-
H5-Rhap^{II}	Me-Rhap^{II}	0.71	0.59	0.6	0.4	-	-	-	-	-	-	-	-	-

Table B.2: ^{1}H, ^{1}H cross-relaxation rates (σ) with associated standard errors (SE), standard error of slope (SES), coefficient of determination (r^2), calculated effective ^{1}H, ^{1}H distances (r_{exp} in Å) from NMR nOe experiments of *S. flexneri* 7a O-Ag. Calculated from ^{1}H,^{1}H-NOESY NMR buildup curves using the PANIC approach. Proton pairs are subdivided into trans-glycosidic, inter-residual, intra-residual, methyl-methine, proton pairs.

Proton Pair		σ	σ SE	σ SES	σR^2	r_{exp}	r_{exp} SE
Trans-glycosidic							
H1-Rha^{II}	H3-Rha^{I}	0.763	0.004	0.016	0.997	2.393	0.029
H1-Rha^{I}	H3-Glc*p*NAc	0.884	0.004	0.020	0.997	2.335	0.029
H1-Glc*p*NAc	H2-Rha^{III}	1.669	0.008	0.019	0.999	2.101	0.026
H1-Glc*p*^{I}	H4-Glc*p*NAc	0.356	0.002	0.015	0.990	2.718	0.035
H1-Glc*p*^{II}	H2-Glc*p*^{I}	0.800	0.004	0.011	0.999	2.375	0.030
Inter-residue							
H1-Glc*p*^{I}	H5-Rha^{I}	0.873	0.004	0.025	0.995	2.341	0.028
H1-Glc*p*^{I}	H3-Glc*p*NAc	0.403	0.002	0.012	0.994	2.662	0.034
H1-Glc*p*^{I}	H1-Glc*p*^{II}	0.683	0.004	0.030	0.988	2.438	0.031
Intra-residue							
H1-Glc*p*NAc	H3-Glc*p*NAc	0.697	0.003	0.015	0.997	2.430	0.030
H1-Glc*p*NAc	H5-Glc*p*NAc	0.431	0.002	0.011	0.996	2.633	0.032
H3-Glc*p*NAc	H5-Glc*p*NAc	0.219	0.001	0.009	0.99	2.948	0.038
Methyl-methine							
Me-Rha^{I}	H1-Glc*p*^{II}	0.295	0.002	0.022	0.969	3.140	0.023
Me-Rha^{I}	H1-Glc*p*^{I}	0.381	0.002	0.008	0.997	3.009	0.021
Me-Rha^{II}	H2-Rha^{I}	0.077	0.000	0.005	0.98	3.926	0.030
Me-Rha*p*^{I}	H4-Rha*p*^{I}	0.723	0.004	0.040	0.982	2.703	0.020
Me-Rha*p*^{II}	H4-Rha*p*^{II}	1.117	0.005	0.022	0.998	2.515	0.017
Me-Rha*p*^{III}	H4-Rha*p*^{III}	0.746	0.004	0.017	0.997	2.689	0.019

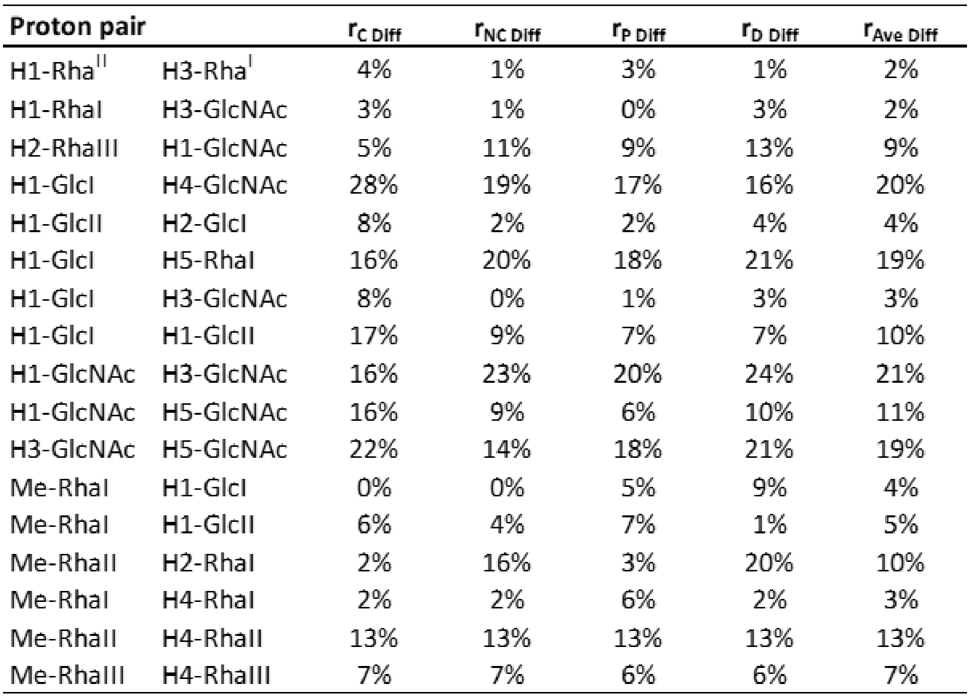
Table B.3 lists the experimental and modelled ^{1}H, ^{1}H distances for the MD simulations of *Sf* 7a. The results show that the difference between the experimental and modelled distances are significantly smaller when a restraint is applied to *β*-D-Glc*p*NAc for deviations away from $^4C_1$. In particular, there is very good agreement between the experimental and modelled H1-Glc*p*$^{\mathrm{I}}$ — H4-Glc*p*NAc and H1-Glc*p*$^{\mathrm{I}}$ — H5-Rha$^{\mathrm{I}}$ distances, which have very large differences in the unrestrained simulations.

Table B.3: ^{1}H, ^{1}H distances calculated from MD simulations (r_{MD}, r_{MDRest}) and NMR nOe experiments (r_{exp} and $r_{expRest}$) in Åfrom the O-Ags of *S. flexneri* 7a. Calculated from ^{1}H,^{1}H-NOESY NMR buildup curves as an average of the normalised classical and PANIC approaches, effective ^{1}H, ^{1}H distances as determined from a modelled 6RU O-Ag of *Sf* 7a with distance restraints applied to the *β*-D-Glc*p*NAc ring for deviations away from $^1C_4$. Computed distances are based on the four inner RUs of the polysaccharide and are subdivided into trans-glycosidic, inter-residual, intra-residual, methyl-methine and reference proton pairs.

Proton Pair		r_{exp}	r_{MD}	% diff.	$r_{expRest.}$	$r_{MDRest.}$	% diff.
Trans-glycosidic							
H1-Rha^{II}	H3-Rha^{I}	2.39	2.32	3%	2.40	2.30	4%
H1-Rha^{I}	H3-Glc*p*NAc	2.34	2.34	0%	2.34	2.23	5%
H2-Rha^{III}	H1-Glc*p*NAc	2.10	2.30	9%	2.11	2.28	8%
H1-Glc*p*^{I}	H4-Glc*p*NAc	2.72	2.29	17%	2.73	2.69	1%
H1-Glc*p*^{II}	H2-Glc*p*^{I}	2.37	2.38	2%	2.38	2.23	7%
Inter-residue							
H1-Glc*p*^{I}	H5-Rha^{I}	2.34	2.79	18%	2.35	2.37	1%
H1-Glc*p*^{I}	H3-Glc*p*NAc	2.66	2.68	1%	2.68	2.67	1%
H1-Glc*p*^{I}	H1-Glc*p*^{II}	2.44	2.28	7%	2.44	2.23	9%
Intra-residue							
H1-Glc*p*NAc	H3-Glc*p*NAc	2.44	2.97	20%	2.44	2.80	14%
H1-Glc*p*NAc	H5-Glc*p*NAc	2.63	2.81	6%	2.64	2.31	13%
H3-Glc*p*NAc	H5-Glc*p*NAc	2.95	3.53	18%	2.96	2.88	2%
Methyl-methine							
Me-Rha^{I}	H1-Glc*p*^{II}	3.14	3.29	5%	3.14	3.20	2%
Me-Rha^{I}	H1-Glc*p*^{I}	3.01	3.23	7%	3.01	3.10	3%
Me-Rha^{II}	H2-Rha^{I}	3.93	3.93	3%	3.93	3.97	1%
Me-Rha*p*^{I}	H4-Rha*p*^{I}	2.70	2.87	6%	2.70	2.88	6%
Me-Rha*p*^{II}	H4-Rha*p*^{II}	2.51	2.86	13%	2.51	2.86	13%
Me-Rha*p*^{III}	H4-Rha*p*^{III}	2.69	2.87	6%	2.68	2.86	6%

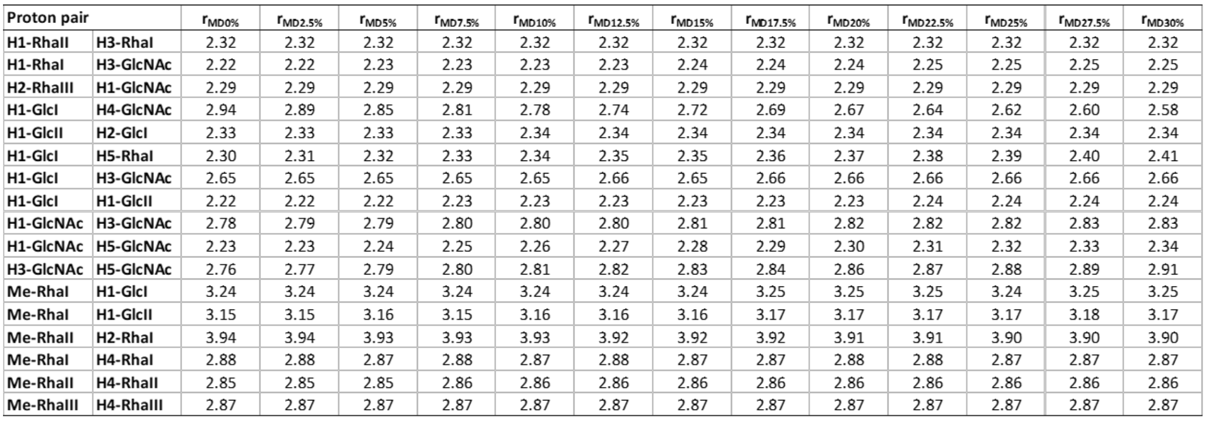
Table B.4: Experimentally derived ^{1}H-^{1}H distances (r_{expt} in Å) for *Sf* 7a calculated by ^{1}H,^{1}H-NOESY NMR buildup curves, effective ^{1}H-^{1}H distances (r_{MD}, $r_{MD0\%}$, and $r_{MD15\%}$ in Å) as determined from the modelled 6RU O-Ag of *Sf* 7a. The effective distance $r_{MD0\%}$ calculated from the entire unrestrained MD trajectory, $r_{MD0\%}$ refers to only the $^{4}C_{1}$ conformation of the Glc*p*NAc residue, while $r_{MD15\%}$ distances were calculated based on an 85:15 ratio of $^{4}C_{1}$ vs. boat/skew conformations. Computed distances are based on the four inner RUs of the 6RU O-Ag and are subdivided into trans-glycosidic, inter-residual, intra-residual, methyl-methine and reference proton pairs. The H1-Glc*p*$^{\mathrm{I}}$ — H4-Glc*p*NAc and H1-Glc*p*NAc — H3-Glc*p*NAc distances are in bold as they have the greatest difference between the experimental and unrestrained MD distances.

Proton Pair		r_{exp}	r_{MD}	$r_{MD0\%}$	$r_{MD15\%}$
Trans-glycosidic					
H1-Rha$^{\mathrm{II}}$	H3-Rha$^{\mathrm{I}}$	2.40	2.32	2.32	2.32
H1-Rha$^{\mathrm{I}}$	H3-Glc*p*NAc	2.34	2.34	2.22	2.24
H2-Rha$^{\mathrm{III}}$	H1-Glc*p*NAc	2.11	2.30	2.29	2.29
H1-Glc*p*$^{\mathrm{I}}$	H4-Glc*p*NAc	**2.73**	**2.29**	2.94	2.72
H1-Glc*p*$^{\mathrm{II}}$	H2-Glc*p*$^{\mathrm{I}}$	2.38	2.33	2.33	2.34
Inter-residue					
H1-Glc*p*$^{\mathrm{I}}$	H5-Rha$^{\mathrm{I}}$	**2.35**	**2.79**	2.30	2.35
H1-Glc*p*$^{\mathrm{I}}$	H3-Glc*p*NAc	2.67	2.68	2.65	2.65
H1-Glc*p*$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{II}}$	2.45	2.28	2.22	2.23
Intra-residue					
H1-Glc*p*NAc	H3-Glc*p*NAc	2.44	2.97	2.78	2.81
H1-Glc*p*NAc	H5-Glc*p*NAc	2.64	2.81	2.23	2.28
H3-Glc*p*NAc	H5-Glc*p*NAc	2.96	3.53	2.76	2.83
Methyl-methine					
Me-Rha$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{II}}$	3.14	3.29	3.24	3.24
Me-Rha$^{\mathrm{I}}$	H1-Glc*p*$^{\mathrm{I}}$	3.01	3.23	3.15	3.16
Me-Rha$^{\mathrm{II}}$	H2-Rha$^{\mathrm{I}}$	3.93	3.82	3.94	3.92
Me-Rha*p*$^{\mathrm{I}}$	H4-Rha*p*$^{\mathrm{I}}$	2.70	2.87	2.88	2.87
Me-Rha*p*$^{\mathrm{II}}$	H4-Rha*p*$^{\mathrm{II}}$	2.51	2.86	2.85	2.86
Me-Rha*p*$^{\mathrm{III}}$	H4-Rha*p*$^{\mathrm{III}}$	2.69	2.87	2.87	2.87



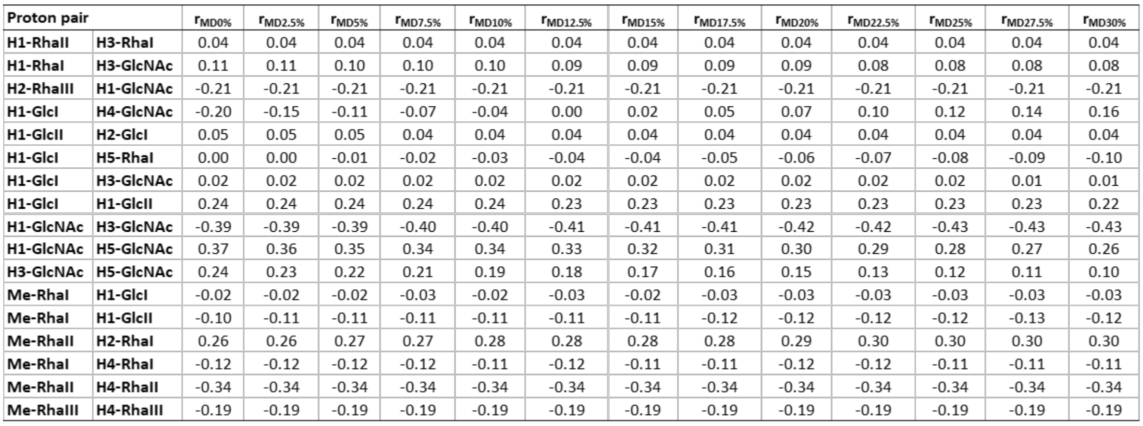
Proton pair		$r_{C\ Diff}$	$r_{NC\ Diff}$	$r_{P\ Diff}$	$r_{D\ Diff}$	$r_{Ave\ Diff}$
H1-Rha^{II}	H3-Rha^{I}	4%	1%	3%	1%	2%
H1-RhaI	H3-GlcNAc	3%	1%	0%	3%	2%
H2-RhaIII	H1-GlcNAc	5%	11%	9%	13%	9%
H1-GlcI	H4-GlcNAc	28%	19%	17%	16%	20%
H1-GlcII	H2-GlcI	8%	2%	2%	4%	4%
H1-GlcI	H5-RhaI	16%	20%	18%	21%	19%
H1-GlcI	H3-GlcNAc	8%	0%	1%	3%	3%
H1-GlcI	H1-GlcII	17%	9%	7%	7%	10%
H1-GlcNAc	H3-GlcNAc	16%	23%	20%	24%	21%
H1-GlcNAc	H5-GlcNAc	16%	9%	6%	10%	11%
H3-GlcNAc	H5-GlcNAc	22%	14%	18%	21%	19%
Me-RhaI	H1-GlcI	0%	0%	5%	9%	4%
Me-RhaI	H1-GlcII	6%	4%	7%	1%	5%
Me-RhaII	H2-RhaI	2%	16%	3%	20%	10%
Me-RhaI	H4-RhaI	2%	2%	6%	2%	3%
Me-RhaII	H4-RhaII	13%	13%	13%	13%	13%
Me-RhaIII	H4-RhaIII	7%	7%	6%	6%	7%

Table B.6: Percentage difference of ^{1}H-^{1}H distances derived from nOe measurements and MD simulations for *Sf* 7a O-Ag.



Proton pair		$r_{MD0\%}$	$r_{MD2.5\%}$	$r_{MD5\%}$	$r_{MD7.5\%}$	$r_{MD10\%}$	$r_{MD12.5\%}$	$r_{MD15\%}$	$r_{MD17.5\%}$	$r_{MD20\%}$	$r_{MD22.5\%}$	$r_{MD25\%}$	$r_{MD27.5\%}$	$r_{MD30\%}$
H1-RhaII	**H3-RhaI**	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32
H1-RhaI	**H3-GlcNAc**	2.22	2.22	2.23	2.23	2.23	2.23	2.24	2.24	2.24	2.25	2.25	2.25	2.25
H2-RhaIII	**H1-GlcNAc**	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29	2.29
H1-GlcI	**H4-GlcNAc**	2.94	2.89	2.85	2.81	2.78	2.74	2.72	2.69	2.67	2.64	2.62	2.60	2.58
H1-GlcII	**H2-GlcI**	2.33	2.33	2.33	2.33	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34
H1-GlcI	**H5-RhaI**	2.30	2.31	2.32	2.33	2.34	2.35	2.35	2.36	2.37	2.38	2.39	2.40	2.41
H1-GlcI	**H3-GlcNAc**	2.65	2.65	2.65	2.65	2.65	2.66	2.65	2.66	2.66	2.66	2.66	2.66	2.66
H1-GlcI	**H1-GlcII**	2.22	2.22	2.22	2.23	2.23	2.23	2.23	2.23	2.23	2.24	2.24	2.24	2.24
H1-GlcNAc	**H3-GlcNAc**	2.78	2.79	2.79	2.80	2.80	2.80	2.81	2.81	2.82	2.82	2.82	2.83	2.83
H1-GlcNAc	**H5-GlcNAc**	2.23	2.23	2.24	2.25	2.26	2.27	2.28	2.29	2.30	2.31	2.32	2.33	2.34
H3-GlcNAc	**H5-GlcNAc**	2.76	2.77	2.79	2.80	2.81	2.82	2.83	2.84	2.86	2.87	2.88	2.89	2.91
Me-RhaI	**H1-GlcI**	3.24	3.24	3.24	3.24	3.24	3.24	3.24	3.25	3.25	3.25	3.24	3.25	3.25
Me-RhaI	**H1-GlcII**	3.15	3.15	3.16	3.15	3.16	3.16	3.16	3.17	3.17	3.17	3.17	3.18	3.17
Me-RhaII	**H2-RhaI**	3.94	3.94	3.93	3.93	3.93	3.92	3.92	3.92	3.91	3.91	3.90	3.90	3.90
Me-RhaI	**H4-RhaI**	2.88	2.88	2.87	2.88	2.87	2.88	2.87	2.87	2.88	2.88	2.87	2.87	2.87
Me-RhaII	**H4-RhaII**	2.85	2.85	2.85	2.86	2.86	2.86	2.86	2.86	2.86	2.86	2.86	2.86	2.86
Me-RhaIII	**H4-RhaIII**	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.87

Table B.7: ^{1}H-^{1}H distances for different boat/skew contributions for *Sf* 7a O-Ag.



Proton pair		$r_{MD0\%}$	$r_{MD2.5\%}$	$r_{MD5\%}$	$r_{MD7.5\%}$	$r_{MD10\%}$	$r_{MD12.5\%}$	$r_{MD15\%}$	$r_{MD17.5\%}$	$r_{MD20\%}$	$r_{MD22.5\%}$	$r_{MD25\%}$	$r_{MD27.5\%}$	$r_{MD30\%}$
H1-RhaII	**H3-RhaI**	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
H1-RhaI	**H3-GlcNAc**	0.11	0.11	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.08
H2-RhaIII	**H1-GlcNAc**	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21	-0.21
H1-GlcI	**H4-GlcNAc**	-0.20	-0.15	-0.11	-0.07	-0.04	0.00	0.02	0.05	0.07	0.10	0.12	0.14	0.16
H1-GlcII	**H2-GlcI**	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
H1-GlcI	**H5-RhaI**	0.00	0.00	-0.01	-0.02	-0.03	-0.04	-0.04	-0.05	-0.06	-0.07	-0.08	-0.09	-0.10
H1-GlcI	**H3-GlcNAc**	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01
H1-GlcI	**H1-GlcII**	0.24	0.24	0.24	0.24	0.24	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.22
H1-GlcNAc	**H3-GlcNAc**	-0.39	-0.39	-0.39	-0.40	-0.40	-0.41	-0.41	-0.41	-0.42	-0.42	-0.43	-0.43	-0.43
H1-GlcNAc	**H5-GlcNAc**	0.37	0.36	0.35	0.34	0.34	0.33	0.32	0.31	0.30	0.29	0.28	0.27	0.26
H3-GlcNAc	**H5-GlcNAc**	0.24	0.23	0.22	0.21	0.19	0.18	0.17	0.16	0.15	0.13	0.12	0.11	0.10
Me-RhaI	**H1-GlcI**	-0.02	-0.02	-0.02	-0.03	-0.02	-0.03	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
Me-RhaI	**H1-GlcII**	-0.10	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	-0.12	-0.12	-0.12	-0.12	-0.13	-0.12
Me-RhaII	**H2-RhaI**	0.26	0.26	0.27	0.27	0.28	0.28	0.28	0.28	0.29	0.30	0.30	0.30	0.30
Me-RhaI	**H4-RhaI**	-0.12	-0.12	-0.12	-0.12	-0.11	-0.12	-0.11	-0.11	-0.12	-0.12	-0.11	-0.11	-0.11
Me-RhaII	**H4-RhaII**	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34	-0.34
Me-RhaIII	**H4-RhaIII**	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19	-0.19

Table B.8: Root mean square difference between NMR and MD ^{1}H-^{1}H distances derived for *Sf* 7a O-Ag.