Molecular Modeling Studies of Carbohydrate Vaccine Antigens

Informing the Future of Vaccine Design

Author:

Nicole Inge Richardson

Supervisors:

Associate Professor Neil Ravenscroft and Professor Michelle Kuttel

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

UNIVERSITY OF CAPE TOWN

August 2023

Plagiarism Declaration

I know the meaning of plagiarism and declare that all the work in the dissertation, save for that which is properly acknowledged, is my own.

I, Nicole Inge Richardson, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorize the University to reproduce for the purpose of research, either the whole or any portion of the contents, in any manner whatsoever.

I hereby present this report in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry at the University of Cape Town.

Nicole Inge Richardson
03 August 2023

Permission to Include Publications

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

1. **N.I. Richardson**, M.M. Kuttel, F. St Michael, C. Cairns and A.D. Cox, Cross-reactivity of *Haemophilus influenzae* type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies, *Glycoconj. J.*, 2021, **38**, 735–746.

2. **N.I. Richardson**, N. Ravenscroft and M.M. Kuttel, Conformational comparisons of *Pasteurella multocida* types B and E and structurally related capsular polysaccharides, *Glycobiology*, 2023, https://doi.org/10.1093/glycob/cwad049.

3. **N.I. Richardson,** N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M.M. Kuttel, Conformational and Immunogenicity Studies of the *Shigella flexneri* Serogroup 6 O-Antigen: The Effect of O-Acetylation, *Vaccines*, 2021, **9**, 432**.**

4. **N.I. Richardson,** M.M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity, *Front. Mol. Biosci.*, 2022, **9**, 961532.

Nicole Inge Richardson

03 August 2023

Table of Contents

Abstract

This thesis delves into the intricate world of carbohydrate-based vaccine antigens by employing molecular dynamics simulations to explore the link between their structure, conformation, and immune function. Through a series of four case studies focused on distinct antigen targets, this research aims to predict potential cross-reactivity and cross-protection, rationalize observed immunological reactivity, and provide valuable insights into key epitopes and mechanisms for antigen-antibody binding.

The case studies encompass the following antigens: *Haemophilus influenzae* types a and b, *Pasteurella multocida* types B and E, *Shigella flexneri* serotype 6, and *Streptococcus pneumoniae* serogroup 10. Each case study investigates the conformational aspects of the target antigens and proposes mechanisms for observed immunological phenomena. The collective findings propose connections between structural features, conformation, and their functional implications in immune responses, drawing parallels across individual case studies to elucidate recurring motifs employed by pathogens such as antigen flexibility, structural modifications, and backbone shielding.

By broadening the applicability of this molecular modeling methodology, this research extends its reach to new target antigens and pathogens, offering a complementary approach to establish structure-function relationships and inform rational vaccine design. The continued application of this methodology to a diverse range of vaccine targets promises to expand the knowledge base in the field, potentially revealing additional features harnessed by pathogens to gain a competitive advantage and evade the immune system. As computational power continues to grow, the cost and time associated with modeling may decrease, further enhancing the accessibility of this methodology for future vaccine development endeavors.

Abbreviations and Symbols

3 RU	three repeating unit
6 RU	six repeating unit
allyl Gal*f*	allyl-β-galactofuranose
AMR	antimicrobial resistance
BSE	block standard averaging
BSA	bovine serum albumin
CDC	Centers for Disease Control and Prevention
CHARMM36	Chemistry at Harvard Macromolecular Mechanics
COOH	carboxylic acid
CPS	capsular polysaccharide
CUDA	Compute Unified Device Architecture
αDGal	α-D-galactopyranose
βDGal	β-D-galactopyranose
βDGal*f*	β-D-galactofuranose
βGalNAc	N-acetyl-β-D-galactosamine
βDGlc*p*	β-D-glucopyranose
ELISA	enzyme-linked immunosorbent assay
EP	epitope
ETD	exocyclic terminal-1,2-diol
Fru*f*	β-D-fructofuranose
Fs	femtosecond
G4C	group 4 carbohydrate
Gal	galactose
GalA	galacturonic acid
GalNAc	N-acetyl-β-D-galactosamine
GC	gas chromatography
GEMS	Global Enteric Multicenter Study
Glc	glucose
GlcNAc	N-acetyl-β-D-glucosamine
Gly	glycine
GMMA	Generalized Modules for Membrane Antigens
GPU	graphics processing unit

GSK	GlaxoSmithKline
GUI	Graphical user interface
Hia	*Haemophilus influenzae* type a
Hib	*Haemophilus influenzae* type b
HibMe	*Haemophilus influenzae* type b methylated at position O2 of ribose
Hid	*Haemophilus influenzae* type d
Hie	*Haemophilus influenzae* type e
hIntL-1	human intelectin-1
HPLC	high performance liquid chromatography
IPD	invasive pneumococcal disease
IUPAC	international union of pure and applied chemists
KDO	3-deoxy-D-*manno*-oct-2-ulopyranosonic acid
KO	D-*glycero*-D-*talo*-oct-2-ulosonic acid
LB	Luria-Bertani
LMIC	low- and middle-income country
LPS	lipopolysaccharide
ManNAcA	N-acetyl-β-D-mannosaminuronic acid
MD	molecular dynamics
Me	methyl
MS	mass spectroscopy
Na^+	sodium ion
NAc	N-Acetyl
NAMD	Nanoscale Molecular Dynamics
NMR	nuclear magnetic resonance
nPT	isothermal-isobaric
O-Ag	O-antigen
OD600	optical density at 600 nm
OH	hydroxyl
OMV	outer membrane vesicle
PBS	phosphate-buffered saline
PCV	Pneumococcal Conjugate Vaccine
PDB	Protein Data Bank
PmB	*Pasteurella multocida* type B
PmE	*Pasteurella multocida* type E

PME	Particle Mesh Ewald
PMF	potential mean force
α Pn	serotype-specific rabbit antisera
Pn10A	*Streptococcus pneumoniae* serotype 10A
Pn10B	*Streptococcus pneumoniae* serotype 10B
Pn10C	*Streptococcus pneumoniae* serotype 10C
Pn10D	*Streptococcus pneumoniae* serotype 10D
Pn10F	*Streptococcus pneumoniae* serotype 10F
PPV	Pneumococcal Polysaccharide Vaccine
PRP	polyribosylribitolphosphate
PS	polysaccharide
PSF	protein structure file
r	end-to-end distance
Rha	rhamnose
Rib	ribose
Rib*f*	ribofuranose
Rib-ol	ribitol
Rib-ol-5P	D-ribitol-5-phosphate
RMSD	root-mean-square deviation
RSCB	Royal Society of Chemistry Biology
RU	repeating unit
SASA	solvent accessible surface area
SBA	serum bactericidal assay
Sf	*Shigella flexneri*
Sf 2a	*Shigella flexneri* serotype 2a
Sf 6	*Shigella flexneri* serotype 6
Sf 6n	non-biological *Shigella flexneri* serotype 6n with Gal instead of GalA
Sf 6-3Ac	*Shigella flexneri* serotype 6 with O-acetylation of RhaIII O3
Sf 6-4Ac	*Shigella flexneri* serotype 6 with O-acetylation of RhaIII O4
Sf Y	*Shigella flexneri* serotype Y
SNFG	Symbol Nomenclature for Glycans
TIP3P	transferable intermolecular potential with 3 points
VMD	Visual Molecular Dynamics
WHO	World Health Organization

Acknowledgements

It is with profound gratitude and deep respect that I wish to acknowledge all those who have contributed to the successful completion of this journey in research. The contributions, guidance, and support of each one have been invaluable, and the accomplishment would not have been possible without them

First and foremost, I would like to thank my supervisors, Associate Professor Neil Ravenscroft and Professor Michelle Kuttel, for their exceptional guidance, support, mentorship, and financial support throughout this journey. Their expertise and encouragement have been invaluable in shaping my research and academic journey.

My sincere thanks to the University of Cape Town for providing funding and to the University of Cape Town's ICTS-HPC for providing the indispensable computational resources and maintaining the HPC cluster that facilitated my research.

I am deeply thankful to the research participants who collaborated with me on various aspects of my research. The team from GSK GVGH in Siena for their invaluable collaboration on *Shigella*. The team from NRC Canada for their collaboration on *Haemophilus influenzae* and insightful discussion on *Pasteurella multocida*. I would also like to express my thanks to John Boyce and Marina Harper from Monash University Melbourne Australia for their insightful discussions and contributions related to *Pasteurella multocida*.

I extend my gratitude to Dr. Jason Hlozek, whose help and guidance in the early stages of my project proved invaluable. His thoughtful advice over countless hours of video calls during the early days of the pandemic was of immense help.

To my partner, Felix Norton, I would like to express my heartfelt gratitude for unwavering support, understanding, and encouragement throughout this journey. Your presence has been a constant source of motivation and strength.

Last, but certainly not least, I acknowledge with profound gratitude the unwavering support and love of my family: Kevin, Claudia, and Justine. Their unflagging belief in my abilities and their continuous encouragement have been instrumental in my academic journey.

This thesis, therefore, is not a mere representation of my work, but a symbol of the collective effort of those who have walked this path with me. I thank you all.

Preface

Please note that for the sake of consistency throughout this thesis, references/bibliography as well as figure and table numbering have been re-formatted. Following feedback from examiners, minor changes have been made compared to published papers. Further, since the publications are included verbatim and are from journals preferring American English spelling, American English spelling is used.

1. Chapter 1

This thesis explores the fascinating field of glycoconjugate vaccines and their potential in combating bacterial diseases. Through a multidisciplinary approach, this research focuses on four separate comparative studies of bacterial carbohydrate antigens aiming to unravel links between structure, conformation, and immunological properties. To lay the groundwork for understanding the research, the introduction first delves into the concept of glycoconjugate vaccines, highlighting their significance in reducing the burden of bacterial diseases and contextualizing the study as well as presenting its key aims. It then delves into the concept of cross-protection and explores the link between conformation and function before considering polysaccharide antigen structure and conformational considerations. The use of molecular modeling methods, specifically molecular dynamics simulations, is introduced as a powerful tool for studying the conformational aspects of carbohydrate-based vaccine antigens with exploration of data interpretation. Finally, the focus turns to the structure of the thesis as the four case studies and relevant background literature are presented, setting the stage for the subsequent detailed exploration of each study in the following chapters.

1.1. Overview

1.1.1. Glycoconjugate Vaccines as Tools Against Bacterial Disease

Glycoconjugate vaccines – comprising a poly- or oligosaccharide carbohydrate antigen chemically conjugated to an immunogenic carrier protein – have proven highly effective over the last half century in significantly reducing the burden of bacterial disease and stimulating long-lasting protection against infection. [1]

Successes such as the *Haemophilus influenzae* type b, *Neisseria meningitidis*, and *Streptococcus Pneumoniae* conjugate vaccines have significantly improved global health by reducing the burden of these diseases, especially in low- middle-income countries (LMICs). [2,3] The usefulness of vaccines in our fight against bacterial disease is especially significant against a background of rising antimicrobial resistance (AMR) and increased difficulty in discovering new antibiotics. [1,2,4]

Despite these successes, AMR is on the rise as is the prevalence of non-vaccine serotypes since the introduction of these vaccines due to phenomena such as serotype replacement. [5] Vaccines with more serotypes (higher valency) may provide greater and continued disease coverage. As an example, the original licensed pneumococcal vaccine which (comprised antigens from seven serotypes) was successful, however, serotype replacement was seen within three years leading to the development of a 13-valent vaccine. [3] Subsequently, additional serotype replacement along with varying epidemiology has motivated the development of third generation vaccines with even higher valency for which 15- and 20-valent vaccines are licensed with 24- and higher valency vaccines in development. [3]

There are also many other bacterial diseases accounting for significant morbidity and mortality against which vaccines can be developed to improve global health. [2] Communicable diseases (specifically neonatal conditions, lower respiratory infections, and diarrheal disease) are among the top causes of mortality; especially in young children (< 5 yrs) and LMICs. [6] In many cases available antibiotics are proving

increasingly ineffective or are not available in resource limited settings, making these pathogens important vaccine targets. [2,4]

Further, veterinary vaccines are an often-overlooked area for vaccine development, with potential for significant food security and economic benefits especially in LMICs where, as with human disease, vaccines and antibiotic treatments are not always available, are expensive, and are difficult to access. [7]

There are still many bacterial disease targets for novel vaccine development and scope for improving current vaccines which would have significant economic impact and work towards sustainable development goals of: good health and wellbeing, no poverty, and - in the case of veterinary vaccines for livestock - zero hunger. [8]

1.1.2. Informing Future Vaccine Design

This thesis is comprised of four publications [9-12] taking the form of case studies (presented here as results chapters) exploring conformational aspects of carbohydrate-based vaccine antigens using molecular dynamics (MD) modeling simulations. In these case studies, we consider groups of antigens that are structurally similar enough for a modeling comparison to provide conformational insights into potential cross-reactivity and immunological function for guiding rational vaccine design.

The antigens chosen for study were primarily informed by vaccines currently under investigation or in development. We add to the existing knowledge in the field by considering structurally similar targets for which either no known modeling data is available or for which the data is very limited and warrants investigation with modern techniques. Additionally, this work explores several different vaccine targets covering four distinct pathogens (including a veterinary pathogen).

The methodology used in this thesis has been previously established, and here we apply it, building on limited (in terms of molecule size and simulation length) early MD studies as well as more recent simulations. [13-18] We extend the methodology by leveraging modern graphics processing unit (GPU)

acceleration to increase the size and simulation time of molecules studied. We also extend the data analysis by including different approaches and visualization strategies as needed.

Given an established methodology is used, the primary contribution of the thesis is in the application of the methodology to new problems to grow the literature and illustrate the applicability of the methodology as well as extending data analysis approaches. This makes the inclusions of publications an appropriate choice for the thesis as the episodic nature of the problems can be tied together with overarching observations and findings that further contribute the literature. Overall, the intention is to use these insights to contribute to linking conformation and function as well as informing the rational basis of vaccine antigen selection and design.

1.2. Background

1.2.1. The Role of Cross-Protection in Glycoconjugate Vaccine Design

A central challenge of vaccine design is achieving the maximum coverage of current and emerging bacterial serotypes with the lowest valency, as each additional serotype component (or antigen) further increases the difficulty in producing and the final cost of the vaccine. [13] Further, each vaccine target and pathogen are different and may not behave as expected based on previous experience.

Reduced valency can be achieved when cross-protection is demonstrated; whereby a specific vaccine antigen elicits an immune response protecting against infection by closely related (non-vaccine) serotypes. [13,19] The related (non-mutually exclusive) phenomenon of cross-reactivity occurs when a vaccine antigen elicits an immune response that recognizes a related (non-vaccine) antigen.

Obtaining evidence for cross-reactivity as an indicator of potential cross-protection between vaccine serotypes has historically been determined by pre-clinical studies and sera fractionation in animal models. Immunological cross-protection in humans is typically only properly established once sufficient data on efficacy is available. [13] Once established, cross-protection can, however, be difficult to rationalize based on structure, as close structural similarity is not a reliable predictor of cross-protection.

It is thus a critical aspect of vaccine design to understand the functional immunological aspects and potential efficacy of target vaccine antigens as early as possible in the vaccine development process to ensure safe, successful, and efficacious vaccine design in a cost- and time-effective manner.

1.2.2. Conformation is Important for Function

A central dogma in biology is the relationship between shape (or conformation) and function. In the context of vaccines, functionality and protection is related to the antigenicity of antigen – the ability to generate antibodies or induce an immune response in an organism. [20] Previous work has shown that small changes

(such as a linkage position or substituent) in the molecular structure of saccharide antigens may have significant effect on altering the conformation and flexibility as well as physical properties and antigenicity. [21,22]

Applying this, we can use structural and conformational similarity between antigens as a predictor of cross-reactivity and potential cross-protection. Conformational analysis can further attempt to answer questions such as: Which epitopes are important in antigenicity? Which epitopes contribute to cross-reactivity? What are the effects of post-translational modifications and side groups on conformation?

1.2.3. Polysaccharide Antigens

The outer capsular polysaccharides (CPSs) and lipopolysaccharides (LPSs) of bacterial pathogens play an important role in their survival and success in evading host immune responses. [23] Thus, these molecules are targeted for inclusion in vaccines as functional antibodies against these polysaccharides would be expected to provide pathogen specific protection. [20]

These polysaccharides are comprised of repeating units (RU) of monosaccharides joined together by covalent glycosidic linkages. [24] The RUs vary in length (usually two to six monosaccharides) and the glycosidic linkages are formed between the anomeric carbon of one monosaccharide and a hydroxyl oxygen of another. [23] While it seems simple enough (especially compared to protein biology), great diversity is present in this class of molecules.

This diversity stems from both the constituent monosaccharides, and the composition of the RU. Monosaccharides can vary in identity, be linear or cyclic, vary in ring size (furanose or pyranose), charge (acidic versus neutral), anomeric configuration (α or β), stereochemistry (D or L), ring linkage position and modifications such as acetylation. [23,24] RUs can vary in monosaccharide composition, length, branching (the inclusion of side groups off the main RU backbone), and the use of non-sugar moieties (such as acetals, amino acids, phosphates) amongst others. [23,24]

Notably, bacteria tend to use monosaccharides uncommon to hosts such as L-rhamnose and L-fucose which further expands the diversity of CPS and LPS. [23] The diversity in glycans across the different phyla is likely driven by varied evolutionary pressures resulting in the development of a glycome that best favors survival. [23] For example, bacteria are evolutionarily distinct from eukaryotes, they inhabit a wider range of ecological niches resulting in different and highly varied survival pressure, they also have shorter generational times, and the ability to exchange genes via plasmid mediated horizontal gene transfer all of which may result in the observed differences in glycome. [23]

1.2.3.1 Polysaccharide Structural Elucidation

Compositional, structural, and linkage elucidation of polysaccharides is primarily performed by wet chemical and physicochemical methods such as high performance liquid chromatography (HPLC), gas chromatography (GC), mass spectroscopy (MS), and nuclear magnetic resonance (NMR). [25] The former provides information on monosaccharide identity and composition as well as preliminary linkage positions while the latter (notably NMR) allows assignment of monosaccharide identity, stereochemistry, and configurations as well as RU composition and linkage positions. [25]

1.2.3.2 Conformational Elucidation of Polysaccharides

Conformation, however, is more difficult to study due to the high flexibility of this molecular class as well as the relatively large molecular weights and high heterogeneity of polysaccharide samples. [21] This makes carbohydrates and polysaccharides (despite their unassuming appearance) notoriously difficult to work with experimentally and makes conformational elucidation by traditional laboratory-based techniques, such as crystallography or NMR, challenging. [13,21,26]

Modern in-silico techniques have provided an alternative avenue to study conformation through molecular modeling simulations. This allows one to model conformational behavior in a biological system and provide insight into immunological features such as cross-protection. [13,21,22,26]

Whilst previously prohibitively expensive and slow to perform, the continual reduction in computational cost with concurrent advances in technology leading to increased computational speed have contributed significantly to the rise of these types of in-silico experiments in recent years. [22,26] Against the backdrop of increasing costs of traditional laboratory-based chemistry (with rising costs of materials, waste disposal, and health and safety requirements) computational experiments become an attractive tool to assist and inform laboratory experimentation as a more sustainable approach for future research in this field.

1.2.4. Molecular Modeling in Vaccine Design

Molecular modeling is a complementary technique that can be applied to understand the conformational aspects of molecules. [13,22,26] Since conformation is crucial in molecular recognition, understanding the conformational aspects of a target antigen can thus provide valuable insight and input into rational vaccine design.[13] Practically this can be achieved by performing a series of step-wise structural modifications to better understand the effect of different components, side groups, and post-translational modifications.

1.2.4.1 Molecular Modeling Simulations

Molecular Dynamics (MD) simulations enable the reproduction of the time dependent behavior of molecular systems. Using calculations, the MD simulations determine the physical properties of the system at a specific point in time, providing valuable information that may be difficult to obtain through other experimental.[22,26]

There are several different models or levels of theory that can be applied. The most accurate simulations use quantum mechanical descriptions in which all the physical properties of the system are described by wave functions. These simulations are, however, computationally expensive. Alternative modeling techniques (such as MD) can use classical mechanics and approximations for greater computational efficiency.[22,26]

1.2.4.2 Molecular Dynamics

MD simulations are concerned with modeling the behavior of molecular systems over time. The simulations are run by calculating the potential energy of the atoms in the system and then integrating the Newton's equations of motion for all atoms over small time steps (typically in the order of femtoseconds). At each time step, the positions and velocities of the particles are updated based on the forces acting on them. This process is repeated many times, allowing the simulation to capture the dynamics of the system over time; these equations and concepts are well described in textbooks and the literature. [22,26-28]

Broadly, molecular simulations require two things: a low-energy starting structure for the molecule of interest (in the desired solvent system) and a force field (which refers to the equations and parameters used to describe the potential energy of the molecule/system in the simulation) used to solve Newton's equations of motion - the mathematical equations that underpin the simulation. [22,26] The force field parameters are determined via ab initio or semi-empirical quantum chemistry calculations on small model systems, or by fitting to experimental data such as NMR, X-ray and electron diffraction, and infrared spectroscopy. [26]

We make use of the CHARMM36 additive force field [29-31] which was developed for biological macromolecules such as lipids, proteins, nucleic acids, and carbohydrates. It has been specifically parameterized for hexopyranose monosaccharide molecules with glycosidic linkages. CHARMM36 is a popular choice for MD simulations of carbohydrates, is well maintained and supported, and is being extended to a Drude polarizable force field. [13] Furthermore, it has been shown to perform well for our molecules of interest (specifically for rhamnose for which some force fields do not perform as well). [13-18,32]

When performing modeling simulations, the computational cost of the simulations increases significantly as the system increases in number of particles. This is because a greater number of particles in the system result in a greater number of inter- and intramolecular force calculations and is especially relevant in solvated systems where the intermolecular forces between molecule and solvent as well as between solvent molecules need to be calculated. Furthermore, for solvated systems the simulation box scales non-linearly

and a doubling in molecule length of 2x would result in a corresponding volume increase of 2^3x for a cubic simulation box. The solvent very quickly becomes the bulk of the calculation and methods such ignoring interactions beyond a certain distance can help reduce the calculation scaling, but there is still a significant increase and the number of particles in a simulation is an important consideration. [26,27] For carbohydrate-based vaccine antigens a representative number of repeats of the RU is used. For short RUs this is relatively cost effective, however, for longer RUs the computational cost increases significantly. [26] The use of GPU acceleration has, thus been a major advance in the field in allowing the modeling of larger systems and with longer simulation times. [33]

In this work we use an established methodology for MD simulations of polysaccharide RUs, [13-18] and were further able to take advantage of recent computational advances - namely the use of more powerful GPUs and software upgrades - which further leverage GPU acceleration of the calculations. [34] These advances were especially relevant to Chapter 5 which was the largest molecular system and longest simulation attempted in this work and allowed these simulations to be performed with relative ease.

1.2.5. Molecular Dynamics Methodology

We used an established methodology to approach these modeling studies, as has been previously described and is briefly explored here as well as in the methods sections of the results chapters. [13-18]

Three repeating unit (3 RU) and six repeating unit (6 RU) chains for MD simulations were built using the CarbBuilder software. [35,36] Starting structures for each molecule were built using low energy glycosidic linkage orientations obtained from potential mean force calculations (calculated on the dihedral angle of the glycosidic linkage). The psfgen tool was used to create protein structure files (PSF) for simulation with the CHARMM36 additive force field for carbohydrates. [29,30]

Where parameters were not available for residues or linkages in the CHARMM36 additive force field, they were added either from the CHARMM36 protein force field (parameters for glycine) or based on parameters already present for these types of molecules (parameters for ribitol and phosphate linkages).

Starting structures were minimized with the Nanoscale Molecular Dynamics (NAMD) program [37] (to ensure the starting structures are low energy) before using the Visual Molecular Dynamics (VMD) software [38] solvation tool to add cubic TIP3P [39] water boxes. When applicable, for charged systems, the VMD and ionization tool was used to neutralize systems with sodium (Na^+) counter ions to ensure the force in the system converge.

MD runs were performed using the NAMD program with CUDA extensions for graphics processor acceleration [40] and the CHARMM36 additive force field.

Periodic boundary conditions equivalent to the cubic box size of the solvated system with wrapping are used to remove boundary effects when the molecules leave the box as they will reappear on the other side of the box. [28,37] The runs proceed following an initial minimization and heating protocol which allows for gradual increases in temperature up to the desired simulation temperature while maintaining system stability. A timestep of 1 femtosecond (fs) was used which is small enough to account for the fastest periodic motion of a system (e.g. bond vibration in the order of ~10 fs) but large enough so as to be more computationally efficient.

1.2.5.1 Potential Energy Calculation Considerations

Long range electrostatics for potential energy calculations were implemented with the Particle Mesh Ewald (PME) summation grid. The PME summation algorithm uses a Fourier transform for the long-range electrostatic contributions which converges much faster and reduces the number of calculations from exponential to logarithmic. [41] Similarly, electrostatic interactions of the 1-2 and 1-3 pairs were excluded from the calculations and a distance-based cut-off for inclusion in the pair list for calculation of non-bonded forces was used. Smoothing functions were applied to both the electrostatics and van der Waals forces with

switching and cut-off distances to gradually truncate (rather than sharply cutoff) long-range contributions to zero.

1.2.5.2 Classical Mechanics Optimizations

A Leap-Frog Verlet integrator was used to integrate the equations of motion in which the short-range non-bonded interactions are calculated every timestep while the full (more numerous) electrostatics calculations were performed every other timestep. The calculation of velocity and position are also performed a half step out of phase. Overall, this significantly reduces the number of calculations required and splits the computational load more evenly. [42]

1.2.5.3 Statistical Mechanics Considerations

Simulations were sampled under constant temperature and pressure - the isothermal-isobaric (nPT) ensemble. Langevin dynamics [43] were used to control the temperature with a damping coefficient to account for omitted degrees of freedom. Nosé-Hoover Langevin piston dynamics were used as a barostat to maintain a target pressure of 1 atm by varying the volume of the simulation cell. [44,45]

1.2.5.4 Approach to Convergence

We addressed convergence visually and using the block standard averaging method described by Grossfield and Zuckerman [46] applied to two observables: end-to-end distance and radius of gyration. Briefly, the method calculated the Block Standard Error (BSE) by dividing the simulation into block of different time lengths and calculating the standard deviation between the average of the observable for that block. From the BSE, correlation time and number of independent variables can be determined and, once block length is greater than correlation time, the blocks can be considered independent of each other, and the simulation can be deemed to have run for long enough.

1.2.6. Interpreting the Data

The outputs of MD simulations are large data files describing the positions of all atoms at each time point (or frame) of the molecule's trajectory. Deriving meaning and results from these data requires subsequent analysis. For our purposes, the analysis is focused around understanding molecular properties and conformation – building on an established analysis methodology expanding and innovating as necessary to establish the conformational rationalization for observed phenomena. [13-18] We provide a brief description of the key analyses performed.

We begin data analysis by gaining an understanding of the flexibility and extension of the molecule from the trajectory by plotting the end-to-end distance versus time for each frame. A similar plot of radius of gyration versus time (a measure of the variation in position of the atoms of the molecule relative to the center of mass or the root mean square deviation (RMSD) [22]) gives an indication of the compactness of the molecule and how this changes over time. The distribution and variation in end-to-end distance and radius of gyration gives important insight into global behaviors, anomalies, and areas for further investigation.

Next, the conformations are examined in detail using clustering analysis. Clusters are created by systematically comparing frames based on RMSD and are grouped based on similarity (within the specified cutoff) to a seed frame that represents each cluster. The seed frame is presented as representative of that cluster. By analysis of the cluster occupancies and shapes of structurally related molecules we can contrast and compare conformational aspects of these molecules such as shape, helical structure, and organization of monosaccharides in 3D space.

The primary source of flexibility in a polysaccharide comes from the freedom implicit in the glycosidic linkages, an analysis of the dihedral angles, phi (ϕ) and psi (ψ), of the glycosidic linkage (defined as used in the methods section of each results chapter) can prove valuable in understanding the clusters and effects of different linkage positions on conformation.

From this point the results of the initial analysis direct further investigations. The nature of the antigen surface is important for antigen-antibody binding, hence, investigating the solvent accessible surface (SASA) in terms of which residues are most exposed for binding (from clustering analysis) as well as in terms of hydrophobicity and hydrophilicity holds value. This is achieved by labeling the different hydrophilic and hydrophobic or neutral groups of the molecule and applying a surface representation to visualize areas that would be accessible to a solvent. This can further be numerically probed by rolling an imaginary water sphere of defined size over the surface and establishing how much of the surface is accessible and what the nature of these areas is. Furthermore, antibodies typically bind small fragments of a polysaccharide rather than the entire molecule making focused conformational and SASA enquiries into smaller sections of the molecule (one to six monosaccharides in length) valuable.

Additional analyses depending on the available information and data in the literature may also be possible.

1.3. Thesis Structure

This is a thesis that includes publications (listed previously) as four results chapters. These publications are included verbatim and formatted for consistency.

Chapter 2 considers the modeling of the highly flexible type a and b *H. influenzae* antigens as well as a synthetic analogue of type b. [9] We observe the presence of a phosphodiester bond in the backbone combined with a linear alditol sugar (ribitol) resulting in extreme flexibility and a potentially less immunogenic antigen compared to a more structurally defined antigen. Further, the lack of conformation suggests that the charged phosphate groups likely play an important role in antibody binding.

Chapter 3 explores veterinary pathogen antigens of *Pasteurella multocida* type B and E which are closely related to type e and d *H. influenzae* antigens. [10] Unlike the *H. influenzae* type a and b antigens from Chapter 1, these antigens exhibit high levels of conformational definition and appear to use an uncommon sugar, fructose, to disguise its immunogenic amino sugar backbone.

Chapter 4 compares the vaccine target antigen, *Shigella flexneri* type 6 with the common backbone of type

Y with special emphasis on the conformational effect of O-acetylation, a common post-translational modification that in some antigens (such as *N. meningitidis* serogroup A, *Salmonella typhi* Vi, and *Staphylococcus aureus* type 5 and 8 CPSs) is critical. [11,47] This antigen is shown to be both flexible and have a very different conformation and helix compared to the common *S. flexneri* backbone with the O-acetylation not being important. In this case, the altered backbone appears to be the primary immune evasion technique.

Chapter 5 explores the *S. pneumoniae* serogroup 10 CPSs which show high levels of asymmetric cross-reactivity. [12] This chapter explores the effect of different conformational epitopes on cross reactivity. We again see the use of a phosphodiester linkage and ribitol sugar in the backbone resulting in high levels of flexibility. Despite differences in backbone linkages and side groups this flexibility is maintained. We therefore investigate the side groups in more detail as the presence of an immunogenic galactofuranose on the backbone amino sugar as a side group and how it may interact with immune factors.

Chapter 6 provides an overview of the preceding chapters and concludes the thesis with discussion of potential implications of the work and future research questions that arise.

Here we provide further context and background for the pathogens of interest considered in these case studies.

1.3.1. *Haemophilus influenzae*

H. influenzae was the leading cause of meningitis disease and resultant mortality, particularly in young children. [48-50] Since the 1990s, a licensed conjugate vaccine (targeting the outer CPS) of *H. influenzae* serotype b (Hib), the most prevalent serotype, has been in use resulting in the epidemiology of the disease changing. [51-54] Disease due to the non-vaccine, *H. influenzae* serotype a (Hia) is now on the rise, especially in Indigenous North American populations despite widespread Hib vaccination. [1,52-55] This suggests that Hib does not cross-protect against Hia and was the motivation for development of a Hia conjugate vaccine which has been successfully tested in pre-clinical studies. [52-54,56] Furthermore, the AMR of *H. influenzae*

has led to it being labeled a priority pathogen by the World Health Organization (WHO). [57]

The disaccharide RU structures of the Hib and Hia CPSs are similar. Both comprise a disaccharide RU with a cyclic monosaccharide and a ribitol phosphate group, however, the cyclic monosaccharide identity (ribofuranose versus glucose) as well as the linkage positions differ between these antigens. [58,59] Understanding the conformational difference brought about by these structural differences can aid in rationalizing the lack of cross-protection observed.

Limited MD studies of the types b antigen have been previously performed; however, these were very short and were performed over two decades ago. [60,61]

1.3.2. *Pasteurella multocida*

P. multocida is a significant veterinary pathogen classified into five serogroups (A, B, D, E, and F) based on the bacterial CPS, which is important for virulence. [62,63] Serogroups B (PmB) and E (PmE) are the primary causative agents of bovine hemorrhagic septicemia, which is associated with significant yearly losses of livestock worldwide, especially in LMIC countries. [64,65] The disease is currently managed, albeit with limited efficacy, by whole-cell vaccination hence, a more effective vaccine would have significant benefit especially since resistance to commonly used antimicrobials has also been noted. [7,64,66,67]

The recent structural elucidation of the PmB and PmE CPSs renews interest in the CPS as a target antigen for a vaccine supported by limited immunological studies. [68-71] The PmB and PmE CPSs are structurally similar to each other and to the *H. influenzae* types e and d CPSs [70] comprising the same amino-sugar disaccharide backbone and fructofuranose side-group, but differing in their glycosidic linkage positions, and a glycine side group in PmB. [70]

Comparative modeling of PmB and PmE can inform the design of a potential *P. multocida* CPS vaccine and identify the function of structural modifications and potential important epitopes.

1.3.3. *Shigella flexneri*

The pathogenic *Shigella* bacteria are a leading cause of diarrheal disease and mortality, disproportionately affecting young children in low-income countries. [72-74] With increasing rates of antibiotic resistance, *Shigella* is a priority pathogen for global health bodies and an effective vaccine is critical to reduce disease burden. [4,57,72,73,75-77] The bacterial LPS O-antigen is the primary target antigen and a multivalent vaccine including *S. flexneri* serotype 6 as a component (as well as serotypes 2a, 3, and *S. sonnei*) has been proposed, with the goal of providing broad protection against *Shigella* disease. [4,74,76,78]

Several studies exploring the conformations of other *S. flexneri* O-antigens (O-Ags) which share a common backbone RU, that of serotype Y, have been performed, but none have considered serotype 6. [14,17] Serotype 6 has a different backbone RU with a different monosaccharide composition (notably the inclusion of a charged, acidic residue) and non-stoichiometric phage-mediated O-acetylation for which the antigenic and immunogenic importance has not yet been established. [79,80]

A systematic modeling study to establish the conformational aspects of serotype 6 relative to serotype Y as well as the effect of the charged acidic moiety and non-stoichiometric O-acetylation on conformation could provide rationalization for cross-reactivity and vaccine design as well as suggest potential binding epitopes.

1.3.4. *Streptococcus pneumoniae*

S. pneumoniae is a significant human pathogen and widespread use of pneumococcal conjugate vaccines has significantly reduced disease, however, the prevalence of non-vaccine serotypes (and emergence of antimicrobial resistant strains) has increased leading to the need for and development of higher valency, next generation vaccines. [4,19,57,81-83]

Serotype 10A is one such serotype, as the most prevalent of serogroup 10 (comprising four main serotypes 10A, 10B, 10C and 10F), that has been targeted for inclusion in these next-generation vaccines. [84-87] The

literature reports considerable puzzling asymmetric cross-reactivity within *S. pneumoniae* serogroup 10 with the reasons for this asymmetry being unclear. [85,88]

The CPSs of serogroup 10 contain the same backbone RU sugars (with ribitol phosphate moieties) differing only in linkage positions and side chain substitutions. [19,85,89] All serotypes contain a galactofuranose monosaccharide which is absent in mammals but occurs in a range of other bacterial pathogens and is often critical for virulence. [90,91] The human innate immune system is able to recognize an epitope on this monosaccharide thus pathogens have evolved to evade the immune system by modifying this epitope. [92-96]

Molecular modeling can be used to correlate the conformational features of these antigens with the observed cross-reactivity data and identify potential cross-protective epitopes and rationalize the mechanisms employed for immune evasion. Understanding this complex cross-reactivity is crucial for establishing the minimal vaccine components to provide for a maximally efficacious vaccine.

1.4. References

1. F. Berti and F. Micoli, Improving efficacy of glycoconjugate vaccines: from chemical conjugates to next generation constructs. *Curr. Opin. Immunol.*, 2020, **65**, 42-49.

2. F. Micoli, P. Costantino and R. Adamo, Potential targets for next generation antimicrobial glycoconjugate vaccines. *FEMS Microbiol. Rev.*, 2018, **42**, 388-423.

3. K. P. Klugman and G. L. Rodgers, Time for a third-generation pneumococcal conjugate vaccine. *Lancet Infect. Dis.*, 2021, **21**, 14-16.

4. U.S Department of Health and Human Services, CDC, Antibiotic Resistance Threats in the United States, 2019, https://www.cdc.gov/drugresistance/Biggest-Threats.html, (Accessed:09 March 2021).

5. R. Dagan, Serotype replacement in perspective. *Vaccine*, 2009, **27**, C22-C24.

6. M. Naghavi, A. A. Abajobir, C. Abbafati, K. M. Abbas, F. Abd-Allah, et al, Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, 2017, **390**, 1151-1210.

7. J. Lubroth, M. M. Rweyemamu, G. Viljoen, A. Diallo, B. Dungu and W. Amanfu, Veterinary vaccines and their use in developing countries. *Rev. - Off. Int. Epizoot.*, 2007, **26**, 179-201.

8. United Nations | Department of Economic and Social Affairs, The 17 goals | Sustainable Development, https://sdgs.un.org/goals, (Accessed:25 April 2023).

9. N. I. Richardson, M. M. Kuttel, F. St Michael, C. Cairns, A. D. Cox and N. Ravenscroft, Cross-reactivity of Haemophilus influenzae type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies. *Glycoconj. J.*, 2021, **38**, 735–746.

10. N. I. Richardson, Neil Ravenscroft and Michelle Mary Kuttel, Conformational comparisons of *Pasteurella multocida* type B and E and structurally related capsular polysaccharides, *Glycobiology*, 2023, https://doi.org/10.1093/glycob/cwad049.

11. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines*, 2021, **9**, 432.

12. N. I. Richardson, M. M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity. *Front. Mol. Biosci.*, 2022, **9**, 961532.

13. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

14. J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, Molecular Modeling of the Shigella flexneri Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a. *Vaccines*, 2020, **8**, 643.

15. M. M. Kuttel, Z. Timol and N. Ravenscroft, Cross-protection in Neisseria meningitidis serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr. Res.*, 2017, **446-447**, 40-47.

16. M. M. Kuttel and N. Ravenscroft, Conformation and cross-protection in Group B Streptococcus serotype III and Streptococcus pneumoniae serotype 14: a molecular modeling study. *Pharmaceuticals*, 2019, **12**, 28.

17. J. Hlozek, N. Ravenscroft and M. M. Kuttel, Effects of Glucosylation and O-Acetylation on the Conformation of Shigella flexneri Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B*, 2020, **124**, 2806-2814.

18. J. Hlozek, M. M. Kuttel and N. Ravenscroft, Conformations of Neisseria meningitidis serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.*, 2018, **465**, 44-51.

19. K. A. Geno, G. L. Gilbert, J. Y. Song, I. C. Skovsted, K. P. Klugman, et al, Pneumococcal capsules and their types: past, present, and future. *Clin. Microbiol. Rev.*, 2015, **28**, 871-899.

20. E. A. Kabat, The nature of an antigenic determinant. *J. Immunol.*, 1966, **97**, 1-11.

21. E. Fadda and R. J. Woods, Molecular simulations of carbohydrates and protein–carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today*, 2010, **15**, 596-609.

22. S. Genheden, A. Reymer, P. Saenz-Méndez, et al, in *Computational Tools for Chemical Biology*, ed. S. Martín-Santamaría, Royal Society of Chemistry, 2017, ch. 1, pp. 1-38.

23. J. H. Prestegard, J. Liu and G. Widmalm, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. H. Schnaar and P. H. Seeberger, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2017.

24. Y. A. Knirel, in *Microbial Glycobiology*, ed. O. Holst, P. J. Brennan and M. von Itzstein, Academic Press, London, Burlington, San Diego, 2009, ch. 4, pp. 57-73.

25. N. Ravenscroft, P. Costantino, P. Talaga, et al, in *Vaccine Analysis: Strategies, Principles, and Control*, ed. B. K. Nunnally, V. E. Turula and R. D. Sitrin, Springer, Berlin, Heidelberg, 2014, ch. 8, pp. 301-381.

26. C. J. Cramer, *Essentials of computational chemistry: theories and models*, John Wiley & Sons, England, 2nd edn., 2013.

27. J. Hlozek, *Molecular modeling of bacterial polysaccharide antigens to inform future vaccine development.*, University of Cape Town, 2021.

28. R. Bernardi, M. Bhandarkar, A. Bhatele, E. Bohm, R. Brunner, 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. H´enin, 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, F. Zhu, *NAMD user's guide version 2.13*, 2018.

29. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Comput. Chem.*, 2009, **5**, 2353-2370.

30. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, et al, Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.*, 2008, **29**, 2543-2564.

31. O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, et al, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *J. Chem. Theory Comput.*, 2011, **7**, 3162-3180.

32. R. D. Lazar, F. B. Akher, N. Ravenscroft and M. M. Kuttel, Carbohydrate Force Fields: The Role of Small Partial Atomic Charges in Preventing Conformational Collapse. *J. Chem. Theory Comput.*, 2022, **18**, 1156-1172.

33. M. Krone, J. E. Stone, T. Ertl and K. Schulten, *in Proceedings of Eurographics Conference on Visualization (EuroVis)* ed. M. Meyer and T. Weinkauf. , Vienna Austria, 2012.

34. NAMD 3.0 Alpha, GPU-Resident Single-Node-Per-Replicate Test Builds, http://www.ks.uiuc.edu/Research/namd/alpha/3.0alpha/, (Accessed:25/04/ 2023).

35. M. M. Kuttel, J. Ståhle and G. Widmalm, CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.*, 2016, **37**, 2098-2105.

36. M. Kuttel, Y. Mao, G. Widmalm and M. Lundborg, *in Proceedings of 2011 IEEE Seventh International Conference on eScience*, ed. Anonymous , 2011.

37. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al, Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 2005, **26**, 1781-1802.

38. W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.*, 1996, **14**, 33-38.

39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983, **79**, 926-935.

40. J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten, Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.*, 2007, **28**, 2618-2640.

41. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.*, 1993, **98**, 10089-10092.

42. W. F. Van Gunsteren and H. J. Berendsen, A leap-frog algorithm for stochastic dynamics. *Mol. Simul.*, 1988, **1**, 173-185.

43. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 1995, **103**, 4613-4621.

44. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.*, 1985, **31**, 1695.

45. S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 1983, **50**, 1055-1076.

46. A. Grossfield and D. M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.*, 2009, **5**, 23-48.

47. F. Berti, R. De Ricco and R. Rappuoli, Role of O-acetylation in the immunogenicity of bacterial polysaccharide vaccines. *Molecules*, 2018, **23**, 1340.

48. R. S. W. Tsang and M. Ulanova, The changing epidemiology of invasive Haemophilus influenzae disease: Emergence and global presence of serotype a strains that may require a new vaccine for control. *Vaccine*, 2017, **35**, 4270-4275.

49. L. Barreto, A. D. Cox, M. Ulanova, M. G. Bruce and R. Tsang, The emerging Haemophilus influenzae serotype a infection and a potential vaccine: Implementation science in action. *Can. commun. dis. rep.*, 2017, **43**, 85-88.

50. M. P. E. Slack, Long Term Impact of Conjugate Vaccines on Haemophilus influenzae Meningitis: Narrative Review. *Microorganisms*, 2021, **9**, 886.

51. A. Zwahlen, J. S. Kroll, L. G. Rubin and E. R. Moxon, The molecular basis of pathogenicity in Haemophilus influenzae: comparative virulence of genetically-related capsular transformants and correlation with changes at the capsulation locus cap. *Microb. Pathog.*, 1989, **7**, 225-235.

52. A. D. Cox, D. Williams, C. Cairns, F. S. Michael, P. Fleming, et al, Investigating the candidacy of a capsular polysaccharide-based glycoconjugate as a vaccine to combat Haemophilus influenzae type a disease: a solution for an unmet public health need. *Vaccine*, 2017, **35**, 6129-6136.

53. A. D. Cox, L. Barreto, M. Ulanova, M. G. Bruce, R. Tsang and on behalf of the conference contributors, *in Proceedings of Implementation science: Developing a vaccine for Haemophilus influenzae serotype a: Proceedings of a workshop*, ed. Anonymous , Canada Communicable Disease Report, 2017.

54. A. D. Cox, R. K. Lee, M. Ulanova, M. G. Bruce and R. Tsang, *Proceedings of a workshop to discuss the epidemiology of invasive Haemophilus influenzae disease with emphasis on serotype a and b in the Americas, 2019*, Elsevier, 2021.

55. M. Ulanova, Invasive Haemophilus influenzae Serotype a Disease in the H. influenzae Serotype b Conjugate Vaccine Era: Where Are We Going? *Clin. Infect. Dis.*, 2020, **73**, e380–e382.

56. A. Shoukat, R. Van Exan and S. M. Moghadas, Cost-effectiveness of a potential vaccine candidate for Haemophilus influenzae serotype ‘a’. *Vaccine*, 2018, **36**, 1681-1688.

57. S. Shrivastava, P. Shrivastava and J. Ramasamy, World health organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *J. Med. Soc.*, 2018, **32**, 76-77.

58. X. Lemercinier and C. Jones, An NMR Spectroscopic Identity Test for the Control of the Capsular Polysaccharide from Haemophilus influenzae Type b. *Biologicals,* 2000, **28**, 175-183.

59. W. Egan, R. Schneerson, K. E. Werner and G. Zon, Structural studies and chemistry of bacterial capsular polysaccharides. Investigations of phosphodiester-linked capsular polysaccharides isolated from Haemophilus influenzae types a, b, c, and f: NMR spectroscopic identification and chemical modification of end groups and the nature of base-catalyzed hydrolytic depolymerization. *J. Am. Chem. Soc.,* 1982, **104**, 2898-2910.

60. C. Hoog, A. Laaksonen and G. Widmalm, Molecular dynamics simulations of the phosphodiester-linked repeating units of the Haemophilus influenzae types c and f capsular polysaccharides. *J. Phys. Chem. B,* 2001, **105**, 7074.

61. M. Maestre and C. S. Pérez, Conformational analysis of diribosylribitol phosphate by NMR spectroscopy and molecular dynamics. *Magn. Reson. Chem.,* 2000, **38**, 123-125.

62. G. R. Carter, in *Methods in microbiology,* ed. T. Bergan, Academic Press, New York, NY, 1984, pp. 247-258.

63. K. L. Heddleston, J. E. Gallagher and P. A. Rebers, Fowl cholera: gel diffusion precipitin test for serotyping Pasteurella multocida from avian species. *Avian Dis.,* 1972, **16**, 925-936.

64. B. A. Wilson and M. Ho, Pasteurella multocida: from Zoonosis to Cellular Microbiology. *Clin Microbiol Rev,* 2013, **26**, 631-655.

65. A. Benikrane and M. C. L. De Alwis, Haemorrhagic septicaemia, its significance, prevention and control in Asia. *Vet. Med. Czech,* 2002, **47**, 234–240.

66. S. F. Capik, H. K. Moberly and R. L. Larson, Systematic review of vaccine efficacy against Mannheimia haemolytica, Pasteurella multocida, and Histophilus somni in North American cattle. *Bov. pract.,* 2021, **55**, 125-133.

67. World Organisation For Animal Health (OIE), OIE Technical Disease Card: Haemorrhagic Septicaemia, https://www.oie.int/app/uploads/2021/09/haemorrhagic-septicemia.pdf, (Accessed:22 November 2021).

68. L. K. Nagy and C. W. Penn, Protection of cattle against experimental haemorrhagic septicaemia by the capsular antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1976, **20**, 249-253.

69. C. W. Penn and L. K. Nagy, Capsular and somatic antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1974, **16**, 251-259.

70. C. W. Penn and L. K. Nagy, Isolation of a protective, non-toxic capsular antigen from Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1976, **20**, 90-96.

71. F. St Michael, C. M. Cairns, P. Fleming, E. V. Vinogradov, J. D. Boyce, et al, The capsular polysaccharides of Pasteurella multocida serotypes B and E: Structural, genetic and serological comparisons. *Glycobiology,* 2021, **31**, 307-314.

72. I. A. Khalil, C. Troeger, B. F. Blacker, P. C. Rao, A. Brown, et al, Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.,* 2018, **18**, 1229-1240.

73. C. Troeger, M. Forouzanfar, P. C. Rao, I. Khalil, A. Brown, et al, Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis. ,* 2017, **17**, 909-948.

74. M. M. Raso, V. Arato, G. Gasperini and F. Micoli, Toward a Shigella Vaccine: Opportunities and Challenges to Fight an Antimicrobial-Resistant Pathogen. *Int. J. Mol. Sci.,* 2023, **24**, 4649.

75. K. L. Kotloff, J. A. Platts-Mills, D. Nasrin, A. Roose, W. C. Blackwelder and M. M. Levine, Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine,* 2017, **35**, 6783-6789.

76. D. E. Atherly, K. D. C. Lewis, J. Tate, U. D. Parashar and R. D. Rheingans, Projected health and economic impact of rotavirus vaccination in GAVI-eligible countries: 2011–2030. *Vaccine,* 2012, **30**, A7-A14.

77. K. L. Kotloff, M. S. Riddle, J. A. Platts-Mills, P. Pavlinac and A. K. M. Zaidi, Shigellosis. *Lancet,* 2018, **391**, 801-812.

78. S. Livio, N. A. Strockbine, S. Panchalingam, S. M. Tennant, E. M. Barry, et al, Shigella Isolates From the Global Enteric Multicenter Study Inform Vaccine Development. *Clin. Infect. Dis.,* 2014, **59**, 933-941.

79. Y. Knirel, Q. Sun, S. Senchenkova, A. Perepelov, A. Shashkov and J. Xu, O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms. *Biochemistry,* 2015, **80**, 901-914.

80. P. Chassagne, C. Fontana, C. Guerreiro, C. Gauthier, A. Phalipon, et al, Structural studies of the O-acetyl containing O-antigen from a Shigella flexneri serotype 6 strain and synthesis of oligosaccharide fragments thereof. *Eur. J. Org. Chem.,* 2013, **2013**, 4085-4106.

81. B. Wahl, K. L. O'Brien, A. Greenbaum, A. Majumder, L. Liu, et al, Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob. Health,* 2018, **6**, e744-e757.

82. M. Moreau and D. Schulz, Polysaccharide based vaccines for the prevention of pneumococcal infections. *J. Carbohydr. Chem.,* 2000, **19**, 419-434.

83. F. Ganaie, J. S. Saad, L. McGee, A. J. van Tonder, S. D. Bentley, et al, A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. *Am. Soc. Microbiol.,* 2020, **11**, 937.

84. C. Jones, Full assignment of the NMR spectrum of the capsular polysaccharide from Streptococcus pneumoniae serotype 10A. *Carbohydr. Res.,* 1995, **269**, 175-181.

85. J. Yang, M. H. Nahm, C. A. Bush and J. O. Cisar, Comparative structural and molecular characterization of Streptococcus pneumoniae capsular polysaccharide serogroup 10. *J. Biol. Chem.*, 2011, **286**, 35813-35822.

86. D. McGuinness, R. M. Kaufhold, P. M. McHugh, M. A. Winters, W. J. Smith, et al, Immunogenicity of PCV24, an expanded pneumococcal conjugate vaccine, in adult monkeys and protection in mice. *Vaccine*, 2021, **39**, 4231-4237.

87. E. Janssens, J. Flamaing, C. Vandermeulen, W. E. Peetermans, S. Desmet and P. De Munter, The 20-valent pneumococcal conjugate vaccine (PCV20): expected added value. *Acta Clin. Belg.*, 2022, **78**, 78-86.

88. J. Henrichsen, Six newly recognized types of Streptococcus pneumoniae. *J. Clin. Microbiol.*, 1995, **33**, 2759-2762.

89. J. Yang, N. Y. Shelat, C. A. Bush and J. O. Cisar, Structure and molecular characterization of Streptococcus pneumoniae capsular polysaccharide 10F by carbohydrate engineering in Streptococcus oralis. *J. Biol. Chem.*, 2010, **285**, 24217-24227.

90. B. Tefsen, A. F. Ram, I. van Die and F. H. Routier, Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology*, 2012, **22**, 456-469.

91. C. Marino, A. Rinflerch and R. M. de Lederkremer, *Galactofuranose antigens, a target for diagnosis of fungal infections in humans*, FSO199, Future Science, 2017.

92. D. A. Wesener, K. Wangkanont, R. McBride, X. Song, M. B. Kraft, et al, Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.*, 2015, **22**, 603-610.

93. C. R. Isabella, *Carbohydrate and bacterial binding specificity of human intelectin-1*, Massachusetts Institute of Technology, 2021.

94. L. L. Kiessling, Chemistry-driven glycoscience. *Bioorg. Med. Chem.*, 2018, **26**, 5229-5238.

95. L. Chen, J. Li and G. Yang, A comparative review of intelectins. *Scand. J. Immunol.*, 2020, **92**, e12882.

96. B. L. Spencer, J. S. Saad, A. T. Shenoy, C. J. Orihuela and M. H. Nahm, Position of O-acetylation within the capsular repeat unit impacts the biological properties of pneumococcal serotypes 33A and 33F. *Infect. Immun.*, 2017, **85**, 132.

1 Chapter 2

2.1 Preface

The following article "Cross-reactivity of *Haemophilus influenzae* **type a** and **b** Polysaccharides: Molecular Modeling and Conjugate Immunogenicity Studies" describes a molecular modeling study of the capsular polysaccharide (CPS) of *H. influenzae* **types b** and **a**. [1] The aim of this work was to generate a conformational model of the molecules to provide a mechanistic understanding for an observed lack of immunological cross reactivity and to rationalize the epidemiology.

It starts by considering the *H. influenzae* **type b** CPS (the antigen used in the first ever glycoconjugate vaccine) which is then compared to the **type a** antigen and a synthetic, cleavage resistant **type b** CPS.

Somewhat unusually for carbohydrates, the **type b** CPS had no defined conformation, behaving rather as a random coil as was described by sedimentation studies. [2] The flexibility is due to the combination of linear sugar (ribitol) and phosphodiester bond in the backbone resulting in high rotational freedom and creating a "hinge" in the backbone. This high flexibility was likely advantageous to the virulence of the bacterium as a poorly defined antigen is more difficult for an antibody to bind. We would expect this to make a poor vaccine antigen, but the highly charged phosphate and its highly solvent exposed nature seemed to dominate the surface and likely explain the success of this antigen as a vaccine target.

The **type a** CPS was very different, showing much greater conformational definition and a zig-zag shape. This illustrated the significant conformational differences brought about by relatively minor structural changes (different linkage of the linear sugar and phosphodiester). The greater conformational definition may explain why **type a** is a less successful pathogen compared to **type b** as we expect a more defined conformation would be easier for an antibody to recognize and bind. The modeling results are validated by immunological data from our collaborators and the conformational differences support the need for a bivalent vaccine to protect against both **type b** and **a**.

The synthetic, cleavage resistant **type b** CPS showed some conformational differences when compared to the native **type b** CPS. Pre-clinical studies, however, suggest these differences may not affect the efficacy as a vaccine antigen and a cleavage resistant antigen is beneficial as it extends the vaccine shelf life and simplifies the vaccine purification and formulation. [3,4]

This work is an example of flexibility brought about by linear sugars and phosphodiester linkages in the backbone and prompted us to consider alternative analysis methods beyond just considering the conformational aspects. We explored the solvent accessible surface and considered potential epitopes that better describe what a potential antibody may see as accessible for binding to derive insights and meaning from the data.

Cross-reactivity of *Haemophilus influenzae* type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies

Nicole I. Richardson[a], Michelle M. Kuttel[b], Frank St. Michael[c], Chantelle Cairns[c], Andrew D. Cox[c], Neil Ravenscroft [a]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

[c] Vaccine and Emerging Infections Research, Human Health Therapeutics Research Centre, National Research Council, Ottawa, ON K1A 0R6, Canada

* Correspondence: Neil Ravenscroft neil.ravenscroft@uct.ac.za

N.I. Richardson, M.M. Kuttel, F. St Michael, C. Cairns and A.D. Cox, Cross-reactivity of *Haemophilus influenzae* type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies, *Glycoconj. J.*, 2021, **38**, 735–746.

2.2 Abstract

Haemophilus influenzae is a leading cause of meningitis disease and mortality, particularly in young children. Since the introduction of a licensed conjugate vaccine (targeting the outer capsular polysaccharide) against the most prevalent serotype, *Haemophilus influenzae* serotype b, the epidemiology of the disease has changed and *Haemophilus influenzae* serotype a is on the rise, especially in Indigenous North American populations. Here we apply molecular modeling to explore the preferred conformations of the serotype a and b capsular polysaccharides as well as a modified hydrolysis resistant serotype b polysaccharide. Although both serotype b and the modified serotype b have similar random coil behavior, our simulations reveal some differences in the polysaccharide conformations and surfaces which may impact antibody cross-reactivity between these two antigens. Importantly, we find significant conformational differences between the serotype a and b polysaccharides, indicating a potential lack of cross-reactivity that is corroborated by immunological data showing little recognition or killing between heterologous serotypes. These findings support the current development of a serotype a conjugate vaccine.

2.3 Introduction

Haemophilus influenzae is a Gram-negative coccobacillus that is a leading cause of meningitis, otitis media, bacteremia, and pneumonia, amongst others, particularly in children younger than five and immune-compromised individuals, with high levels of morbidity and disability. [5-8] There are six encapsulated serotypes, a – f, based on the capsular polysaccharide (CPS) structure as well as unencapsulated, non-typeable *H. influenzae* (NTHi) strains. [5,9,10] *H. influenzae* type b (Hib) was the leading cause of invasive *H. influenzae* disease prior to the introduction of conjugate vaccines in the 1980s and 1990s. [6,7,11,12] The successful introduction of Hib vaccination in many regions has changed the epidemiology of *H. influenzae* disease and resulted in the near elimination of Hib disease. Most invasive disease is now caused by non-b serotypes and NTHi strains, with Hia appearing as a leading cause of disease in Indigenous children in North America. [5-7,12-14] The emergence of Hia disease despite widespread Hib vaccination suggests that Hib does not cross-protect against Hia and is the motivation for development of a Hia conjugate vaccine. This is aligned with the WHO priority in their "Defeating meningitis by 2030: a global roadmap", particularly as the distribution and burden of disease is underestimated in low- and middle-income countries (LMICs). [8,10,13]

Licensed conjugate vaccines against Hib – based on the immunogenic CPS conjugated to different carrier proteins – are available and implemented in national immunization programs of all countries except for China and the Russian Federation. [8,15,16] For Hia, a promising candidate conjugate vaccine – which is expected to be cost effective in affected areas – has been developed and tested in pre-clinical studies; it showed immunogenicity, functional antibody stimulation (by serum bactericidal assay), and protective titers equivalent to that of the PedVaxHIB® vaccine (OMPC carrier). [6,12,14,17,18] A North American consortium including researchers at McGill University, Dalhousie University, National Research Council, Canadian Immunization Research Network, Canadian Glycomics Network and Inventprise LLC has identified funds to support GMP clinical lot production, GLP-toxicology studies and a Phase 1 clinical trial

of the candidate Hia conjugate vaccine that is scheduled to take place in 2022. [A. Cox, personal communication]

The disaccharide repeating unit structures of the Hib and Hia CPS are similar. The Hib repeating unit (RU) – also commonly known as known as polyribosylribitolphosphate (PRP) – contains ribose, in the *β*-furanose form, (Rib*f*) as well as ribitol phosphate (Rib-ol) linked by a 1,5-phosphodiester bond (**Figure 2.1A**), whereas Hia contains *β*-glucopyranose (Glc) and Rib-ol linked via a 4,5-phosphodiester bond (**Figure 2.1C**). [19,20] Ribitol phosphate is not present in the CPS of the other *H. influenzae* serotypes.[9,21]

Hib vaccines have poor stability due to the susceptibility of the PRP to hydrolytic cleavage of the phosphodiester bond through nucleophilic attack from O2 of Rib*f* – which is further catalyzed by the presence of divalent cations and some adjuvants. [3,4,22] This must be circumvented when designing the purification of the CPS, its conjugation and, particularly, its formulation in liquid combination vaccines. [23] One solution could be the use of a stabilized Hib antigen. The first example of a conjugate vaccine with a fully synthetic carbohydrate moiety is the QuimiHib made in Cuba; [8,15,16] more recently a modular synthetic route was proposed which identified an octasaccharide lead vaccine candidate. [24] The same group showed that methylation at O2 of Rib*f* (see **Figure 2.1B**, HibMe) provides a method for preventing hydrolytic cleavage, thus improving vaccine stability even in the presence of basic aqueous media or aluminum salts, as well as at elevated temperatures, while still inducing an immune response against Hib. [3,4]

As cross-protection between antigens is expected to require both chemical and conformational similarity, [25] molecular modeling can provide insight into the potential for cross-protection between the *H. influenzae* CPSs and stabilized Hib. While the structures of *H. influenzae* type a and b polysaccharides have been fully characterized through wet-chemical assays and NMR, [20,26-29] conformational studies have been limited. Physicochemical studies of Hib CPS and a size-reduced activated intermediate showed that Hib behaves as a highly flexible random coil. [2] Two decades ago, short molecular dynamics (MD) simulations of one RU of the Hib CPS showed a very flexible phosphodiester linkage. [29,30] To date, we are not aware of any

conformational studies on CPS of Hia or methyl stabilized Hib (HibMe).

To provide insight into potential cross-reactivity between the antigens, in this work we ran microsecond MD simulations of 6 RU of Hib as well as the vaccine candidates Hia and HibMe (**Figure 2.1**) in order to compare the flexibility and dominant conformations of the polysaccharide chains. For experimental corroboration, we also investigated the cross-reactivity between Hib and Hia by performing immunological studies in rabbits.

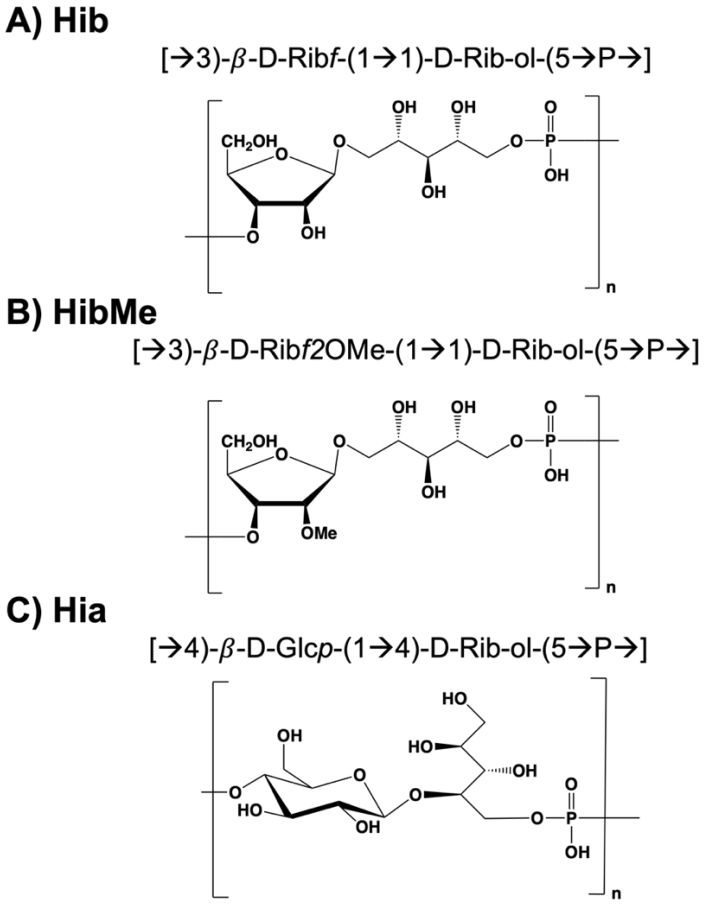


Figure 2.1: *Primary line structures and repeating units of the CPS modeled in this work:* ***(A)*** *H. influenzae type b (Hib),* ***(B)*** *non-biological stabilized Hib (HibMe) with a methyl group at O2 of Ribf,* ***(C)*** *H. influenzae type a (Hia).* [9,20,21,25,26]

2.4 Materials and Methods

2.4.1 Molecular Simulations

2.4.1.1 Molecular dynamics

Simulations were run using NAMD (version 2.13) [35] with CUDA extensions for graphics processor acceleration. [36] The CHARMM36 additive force field used to model the carbohydrates [37,38] and TIP3P to model the water molecules. [39] To represent missing glycosidic linkages, we added patches to the CHARMM36 force field for carbohydrates, as well as the necessary bond, angle and dihedral parameters. These were copied from similar linkages in the existing force field. Meso-ribitol, which has a number of alditols, is available in the CHARMM carbohydrate force field. Phosphodiester linkage parameters were adapted from those we previously added to the force field. [40] The initial coordinate files (.pdb format) and associated molecular force field files (protein structure file format, .psf) for all systems modeled are provided as online resources 1–6.

Starting structures for each molecule (**Supplementary Figure 2.1** and **Figure S2.1**) were built with CarbBuilder [41] using the default glycosidic linkages designated by CarbBuilder: for the β-linkages ϕ, ψ = 40, 0 and for the phosphodiester linkages ϕ, ψ = -40, 0. These initial structures were then minimized with NAMD for 10 000 steps at 300 K.

Chain length is an important consideration when modeling CPSs, as a short chain may have insufficient molecular flexibility, while long chains are more computationally expensive to model. On the basis of our previous work, we consider a 6 RU chain to be representative of the behavior of the longer polysaccharide. Further, antibodies bind small fragments of the CPS between one and seven residues in length [20] corresponding to between one and four RU in the case of *H. influenzae* type a and b, making a 6 RU model sufficient to explore antibody binding epitopes.

Minimized structures were solvated using the Visual Molecular Dynamics (VMD) software [42] to add cubic water boxes of 60 Å per side for Hib/HibMe and 72 Å per side for Hia. Each system was ionized with sodium (Na^+) counter ions to neutralize the system (3 ions for the 3RU, 6 for the 6 RU) at an ionic concentration of 0.027 mol/L and 0.046 mol/L for the 72 Å and 60 Å boxes respectively. Initial minimization and heating protocols comprised 5 K incremental temperature reassignments beginning at 10 K up to 300 K with 5 000 steps of NAMD minimization and 2 000 steps of MD at each temperature reassignment.

Periodic boundary conditions equivalent to the cubic box size were employed for the solvated simulation with wrapping on. Long range electrostatics were implemented with Particle Mesh Ewald summation (PME) on a 60 Å (Hib/ HibMe) or 72 Å (Hia) grid size. [43] Atoms were not held fixed, and the initial center of mass motion was off. The 1–3 pairs were excluded from non-bonded interactions, 1–4 interactions were not scaled, and the dielectric constant was set to 1. Smoothing functions were applied to both the electrostatics and van der Waals forces with switching and cut of distances of 12 Å and 15 Å, respectively.

A Leap-Frog Verlet integrator was used to integrate the equations of motion over a step size of 1 fs. A distance of 18 Å was used as the cut-off for inclusion in the pair list for calculation of non-bonded forces. The short-range non-bonded interactions were calculated every 1 fs, full electrostatics calculations were performed every 2 fs, and atoms were reassigned every 10 fs. [44]

Simulations were sampled under isothermal-isobaric (nPT) ensemble. Langevin dynamics were used to control the temperature with a damping coefficient of 5/ps. Nosé Hoover Langevin piston dynamics were used as a barostat to maintain a target pressure of 1 atm. Variable system volume was used with a piston period of 100 fs and decay of 50 fs. Post equilibration (200 ns), simulations underwent production runs of 1 300 ns as was required for the different models to converge.

2.4.1.2 Convergence

We addressed convergence using block standard averaging [45] applied to two metrics: end-to-end distance

and radius of gyration (**Supplementary Figure S2.2**). Block standard averaging was implemented with in-house Python scripts, as previously described. [31,45]

For all simulations, the blocked standard error (BSE) reached plateaus for both metrics, indicating convergence. The simulation lengths were large multiples of the correlation times for end-to-end distance (Hib, 11 ns; HibMe, 11 ns; Hia, 24 ns) and radius of gyration (Hib, 11 ns; HibMe, 11 ns; Hia, 25 ns). Further, the numbers of independent samples were >> 1 for both the end-to-end distance (Hib, 143; HibMe, 131; Hia, 64) and the radius of gyration (Hib, 131; HibMe, 142; Hia, 60). Our designated equilibration time of 200 ns is therefore much greater than the correlation time indicating that the properties of the system are not correlated or related on this timescale and that the simulation length should be sufficient.

2.4.1.3 Data analysis

Output trajectories were extracted every 25 ps and analyzed at 250 ps intervals. Inter-atomic distances and dihedral angles were measured using VMD, with data analyses performed with in-house Python scripts and plots generated using Matplotlib. [46]

The end-to-end distance, *r*, was measured from C1 of Rib-ol at the non-reducing end, to C1 of Rib*f* (for Hib) or Glc (for Hia) at the reducing end, thus excluding the highly flexible terminal residues.

The glycosidic linkages Rib*f* (1→1) Rib-ol and Glc*p* (1→4) Rib-ol are described by two dihedral angles ϕ = H_1-C_1-O_1- C_x' and ψ = C_1-O_1-C_x'-H_x', which are equivalent to ϕ_H and ψ_H in IUPAC nomenclature. [47,48] The phosphodiester linkages were defined as: ϕ = H_x-C_x-O_x-P, ψ = C_x-O_x-P-O_y, ω = O_x-P-O_y-C_y and ε = P-O_y-C_y-H_y as we have previously done for this type of linkage. [49] All dihedral angles were measured as a combination of the dihedrals from the central RUs, three and four thus providing a sample of the most central angles.

The most common chain conformations for each simulation were determined by clustering the simulation frames (post equilibration) into families and calculating the relative occupancies of each family. Clusters

comprising less than 2 % of the production run (post equilibration) were discarded. Clustering was performed using the WMC PhysBio plug-in for VMD's built-in cluster command. [50] Prior to clustering, the molecules were aligned on the ring and linkage atoms of the central Rib-ol residue of RU 3 and Rib*f* residue of RU 4. Clustering was performed as an RMSD ft to the ring and linkage atoms of the central 4 RU of the 6 RU chains (excluding the highly flexible terminal repeating units) with a cut-off of 5.5 Å.

Hydrophilic/hydrophobic regions of the molecular surface were analyzed using VMD's built-in "measure sasa" command. The solvent accessible surface area (SASA) analysis was performed by probing first hydrophilic regions (comprising hydroxyl groups, carbonyl groups, ring oxygens, phosphorous, and linkage oxygens) and then hydrophobic/ neutral regions (comprising methyl groups, CH_2 groups, ring carbons, and ring protons) of the molecule using a van der Waal's radius of 1.4 Å—analogous to that of water. The ratio of hydrophilic to hydrophobic/neutral regions was then calculated to determine the percentage hydrophilic surface area available for potential antibody binding.

Molecular conformations were visualized using VMD, with the PaperChain visualization algorithm used to highlight carbohydrate rings. [51] and the hydrophilic and hydrophobic surfaces were visualized with the Quicksurf visualization algorithm. [52]

2.4.2 Immunological studies

CPS from Hia and Hib were isolated from cell supernatants following fermenter growth, as described previously for Hia, however, no RNase treatment was utilized for Hib CPS isolation. [12] CRM_{197} carrier protein was prepared as described previously. [12] Conjugates of purified Hia and Hib CPS were prepared via direct reductive amination as described previously for the Hia conjugates.[8]

Three New Zealand female white rabbits (Hia: RCAV1-3), (Hib: RCBV1-3) were immunized subcutaneously with each conjugate at a 25 μg dose with the same prime and two boost strategy and adjuvanting with incomplete Freund's adjuvant as described previously. [12]

Final bleed day 70 (D70) rabbit sera from Hia and Hib conjugates were screened against BSA conjugates of Hia and Hib PS in ELISA as described previously. [12] Pre-immune sera were employed as the negative control.

Final bleed (D70) rabbit sera from Hia and Hib conjugates were screened for functionality in a serum bactericidal assay (SBA) against Hia strain (NRCC # 6753) and Hib strain (NRCC # 4274) as described previously. [12] Pre immune sera were employed as the negative control.

2.5 Results

We begin with a comparison of the flexibilities of the Hib, HibMe, and Hia CPS chains, then the molecular conformations, followed by the minimal binding epitopes and the characteristics of the molecular surface. We then discuss the results of our immunological studies in the light of the modeling results.

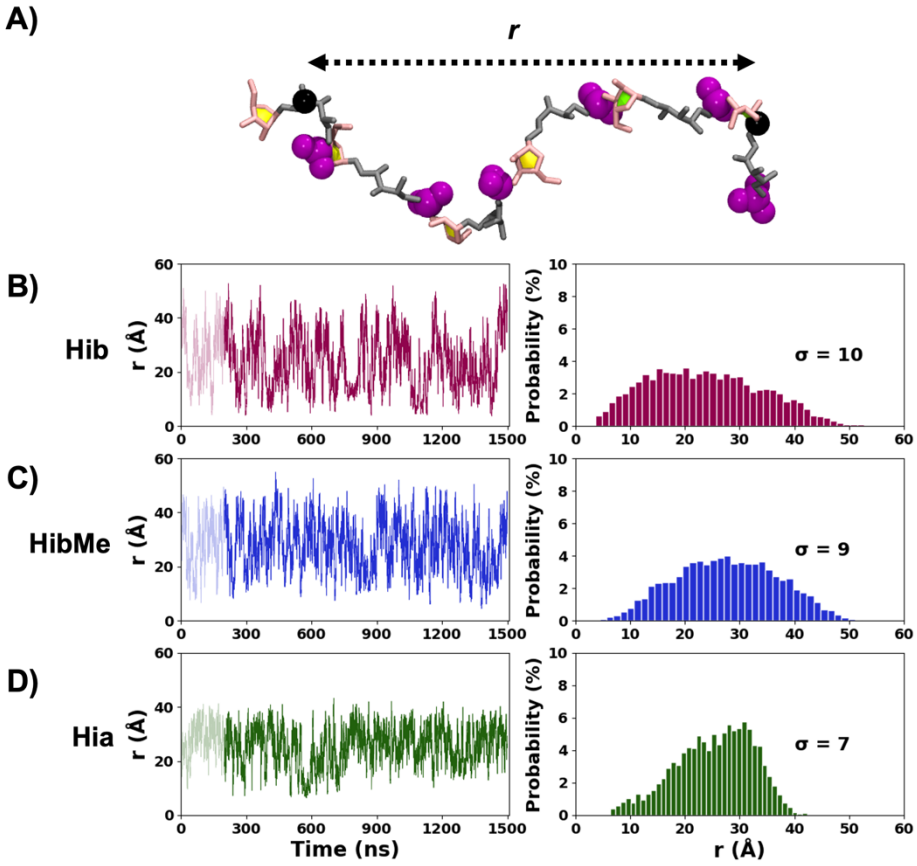


Figure 2.2: *Time series of the end-to-end distance, r, in the 6 RU CPS chains. Here r is defined as the distance (Å) from C1 of Rib-ol in RU1 to C1 of Ribf in RU 6 for Hib and HibMe* ***(A)****, or the equivalent C1of Glc in RU 6 (for Hia). Time series (left column) and histograms (right column) are plotted for r in* ***(B)*** *Hib,* ***(C)*** *HibMe, and* ***(D)*** *Hia. Histograms are labelled with the standard deviation, σ.*

2.5.1 Chain flexibility

Time series graphs of the molecular end-to-end distance, *r*, over the course of the MD simulations provide a simple metric for comparison of the molecular flexibility and extension of the 6 RU CPS chains. The *r*

time series plots (**Figure 2.2**, left column) and corresponding histograms (**Figure 2.2**, right column) reveal that all three molecules are very flexible, with rapid transition between a wide range of *r* values. The similarity of the plots for Hib (**Figure 2.2B**) and HibMe (**Figure 2.2C**) indicate that the O-methyl substitution on Hib does not affect the flexibility of the chain substantially: Hib and HibMe have the same range of *r* (5–55 Å), although the *r* histograms reveal that Hib has a greater incidence of conformations with $r < 10$ Å. However, Hib and HibMe together are considerably more flexible than Hia (**Figure 2.2D**). The range of *r* is considerably broader for Hib and HibMe than for Hia (5–45 Å), which lacks the extreme extensions of the chain. Further, the more compact distribution of Hia (σ = 7 Å), with a clear peak at approximately 30 Å, indicates that it is more conformationally defined than either Hib (σ = 10 Å) or HibMe (σ =9 Å).

2.5.2 Molecular conformations

The flexibility of the CPS antigens is further demonstrated in a comparison of the dominant CPS chain conformations for Hib, HibMe, and Hia (**Figure 2.3**). Hib has no dominant conformation, with the highest occupation for a conformational family at 7 % of simulation time (**Figure 2.3A**, Hib1). The wide range of collapsed (**Figure 2.3A**: Hib1, Hib2, Hib8) and extended (**Figure 2.3A**: Hib3-7 and Hib9-10) conformations with low occupancies (< 7 %) for Hib is consistent with the random coil classification of this antigen on the basis of sedimentation studies. [2] The conformations of HibMe (**Figure 2.3B**) are very similar to Hib, although with more extended chain conformations: O-methyl substitution of the Rib*f* residue removes the compact conformations of Hib. However, there remains a wide range of conformations of the chain, which also behaves as a random coil.

In contrast, the conformations of Hia are much more defined and regular (**Figure 2.3C**), with the primary cluster comprising 36 % of the simulation (**Figure 2.3C**, Hia1). Zig-zag conformations of the chain predominate (**Figure 2.3C**, Hia1, Hia5, Hia6), with bends at the Rib-ol of each RU. More compact conformations with accordion-style folding of the zig-zag do occur with relatively high frequency

(**Figure 2.3C**, Hia2, Hia4, Hia8, Hia9), but very extended conformations of the chain are absent. The glycosidic linkages (**Supplementary Figure S2.3** and **S2.4**) in the Hia simulations are more constrained and have different orientations than in Hib. Overall, Hia therefore has markedly different conformations and dynamics from Hib.

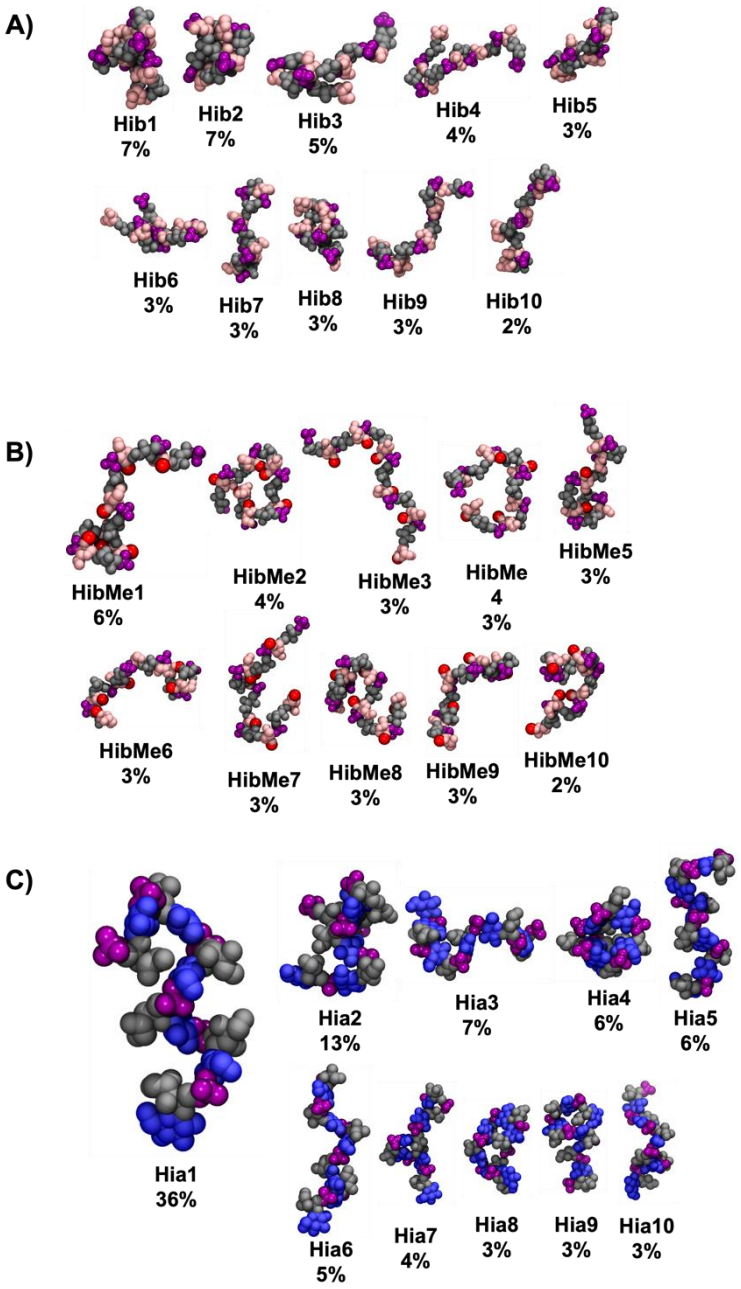


Figure 2.3: *Conformational families (seed frame for each cluster is shown) for the middle 4 RU of the 6 RU CPS chains:* ***(A)*** *Hib,* ***(B)*** *HibMe and* ***(C)*** *Hia. Families are scaled to illustrate differences in their relative percentage occupancies. Residues are colored: Ribf, pink; Rib-ol, gray; Glc, blue. The phosphate groups are highlighted in*

purple and the methyls in red.

2.5.3 Binding epitopes

As the CPS regions bound by antibodies comprise one to seven residues, [33] it is useful to compare the 6 RU CPS chains on this length scale. **Figure 2.4** shows the three main conformations of the central 4 RU segment (comprising 2 RU) for the Hib, HibMe, and Hia CPS chains. Overall, the three primary conformations of the 2 RU segments are much more consistent with each other than the conformations of the 6 RU chains.

At this scale it is clear that the flexible phosphodiester linkage acts as a hinge between repeating units for all three CPS chains. However, the conformations of the 2 RU segment are quite different across the three antigens. The primary conformations of Hib are compact, with sharp bends exposing the phosphate group (**Figure 2.4A**), whereas the primary conformations for HibMe are more extended, with the O-methyl group quite exposed on the chain (**Figure 2.4B**). For Hib (**Figure 2.4A**) the sharp bends enable conformational collapse and stabilizing interactions between the Rib-ol and opposing sides of the chain (**Figure 2.4D**) likely increase the preference for collapsed over extended conformations in this CPS. For HibMe the bulky methyl group at O2 of Rib*f* prevents collapse and more extended conformations therefore predominate.

The more moderate bends in the Hia conformations (**Figure 2.4C**) are due to the constrained 4,5 linked Rib-ol. Here the phosphate hinge exposes both the phosphate groups as well as the bulky C1-C3 portion of the Rib-ol for possible antibody binding (**Figure 2.4F**).

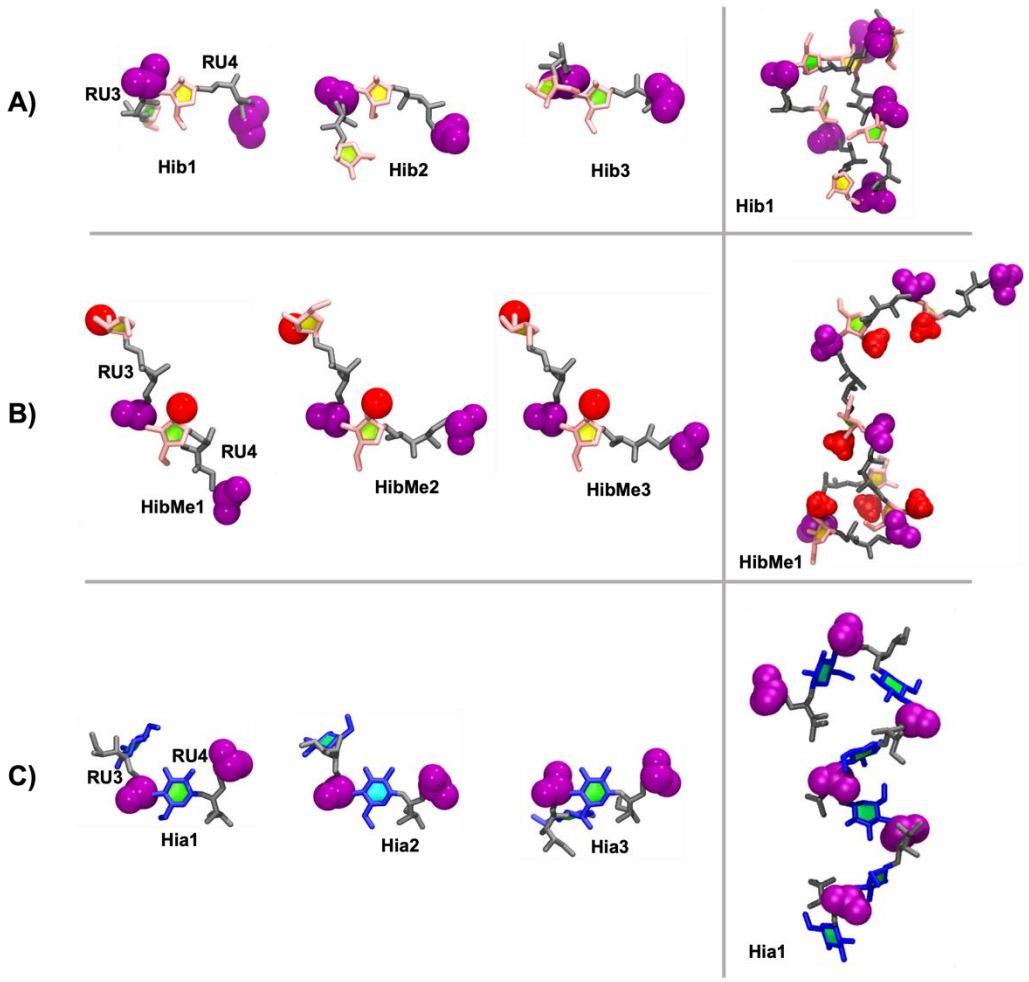


Figure 2.4: *Four residue sections of the CPS chains comprising RU 3 and RU 4 for the three primary conformational clusters:* ***(A)*** *Hib,* ***(B)*** *HibMe and* ***(C)*** *Hia. The full primary clusters for Hib, HibMe, and Hia are shown in the last column. Residues are colored: Ribf, pink; Rib-ol, gray; Glc, blue. The phosphate groups are highlighted in purple and the methyls in red.*

Comparison of the solvent exposed hydrophilic and hydrophobic regions of the molecular surfaces further highlights significant differences between the CPS antigens that may affect antibody binding. Time series and histograms of the solvent exposed hydrophilic surface area for step in the simulation are plotted in **Figure 2.5**. As expected, the surface of the Hib CPS (**Figure 2.5A**) is largely hydrophilic (average of 72 %) – due primarily to the bulky charged phosphate groups. However, the addition of the hydrophobic methyl group in HibMe greatly reduces the hydrophilicity of the exposed molecular surface to an average of 57 % (**Figure 2.5B**). This can be rationalized by considering that the methyl group in HibMe is relatively bulkier

(4 atoms) compared to the OH in Hib (2 atoms), thus the methyl substituent would contribute more to the hydrophobicity of the molecular surface in HibMe than the OH contributes to hydrophilicity of the surface in Hib. In addition to the conformational differences in the chain, this difference in the properties of the surface is another factor to consider that may affect cross-reactivity between the two antigens.

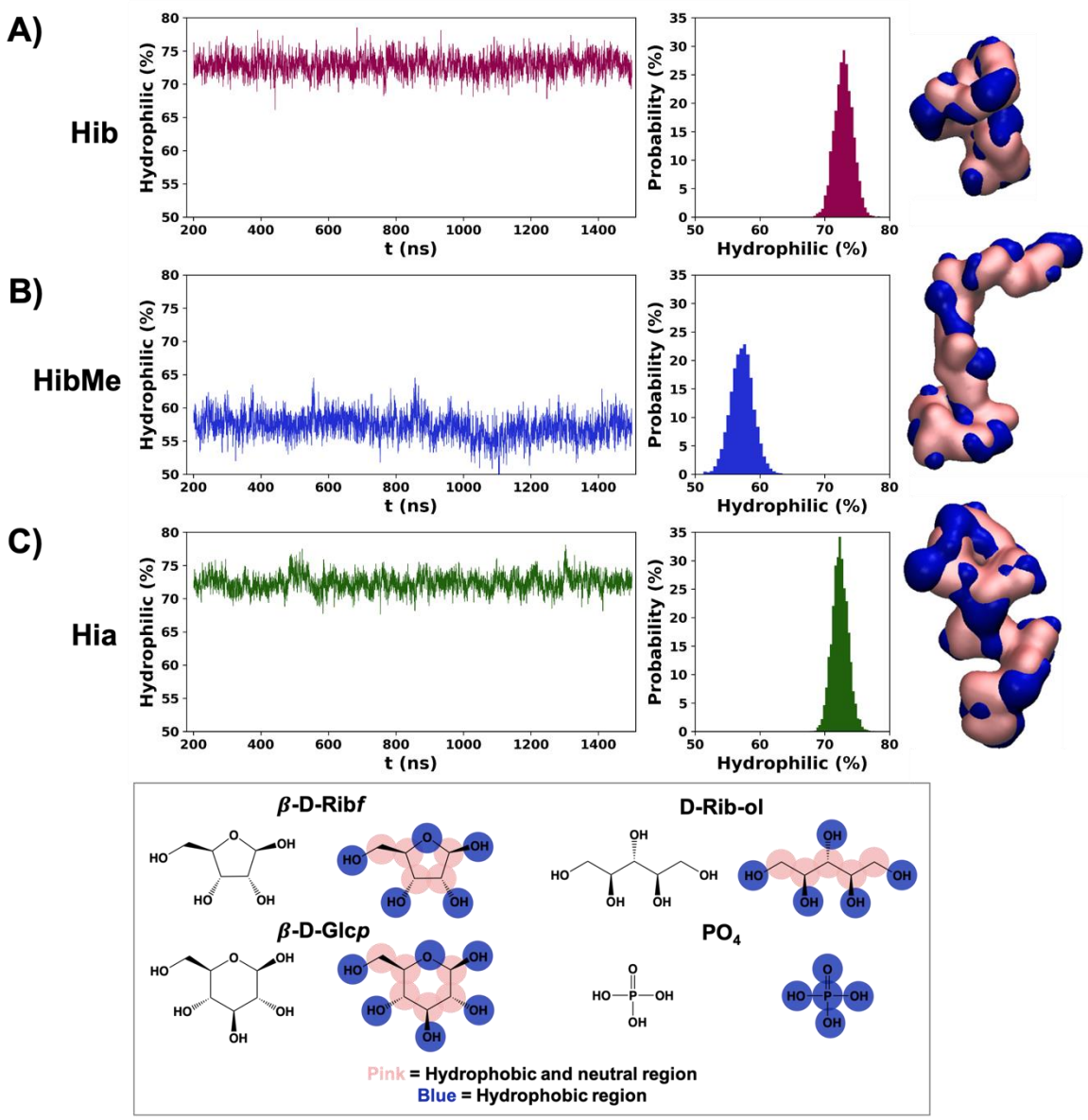


Figure 2.5: *Time series plots and histograms of the percentage of the solvent accessible surface area that is hydrophilic in the 6 RU simulations of* ***(A)*** *Hib,* ***(B)*** *HibMe, and* ***(C)*** *Hia. Examples of the molecular surfaces for each CPS with hydrophilic (blue) and hydrophobic (pink) patches indicated are shown in the right column; the key and classification for each RU moiety is given in the lower panel.*

Interestingly, the analysis shows that Hia (**Figure 2.5C**) has the same relative hydrophilicity as Hib, an average of 72 % (with a narrower distribution), despite the significant structural and conformational differences between these antigens. This is consistent with a comparison of the hydrophilic surfaces of the Rib*f* (**Supplementary Figure S2.5A**, 69 %) and Glc (**Supplementary Figure S2.5B**, 69 %) residues—which, given the common Rib-ol residue (**Supplementary Figure S2.5C**) contributes to the overall hydrophilicity of the surface.

2.5.4 Immunological studies

To investigate cross-reactivity between Hib and Hia, D70 rabbit sera derived from immunizations with three doses of the Hia and Hib conjugates were screened in ELISA for recognition of the homologous and heterologous CPS (**Figure 2.6**). Only the homologous CPS was recognized in each case, confirming a lack of cross-reactivity between the Hia and Hib CPS.

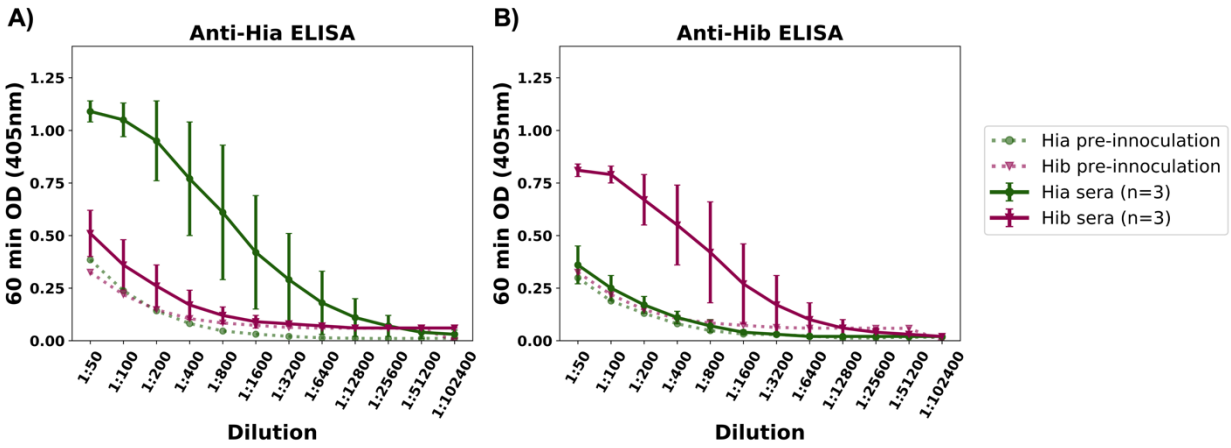


Figure 2.6: *ELISA screening of rabbit sera from Hia or Hib conjugate immunized rabbits (n=3) for recognition of:* ***(A)*** *Hia-BSA conjugate or* ***(B)*** *Hib-BSA conjugate. Homologous serotype recognition is seen, but no cross-recognition of the heterologous serotype. In all cases the pre-immunization sera were used as the control.*

The same D70 sera were utilized in an SBA assay to assess for functionality of the sera that supports complement mediated killing. As can be seen in **Figure 2.7**, the Hia conjugate derived sera could only facilitate the killing of the Hia strain and not killing of the Hib strain. The same reciprocal results were

obtained when the Hib conjugate derived sera were examined, confirming the inability of sera specific to one serotype to recognize nor facilitate the killing of the heterologous serotype strains.

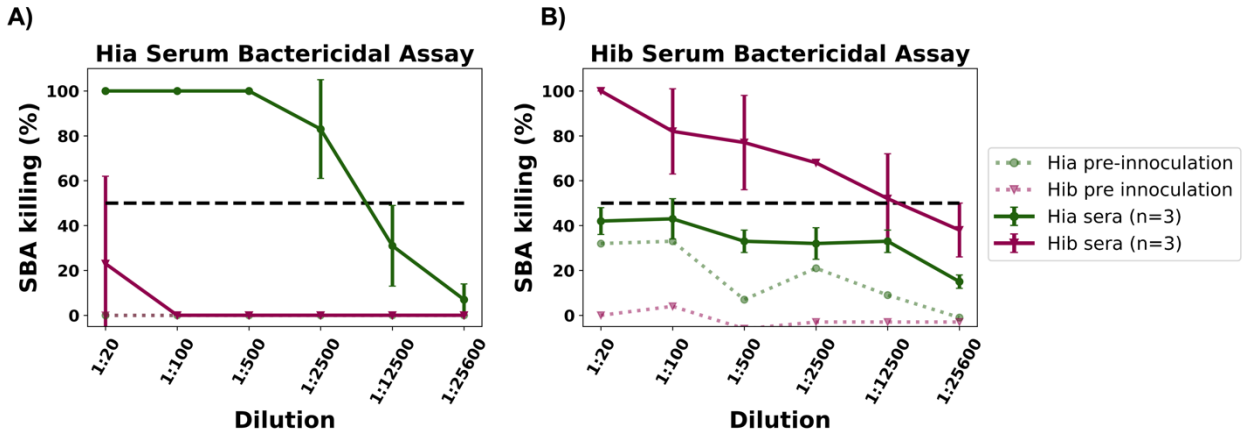


Figure 2.7: *Serum bactericidal assay (SBA) titration curves of rabbit sera from Hia or Hib conjugate immunized rabbits (n=3) against:* ***(A)*** *Hia bacterial strain # 6753 or* ***(B)*** *Hib bacterial strain # 4274. We see killing of the homologous serotype, but a lack of cross-killing of the heterologous serotype. In all cases, pre-immunization sera were used as the control.*

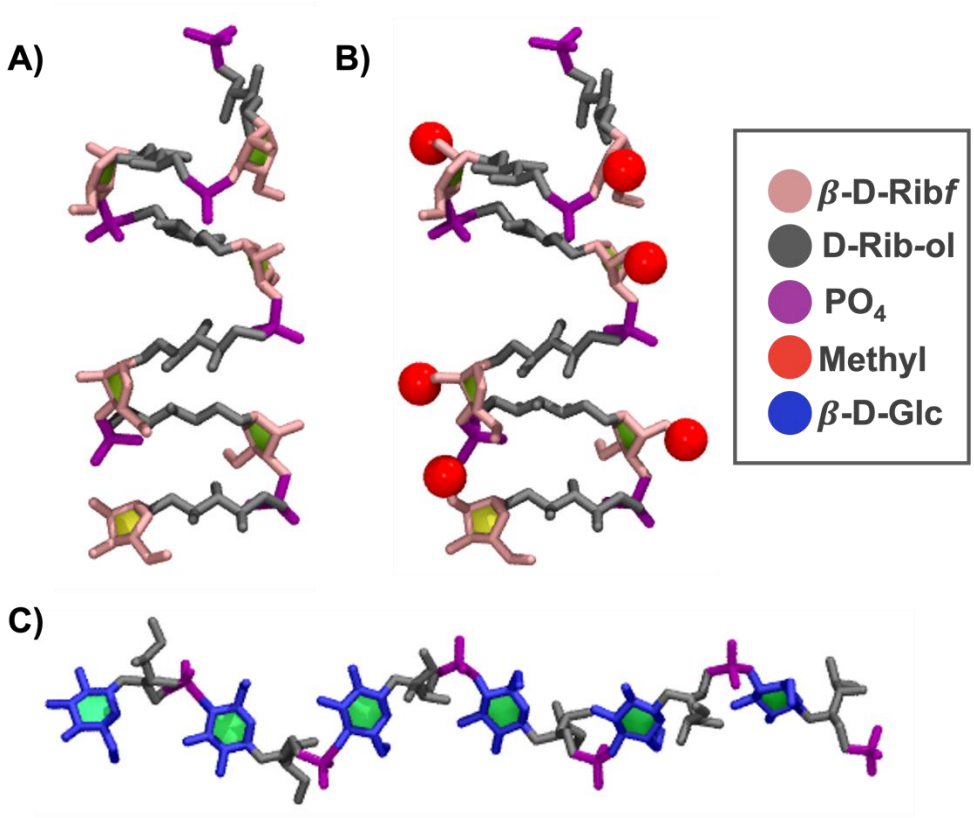
2.6 Discussion

We find that Hib and the modified HibMe CPS antigens both behave as unstructured random coils in solution. However, we show that there are also some differences between the Hib and HibMe saccharides that may affect antibody binding: Hib has a range of collapsed, globular conformations that do not occur in HibMe, the local conformation of the chain in Hib has more extreme bends than HibMe, and the molecular surface of Hib is much more hydrophilic than HibMe. However, despite the differences in conformation and molecular surface, a modified HibMe conjugate vaccine raised antibodies that recognized the natural Hib antigen. [4] This suggests that the presence of immuno-dominant charged phosphate [53] is the focus of antibody recognition and binding.

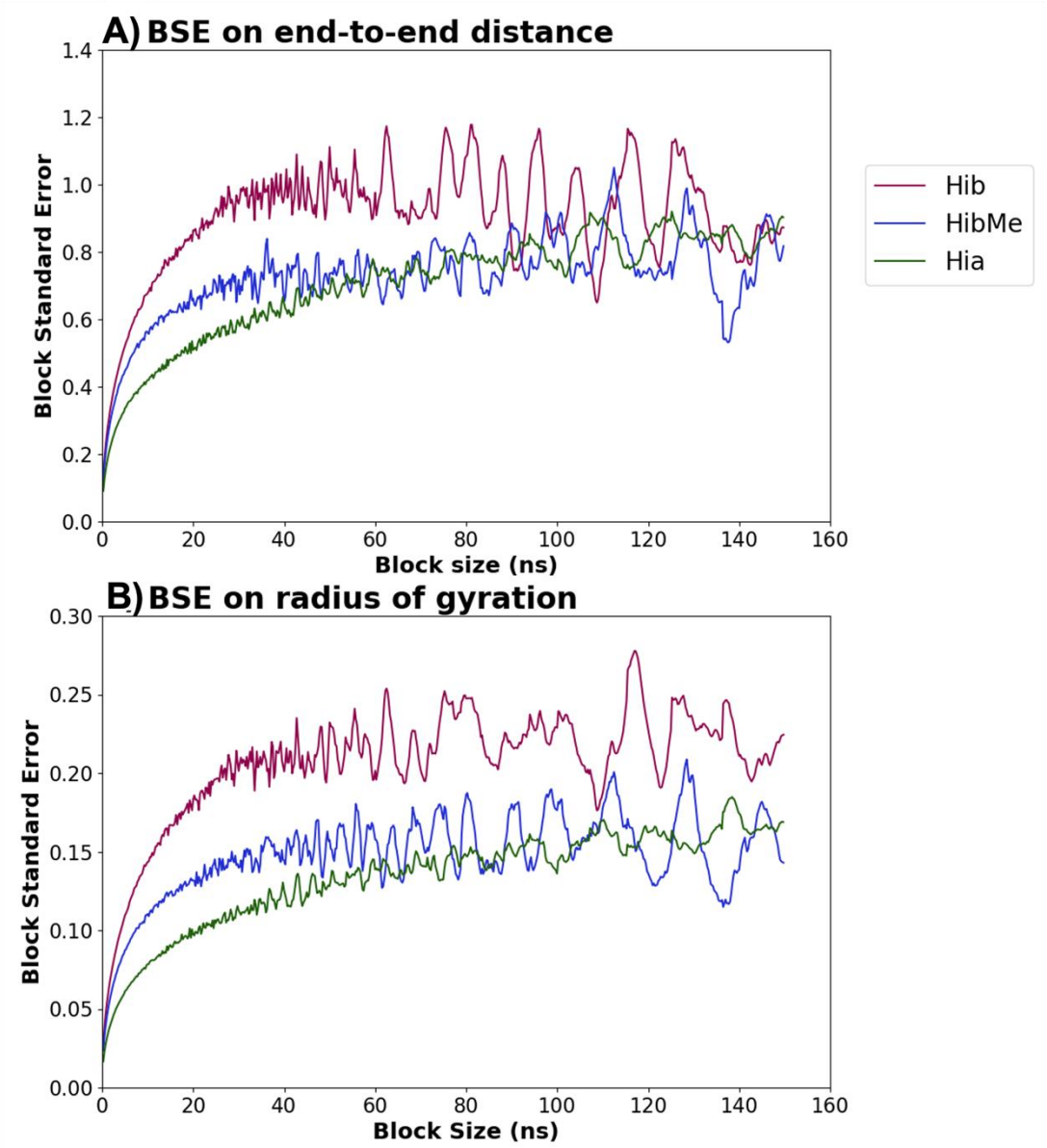
In contrast to Hib, the flexible Hia CPS forms relatively defined zig-zag conformations. We find significant differences between the Hia and Hib CPS antigens in flexibility, conformation, and minimal binding epitope, which does not support the existence of cross-reactivity between them. The immunological data confirms this: sera derived from Hia CPS conjugates cannot recognize nor kill Hib strains and sera derived from Hib CPS conjugates cannot recognize nor kill Hia strains. This is also consistent with clinical data showing an increase in invasive disease caused by serotype a in infants and children, even after receiving a full childhood serotype b conjugate vaccine regimen. [54,55] Based on the limited immunogenicity studies detailed herein, we also note that Hia appears to elicit a stronger immune response than the highly flexible Hib; it is interesting to postulate that the more conformationally defined Hia CPS is a stronger antigen than Hib.

Future work is focused on the development of a Hia glycoconjugate vaccine to protect all infants and children from this disease; funding has been secured for Phase 1 clinical trial of a prospective Hia vaccine by a North American consortium.

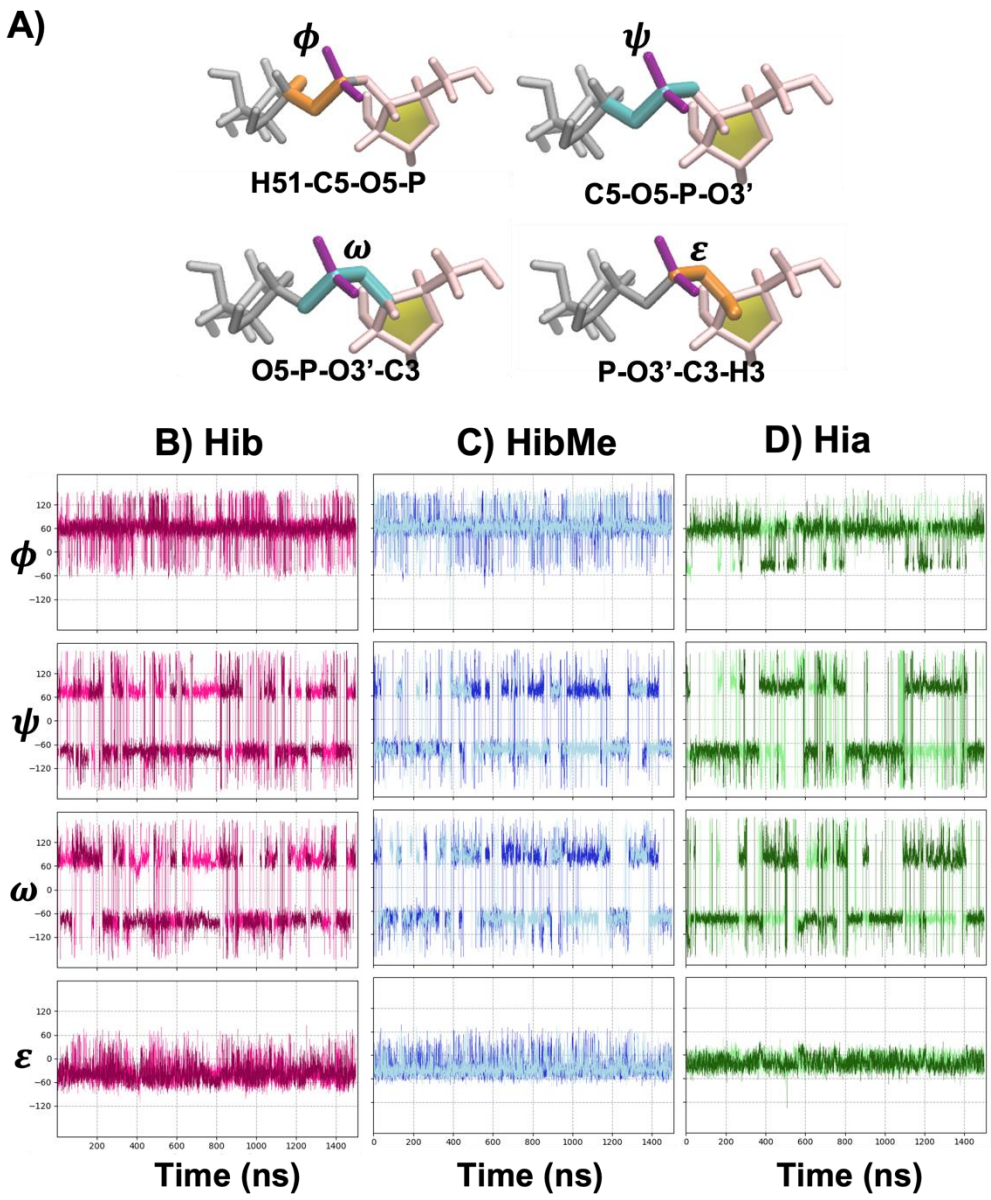
2.7 Supplementary Material



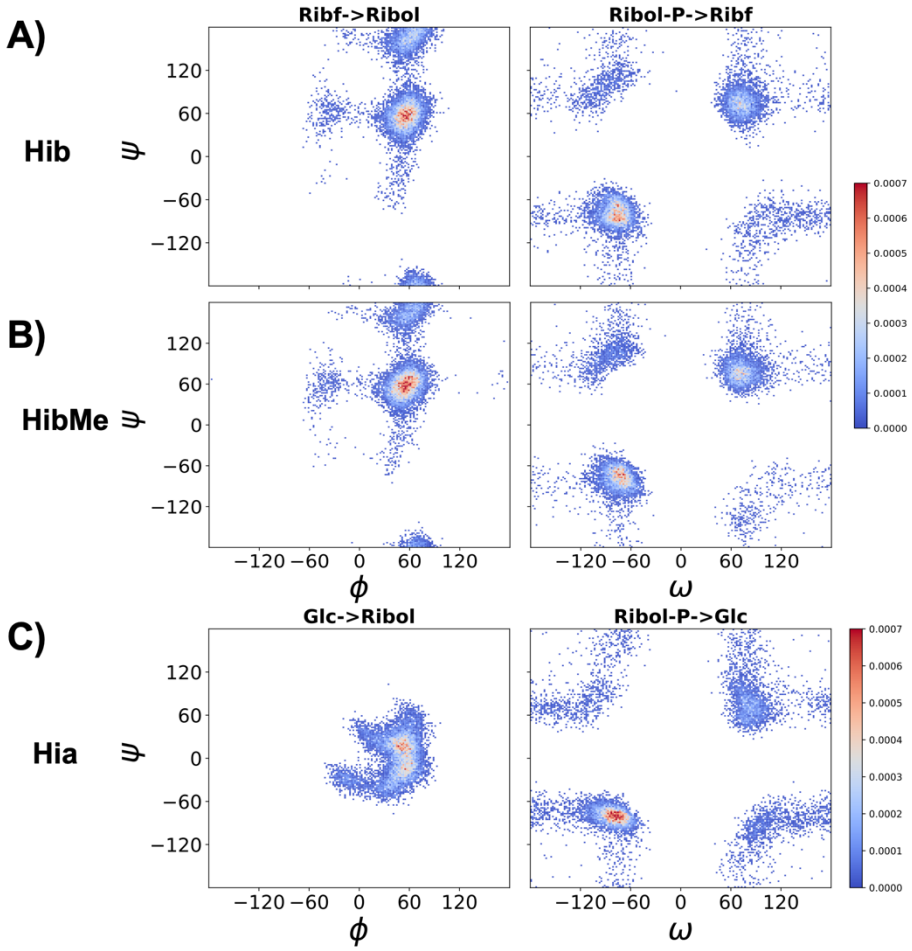
Supplementary Figure S2.1: *Starting conformations of the 6 RU CPS of:* ***(A)*** *Hib,* ***(B)*** *HibMe, and* ***(C)*** *Hia. The Licorice, VDW, and PaperChain representations are used. Structures were built using CarbBuilder, solvated and ionized using VMD, and minimized using NAMD. Water and counter ions are not shown.*



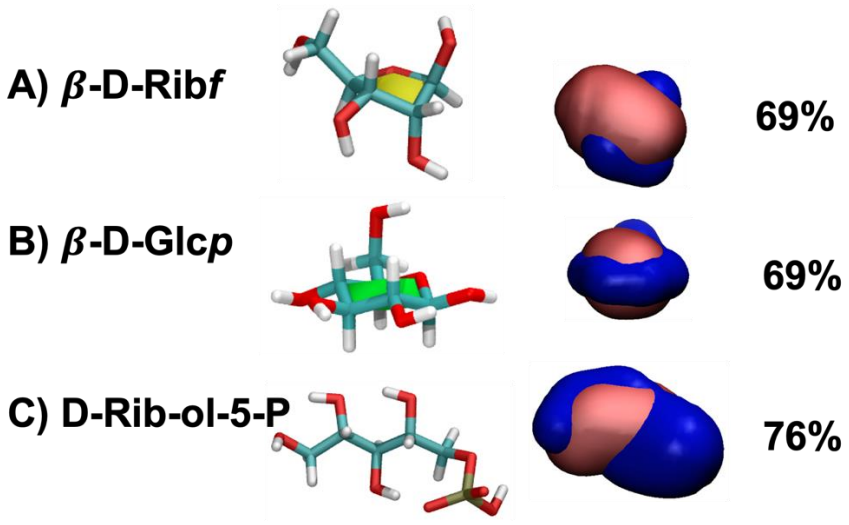
Supplementary Figure S2.2: *Block standard averaging analysis calculated for the time series of* ***(A)*** *end-to-end distance and* ***(B)*** *radius of gyration of the 6RU chains. The blocked standard error (BSE) reaches a plateau with increased block size (ns) for all the CPSs indicating convergence. Further analysis using the approximate BSE reveals correlation times < 25 ns (much less than the simulations times of 1 300 ns) and numbers of independent samples >> 1 which support the convergence of the simulation.*



Supplementary Figure S2.3: *Time series scatter plots for the phi (ϕ), psi (ψ), omega (ω), and epsilon (ε) dihedral angles of the phosphodiester linkage for* ***(B)*** *Hib,* ***(C)*** *HibMe, and* ***(D)*** *Hia. Points from both central repeating units (RU 3 and RU 4) were included to broadly sample the backbone behavior each (each assigned a different shade). We can see that the ω and ε angles are largely constant for all. The phosphodiester linkages were defined as in* ***(A)*** *by: ϕ = H_x-C_x-O_x-P, ψ = C_x-O_x-P-O_y, ω = O_x-P-O_y-C_y and ε = P-O_y-C_y-H_y as we have previously done for this type of linkage.* [48]



Supplementary Figure S2.4: *Heat map representation of scatter plots for the phi (ϕ), psi (ψ) dihedral angles of the glycosidic linkage (left column) and psi (ψ), omega (ω) dihedral angles for the phosphodiester linkage (right column) for:* ***(A)*** *Hib,* ***(B)*** *HibMe, and* ***(C)*** *Hia. Points from both central repeating units (RU 3 and RU 4) were included to broadly sample the backbone behavior. The color scale indicates the relevant occupancy of the dihedral angles during the simulation with red being high occupancy, and blue low occupancy. The glycosidic linkages (left column) are defined by two dihedral angles, ϕ = H_1-C_1-O_1-C_x' and ψ = C_1-O_1-C_x'-H_x', which are equivalent to ϕ_H and ψ_H in IUPAC nomenclature.* [47,48] *Only the variable ψ and ω angles are compared for the phosphodiester linkage (right column) since the ϕ and ε angles are very consistent and similar between all the CPSs* ***(Supplementary Figure S2.3)***



Supplementary Figure S2.5: *Licorice representation and solvent accessible surface area that is hydrophilic for static structure of monosaccharides* ***(A)*** *Ribf,* ***(B)*** *Glc, and* ***(C)*** *Rib-ol with hydrophilic (blue) and hydrophobic (pink) patches.*

2.8 References

1. N. I. Richardson, M. M. Kuttel, F. St Michael, C. Cairns, A. D. Cox and N. Ravenscroft, Cross-reactivity of Haemophilus influenzae type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies. *Glycoconj. J.*, 2021, **38**, 735–746.

2. A. S. Abdelhameed, G. G. Adams, G. A. Morris, F. M. Almutairi, P. Duvivier, et al, A glycoconjugate of Haemophilus influenzae Type b capsular polysaccharide with tetanus toxoid protein: hydrodynamic properties mainly influenced by the carbohydrate. *Sci. Rep.*, 2016, **6**, 1-11.

3. US Pat., US20190153015A1, 2019.

4. M. Emmadi, M. P. Lisboa, B. Monnanda, et al., A Liquid Stable Biological Active Semi-Synthetic Glycoconjugate Vaccine against Haemophilus influenzae type b, unpublished work.

5. R. S. W. Tsang and M. Ulanova, The changing epidemiology of invasive Haemophilus influenzae disease: Emergence and global presence of serotype a strains that may require a new vaccine for control. *Vaccine*, 2017, **35**, 4270-4275.

6. L. Barreto, A. D. Cox, M. Ulanova, M. G. Bruce and R. Tsang, The emerging Haemophilus influenzae serotype a infection and a potential vaccine: Implementation science in action. *Can. commun. dis. rep.*, 2017, **43**, 85-88.

7. M. P. E. Slack, Long Term Impact of Conjugate Vaccines on Haemophilus influenzae Meningitis: Narrative Review. *Microorganisms*, 2021, **9**, 886.

8. Defeating Meningitis by 2030: global road map, https://www.who.int/initiatives/defeating-meningitis-by-2030, (Accessed:03 May 2021).

9. L. Kenne and B. Lindberg, in *The polysaccharides*, ed. G. O. Aspinall, Academic Press, 1983, ch. Bacterial polysaccharides, pp. 287-363.

10. M. Ulanova, Invasive Haemophilus influenzae Serotype a Disease in the H. influenzae Serotype b Conjugate Vaccine Era: Where Are We Going? *Clin. Infect. Dis.*, 2020, **73**, e380–e382.

11. A. Zwahlen, J. S. Kroll, L. G. Rubin and E. R. Moxon, The molecular basis of pathogenicity in Haemophilus influenzae: comparative virulence of genetically-related capsular transformants and correlation with changes at the capsulation locus cap. *Microb. Pathog.*, 1989, **7**, 225-235.

12. A. D. Cox, D. Williams, C. Cairns, F. S. Michael, P. Fleming, et al, Investigating the candidacy of a capsular polysaccharide-based glycoconjugate as a vaccine to combat Haemophilus influenzae type a disease: a solution for an unmet public health need. *Vaccine*, 2017, **35**, 6129-6136.

13. M. Ulanova MD and R. S. W. Tsang Dr, Haemophilus influenzae serotype a as a cause of serious invasive infections. *Lancet Infect. Dis.*, 2014, **14**, 70-82.

14. A. D. Cox, R. K. Lee, M. Ulanova, M. G. Bruce and R. Tsang, *Proceedings of a workshop to discuss the epidemiology of invasive Haemophilus influenzae disease with emphasis on serotype a and b in the Americas, 2019*, Elsevier, 2021.

15. V. Verez-Bencomo, V. Fernandez-Santana, E. Hardy, M. E. Toledo, M. C. Rodríguez, et al, A synthetic conjugate polysaccharide vaccine against Haemophilus influenzae type b. *Science;* 2004, **305**, 522-525.

16. A. E. Zarei, H. A. Almehdar and E. M. Redwan, Hib Vaccines: Past, Present, and Future Perspectives. *J. Immunol. Res.*, 2016, **2016**, 7203587.

17. A. D. Cox, L. Barreto, M. Ulanova, M. G. Bruce, R. Tsang and on behalf of the conference contributors, *in Proceedings of Implementation science: Developing a vaccine for Haemophilus influenzae serotype a: Proceedings of a workshop*, ed. Anonymous, Canada Communicable Disease Report, 2017.

18. A. Shoukat, R. Van Exan and S. M. Moghadas, Cost-effectiveness of a potential vaccine candidate for Haemophilus influenzae serotype 'a'. *Vaccine,* 2018, **36**, 1681-1688.

19. X. Lemercinier and C. Jones, An NMR Spectroscopic Identity Test for the Control of the Capsular Polysaccharide from Haemophilus influenzae Type b. *Biologicals,* 2000, **28**, 175-183.

20. W. Egan, R. Schneerson, K. E. Werner and G. Zon, Structural studies and chemistry of bacterial capsular polysaccharides. Investigations of phosphodiester-linked capsular polysaccharides isolated from Haemophilus influenzae types a, b, c, and f: NMR spectroscopic identification and chemical modification of end groups and the nature of base-catalyzed hydrolytic depolymerization. *J. Am. Chem. Soc.*, 1982, **104**, 2898-2910.

21. Y. Ovodov, Bacterial capsular antigens. Structural patterns of capsular antigens. *Biochemistry*, 2006, **71**, 937-954.

22. A. W. Sturgess, K. Rush, R. J. Charbonneau, J. I. Lee, D. J. West, et al, Haemophilus influenzae type b conjugate vaccine stability: catalytic depolymerization of PRP in the presence of aluminum hydroxide. *Vaccine,* 1999, **17**, 1169-1178.

23. J. P. Hennessey Jr, P. Costantino, P. Talaga, et al, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. K. Prasad, ACS Publications, 2018, pp. 323-385.

24. J. Y. Baek, A. Geissner, D. C. K. Rathwell, D. Meierhofer, C. L. Pereira and P. H. Seeberger, A modular synthetic route to size-defined immunogenic Haemophilus influenzae b antigens is key to the identification of an octasaccharide lead vaccine candidate. *Chem. Sci.*, 2018, **9**, 1279-1288.

25. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

26. R. A. Byrd, W. Egan, M. F. Summers and A. Bax, New N.M.R.-spectroscopic approaches for structural studies of polysaccharides: Application to the Haemophilus influenzae type a capsular polysaccharide. *Carbohydr. Res.*, 1987, **166**, 47-58.

27. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of two capsular polysaccharides elaborated by different strains of *Haemophilus influenzae* type e. *Carbohydr. Res.,* 1981, **88**, 77-84.

28. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of the capsular polysaccharide elaborated by Haemophilus influenzae type d. *Carbohydr. Res.,* 1981, **97**, 285-291.

29. C. Hoog, A. Laaksonen and G. Widmalm, Molecular dynamics simulations of the phosphodiester-linked repeating units of the Haemophilus influenzae types c and f capsular polysaccharides. *J. Phys. Chem. B,* 2001, **105**, 7074.

30. M. Maestre and C. S. Pérez, Conformational analysis of diribosylribitol phosphate by NMR spectroscopy and molecular dynamics. *Magn. Reson. Chem.,* 2000, **38**, 123-125.

31. J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, Molecular Modeling of the Shigella flexneri Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a. *Vaccines,* 2020, **8**, 643.

32. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines,* 2021, **9**, 432.

33. E. A. Kabat, The nature of an antigenic determinant. *J. Immunol.,* 1966, **97**, 1-11.

34. N. Ravenscroft, G. Averani, A. Bartoloni, S. Berti, M. Bigio, et al, Size determination of bacterial capsular oligosaccharides used to prepare conjugate vaccines. *Vaccine,* 1999, **17**, 2802-2816.

35. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al, Scalable molecular dynamics with NAMD. *J. Comput. Chem.,* 2005, **26**, 1781-1802.

36. J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten, Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.,* 2007, **28**, 2618-2640.

37. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, et al, Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.,* 2008, **29**, 2543-2564.

38. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Comput. Chem.,* 2009, **5**, 2353-2370.

39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.,* 1983, **79**, 926-935.

40. M. M. Kuttel, G. E. Jackson, M. Mafata and N. Ravenscroft, Capsular polysaccharide conformations in pneumococcal serotypes 19F and 19A. *Carbohydr. Res.,* 2015, **406**, 27-33.

41. M. M. Kuttel, J. Ståhle and G. Widmalm, CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.,* 2016, **37**, 2098-2105.

42. W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.*, 1996, **14**, 33-38.

43. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.*, 1993, **98**, 10089-10092.

44. W. F. Van Gunsteren and H. J. Berendsen, A leap-frog algorithm for stochastic dynamics. *Mol. Simul.*, 1988, **1**, 173-185.

45. A. Grossfield and D. M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.*, 2009, **5**, 23-48.

46. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, 2007, **9**, 90-95.

47. R. S. Haltiwanger, Symbol nomenclature for glycans (SNFG). *Glycobiology*, 2016, **26**, 217.

48. S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, et al, Updates to the symbol nomenclature for glycans guidelines. *Glycobiology*, 2019, **29**, 620-624.

49. J. Hlozek, M. M. Kuttel and N. Ravenscroft, Conformations of Neisseria meningitidis serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.*, 2018, **465**, 44-51.

50. Gracia Luis, *WMC PhysBio Clustering*, Weill Cornell Medical College, 2012.

51. S. Cross, M. M. Kuttel, J. E. Stone and J. E. Gain, Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.*, 2009, **28**, 131-139.

52. M. Krone, J. E. Stone, T. Ertl and K. Schulten, *in Proceedings of Eurographics Conference on Visualization (EuroVis)* ed. M. Meyer and T. Weinkauf, Vienna Austria, 2012.

53. S. Zhang, M. Sella, J. Sianturi, P. Priegue, D. Shen and P. H. Seeberger, Discovery of oligosaccharide antigens for semi-synthetic glycoconjugate vaccine leads against Streptococcus suis serotypes 2, 3, 9 and 14. *Angew. Chem. Int. Ed.*, 2021, **60**, 14679-14692.

54. H. M. Soeters, S. E. Oliver, I. D. Plumb, A. E. Blain, T. Zulz, et al, Epidemiology of Invasive Haemophilus influenzae Serotype a Disease—United States, 2008–2017. *Clin. Infect. Dis.*, 2020, **73**, e371-379.

55. H. M. Soeters, A. Blain, T. Pondo, B. Doman, M. M. Farley, et al, Current epidemiology and trends in invasive Haemophilus influenzae disease—United States, 2009–2015. *Clin. Infect. Dis.*, 2018, **67**, 881-889.

3 Chapter 3

3.1 Preface

The following article “Conformational comparisons of *Pasteurella multocida* type B and E and structurally related capsular polysaccharides” describes a systematic molecular modeling study of the capsular polysaccharide (CPS) of the veterinary pathogen, *Pasteurella multocida* types B (PmB) and E (PmE) as well as *H. influenzae* types d (Hid) and e (Hie). [1]

Interest in this problem was sparked following the recent structural elucidation of the PmE and PmB CPSs and the finding of close structural similarity to the Hid and Hie CPSs. [2] This renewed interest in CPS as a target for new vaccines against PmE and PmB disease as the current whole-cell vaccines have limited efficacy. [3-5]

The CPSs all share a common, immunogenic amino-sugar disaccharide backbone with different linkage positions and vary in side group substitution with fructofuranose (an unusual monosaccharide in bacteria) and/or an amino acid moiety. [2,6,7] A stepwise systematic approach to the modeling of the four molecules was employed: starting with backbones and building up to fully substituted molecules. The aim of which was to fully explore conformational differences brought about by each of the subtle structural differences and investigate the potential for cross-reactivity.

These molecules showed high conformational definition which only increased with side-group addition (unlike the flexible *H. influenzae* types a and b studied in Chapter 2). We observe small structural differences having a significant impact on both the chain conformation and the exposed potential antibody-binding epitopes. A single backbone linkage difference (3-linked versus 4-linked) results in a change in helical handedness of the molecule presenting completely different potential binding epitopes.

The high conformational definition would be considered disadvantageous for the pathogen as high conformational definition presents a better (less variable) target for antibodies and would result in greater

immune recognition and clearance. However, we see the use of fructofuranose and/or glycine side chains to shield these conformationally defined CPS backbones that are also immunogenic due to the presence of amino-sugars. The side chains further act to disrupt potential binding epitopes – a proposed strategy for immune evasion.

Overall, based on our modeling and available immunological studies, we expect limited potential for cross-reactivity between *P. multocida* type B and E, suggesting that a bivalent CPS-based vaccine will be necessary to provide adequate protection against both serogroups.

In this work, the highly conformationally defined antigens showed significant conformational differences with a backbone shielding mechanism employed by the pathogen to disguise an immunogenic backbone. We considered the potential binding epitopes to propose mechanistic rationalization for the specific backbone structure employed and how it affects conformation and function.

Conformational comparisons of *Pasteurella multocida* type B and E and structurally related capsular polysaccharides

Nicole I. Richardson[a], Neil Ravenscroft[a], Michelle M. Kuttel[b*]

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

[*] Correspondence: Michelle Kuttel: mkuttel@cs.uct.ac.za

N.I. Richardson, N. Ravenscroft and M.M. Kuttel, Conformational comparisons of *Pasteurella multocida* types B and E and structurally related capsular polysaccharides, *Glycobiology*, 2023, https://doi.org/10.1093/glycob/cwad049.

3.2 Abstract

Pasteurella multocida, an encapsulated gram-negative bacterium, is a significant veterinary pathogen. *P. multocida* is classified into five serogroups (A, B, D, E, and F) based on the bacterial capsular polysaccharide (CPS), which is important for virulence. Serogroups B and E are the primary causative agents of bovine hemorrhagic septicemia, which is associated with significant yearly losses of livestock worldwide, especially in low- and middle-income countries. *P. multocida* disease is currently managed by whole-cell vaccination, albeit with limited efficacy. CPS is an attractive antigen target for an improved vaccine: CPS-based vaccines have proven highly effective against human bacterial diseases and could provide longer-term protection against *P. multocida*. The recently elucidated CPS repeat units of serogroups B and E both comprise a ManNAcA/GlcNAc disaccharide backbone with Fru*f* side chain, but differ in their glycosidic linkages, and a glycine side chain in serogroup B. Interestingly, the *Haemophilus influenzae* types e and d CPS have the same backbone residues. Here comparative modeling of *P. multocida* serogroups B and E and *H. influenzae* types e and d CPS identifies a significant impact of small structural differences on both the chain conformation and the exposed potential antibody-binding epitopes. Further, addition of Fru*f* and/or glycine side chains shields the immunogenic amino-sugar CPS backbone – a possible common strategy for immune evasion in both *P. multocida* and *H. influenzae*. The lack of common epitopes suggests limited potential for cross-reactivity between *P. multocida* type B and E, indicating that a bivalent CPS-based vaccine will be necessary in order to provide adequate protection against both serogroups.

3.3 Introduction

Pasteurella multocida is a widespread pathogenic encapsulated Gram-negative bacterium that primarily infects animals. [8,9] *P. multocida* causes a range of diseases in different host species: from fowl cholera in poultry, through hemorrhagic septicemia in cattle and buffaloes, atrophic rhinitis in swine, to snuffles in rabbits. Further, exposure to animals (e.g. through licks, bites, scratches) causes occasional severe zoonotic infections with *P. multocida* in humans. [4] Bovine hemorrhagic septicemia disease due to *P. multocida* is of particular concern due to the significant economic losses incurred worldwide, especially in low- and middle- income countries (LMICs) in Asia and sub-Saharan Africa where veterinary care is limited. [10] As antibiotics are only effective if started in the very early stages of disease – a challenge in environments with limited resources [5,10] – hemorrhagic septicemia is primarily managed through relatively inexpensive whole cell killed veterinary vaccines. However, these provide limited short-term protection of 6 - 12 months; [3-5] more effective vaccines against *P. multocida* could therefore have significant benefit.

P. multocida is classified into five serogroups (A, B, D, E, and F) based on the capsular polysaccharide (CPS) and 16 serovars based on cell wall lipopolysaccharide (LPS). [11,12] The capsular serogroups largely determine the hosts and diseases, with serogroups B and E associated with bovine hemorrhagic septicemia, serogroups A and F with fowl cholera, and serogroups D and A with porcine atrophic rhinitis. [4] The CPS is an attractive vaccine antigen target, as it is essential for virulence of *P. multocida* and inhibits phagocytic uptake. [13,14]

The recent structural elucidation of the *P. multocida* serogroup B (PmB, **Figure 3.1A**) and E (PmE, **Figure 3.1B**) [2] CPSs renews interest in the CPS as a target antigen for a bovine vaccine against *P. multocida*. The CPSs in these serogroups are structurally similar (**Table 3.1**). The repeating unit comprises N-acetyl-β-D-mannosaminuronic acid (ManNAcA), N-acetyl-β-D-glucosamine (GlcNAc) and β-D-fructofuranose (Fru*f*): ManNAcA and GlcNAc form a disaccharide backbone, with a Fru*f* side chain linked to ManNAcA. [2] The serogroups differ in three respects (indicated by shaded boxes in **Figure 3.1**): the

GlcNAc $\beta(1\rightarrow3)$ linkage in PmE is replaced with a $\beta(1\rightarrow4)$ linkage in PmB; the Fru*f* side chain $\beta(2\rightarrow4)$ linkage in PmE is replaced with a $\beta(2\rightarrow3)$ linkage in PmB; and PmB contains an N-linked glycine (Gly) side chain at C6 of ManNAcA.

Table 3.1. *Line Structures of P. multocida Type B (PmB) and E (PmE) as well as H. influenzae Type e (Hie) and d (Hid) Repeating Unit CPSs. (Side groups indicated in bold.)*

Organism	**Antigen repeating unit structure in CASPER format**
PmE	{→3) [***β*DFru*f*(2→4)**] *β*DMan*p*NAcA(1→3) *β*DGlc*p*NAc(1→}$_n$
PmB	{→4) [***β*DFru*f*(2→3)]** *β*DMan*p*NAcA6**Gly**(1→3) *β*DGlc*p*NAc(1→}$_n$
Hie	{→4) [***β*DFru*f*(2→3)**] *β*DMan*p*NAcA (1→3) *β*DGlc*p*NAc(1→}$_n$
Hid	{→3) *β*DMan*p*NAcA6**X***(1→4) *β*DGlc*p*NAc(1→}$_n$

* X represents a variable amino acid moiety: either L-alanine, L-serine, or L-threonine in the proportions 2:2:1.

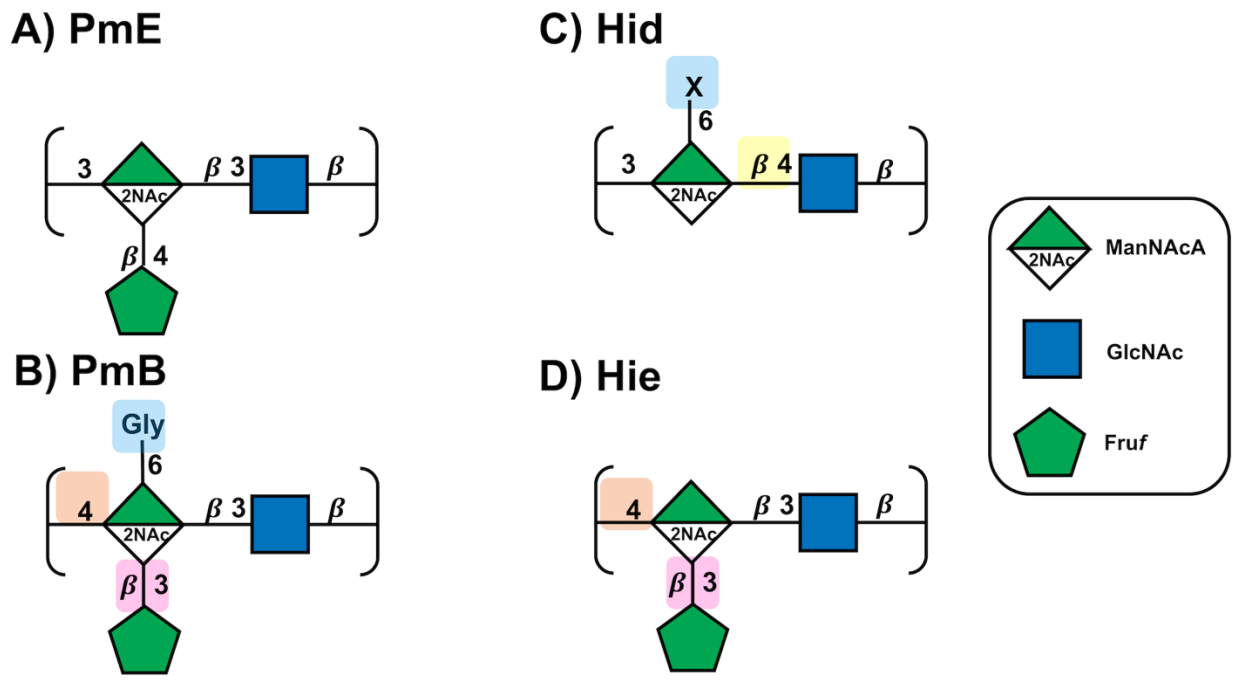


Figure 3.1: *CPS structures for P. multocida type E (PmE) and B (PmB) and H. influenzae type d (Hid) and e (Hie) represented with the SNFG (Symbol Nomenclature for Glycans) symbols.* [15-17] *Differences in repeating units relative to PmE are indicated by color shading of the background.*

Interestingly, PmE and PmB are structurally similar to the *Haemophilus influenzae* type e (Hie, **Figure 3.1C**) and d (Hid, **Figure 3.1D**) CPS, a possible example of convergent evolution. *H. influenzae* is a leading cause of meningitis, otitis media among other infections in children. [1] Type b was the primary cause of disease prior to the development of a vaccine, however, we are now seeing an increase in non-vaccine disease especially due to type a with Hie and type f also showing increases. [18] Hie is very similar to PmB, lacking only the glycine substituent [19,20] while Hid has a different backbone glycosidic linkage pattern (3-linked ManNAcA and 4-linked GlcNAc), lacks a Fru*f* substituent, and has a variable N-linked amino acid substitution pattern on C6 of ManNAcA (L-alanine, L-serine, L-threonine in proportion 2:2:1). [21,22]

An understanding of the potential for cross-reactivity (and hence cross-protection) between related target antigens – which is believed to have both structural and conformational aspects – may assist in the design of a minimal valency vaccine that provides maximum disease coverage. [23] The structural similarity between PmE and PmB points to the potential for cross-reactivity between these antigens, which would allow for a cheaper monovalent bovine vaccine. Early work using purified CPS in cattle challenge and passive mouse protection studies supports low levels of heterologous cross-reactivity and protection with good homologous serotype protection. [24-26] A more recent study, in which the CPS structures were elucidated for the first time, showed cross-reactivity between a PmB conjugate vaccine and PmE CPS, with limited cross-protection against PmE bacteria in mice. [2] The reciprocal study, however, was not performed and further work is still underway to extend these data. [2] In the absence of more detailed experimental information on cross-reactivity, molecular modeling can provide potentially useful information on which epitopes are exposed and are therefore likely to be recognized by antibodies. [1,27] While not a current vaccine target, modeling the structurally related Hie and Hid CPS provides insight into the conformational effect of structural differences.

Here we report a comparative modeling study to assess the conformational impact of the different structural features in the similar PmE, PmB, Hie, and Hid CPS antigens. We first compare the unsubstituted CPS

backbones (PmE Δ Fru*f*, PmB Δ Fru*f* Δ Gly, and Hid Δ X – where X is the variable amino acid described above) to determine the effect of changes in the glycosidic linkages on the chain conformation (**Figure 3.2**, top panel). Next, we investigate the impact of Fru*f* side chains on the backbone conformation (PmE and PmB Δ Gly (Hie), **Figure 3.2**, middle panel). Finally, the effect of the Gly amino acid side chain was established for PmB (**Figure 3.2**, bottom panel).

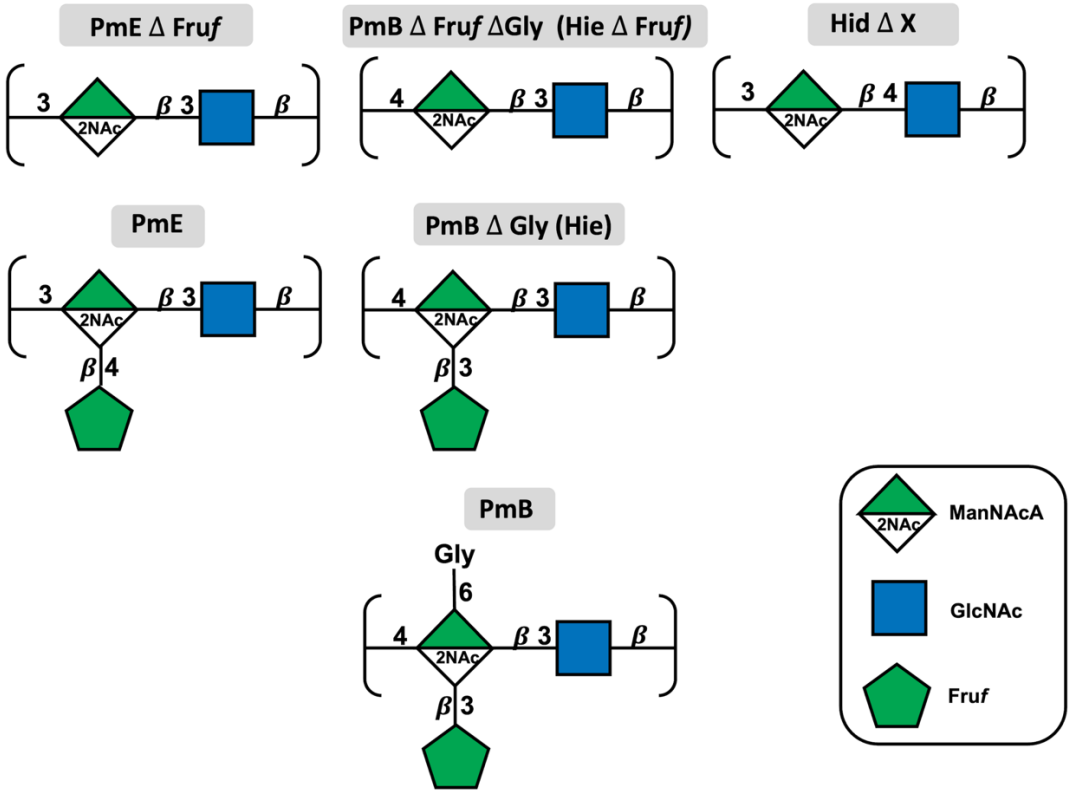


Figure 3.2: *CPS molecules of P. multocida type E (PmE) and B (PmB) and H. influenzae type d (Hid) and e (Hie) modeled in this work (SNFG symbol representation). The top panel shows the unsubstituted backbone molecules: PmE without Fruf (PmE Δ Fruf), PmB without Fruf and Gly (PmB Δ Fruf Δ Gly equivalent to Hie without Fruf) and Hid without the variable amino acid substituent (L-alanine, L-serine, or L-threonine in the proportions 2:2:1) represented by X (Hid Δ X). The middle panel shows the modeled Fruf substituted molecules: PmE wild type, PmB without Gly (PmB Δ Gly; equivalent to Hie wild type). The bottom panel shows the wild type PmB, with the Gly substituent (PmB).*

3.4 Materials and Methods

We followed our established systematic approach to the modeling of polysaccharides, as previously described. [1,27,28] For each of the six molecules of interest (**Figure 3.1**) we built CPS chains of three repeating units (3 RU) and six repeating units (6 RU).

Chain length is an important consideration when modeling CPSs, as a short chain may have insufficient molecular flexibility, while long chains are more computationally expensive to model. On the basis of our previous work, we consider a 6 RU chain to be representative of the behavior of the longer polysaccharide. Further, antibodies bind small fragments of the CPS between one and seven residues in length [29] corresponding to between one and four RU in the case of *P. multocida*, making a 6 RU model sufficient to explore antibody binding epitopes. We perform 3 RU simulations to ensure the behaviors of the short 3 RU and longer 6 RU chains are consistent, but here we present only the 6 RU results (3 RU results available in the Supplementary Material).

Molecular dynamics (MD) simulations in aqueous solution for each of the chains followed a protocol of initial system equilibration (200 ns) and then a production run of 800 – 1 800 ns (3 RU simulations) or 1 300 – 2 300 ns (6 RU simulations) respectively, as required for convergence. Data analyses were performed on these simulation trajectories, as described below.

3.4.1 Molecular dynamics

The 3 RU and 6 RU chains were built using our CarbBuilder software (version 1.2.42). [30,31] Starting structures for each molecule were built using low energy glycosidic linkage orientations obtained from potential mean force calculations (**Supplementary Figure S3.1**), as per our previous work [28] The psfgen tool was used to create protein structure files (PSF) for simulation with the CHARMM36 additive force field for carbohydrates. [32,33] For the PmB molecule with Gly we used the CHARMM36 protein force field parameters for glycine, which are compatible with the CHARMM36 additive force field. [34] We modeled

only the 6 RU model for this molecule. Starting structures were minimized with the Nanoscale Molecular Dynamics (NAMD) program (version 2.13) for 10 000 steps at 300 K. Minimized structures were solvated using the Visual Molecular Dynamics (VMD) software [35] solvation and ionization tools to add cubic TIP3P [36] water boxes of 50 Å per side for the 3 RU systems and 70 Å per side for the 6 RU systems. Systems were then neutralized with one sodium (Na^+) counter ion per repeating unit (three ions for the 3 RU, six for the 6 RU). Initial minimization and heating protocols comprised 5 K incremental temperature reassignments beginning at 10 K up to 300 K with 500 steps of NAMD minimization and 8 000 steps of MD at each temperature reassignment. Solvated and ionized structures (PDB and PSF files) for each 6 RU system are available as Supplementary Material.

Simulations of 3 RU and 6 RU were run using NAMD (version 2.13) [37] with CUDA extensions for graphics processor acceleration. [38]

Periodic boundary conditions equivalent to the cubic box size were employed for the solvated simulation with wrapping on. Long range electrostatics were implemented with the Particle Mesh Ewald summation grid spacing set to 1. [39] Atoms were not held fixed, and the initial center of mass motion was turned off. The 1-3 pairs were excluded from non-bonded interactions, 1-4 interactions were not scaled, and the dielectric constant was set to 1. Smoothing functions were applied to both the electrostatics and van der Waals forces with switching and cut-off distances of 10 Å and 12 Å, respectively.

A Leap-Frog Verlet integrator was used to integrate the equations of motion over a step size of 1 fs. A distance of 15 Å was used as the cut-off for inclusion in the pair list for calculation of non-bonded forces. The short-range non-bonded interactions were calculated every 1 fs, full electrostatics calculations were performed every 2 fs, and atoms were reassigned every 10 fs. [40]

Simulations were sampled under the isothermal-isobaric (nPT) ensemble. Langevin dynamics [41] were used to control the temperature with a damping coefficient of 5/ps. Nosé-Hoover Langevin piston dynamics were used as a barostat to maintain a target pressure of 1 atm. [42,43] Variable system volume was used with a piston

period of 100 fs and decay of 50 fs. Simulations of 1 500 ns (2 500 ns for PmB Δ Gly and PmB) were performed for the 6 RU systems comprising 200 ns of equilibration and 1 300 ns (2 300 ns for PmB Δ Gly and PmB) of production run as was required for convergence (**Supplementary Figure S3.2**). The 3 RU systems were run to 1 000 ns, except for PmE Δ Fru*f* which was run to 2 000 ns, as required for convergence (data available as **Supplementary**, **Figure S3.3** and **Figure S3.4**).

3.4.2 Convergence

We addressed convergence using block standard averaging [44] applied to two metrics: end-to-end distance and radius of gyration (**Supplementary Figure S3.2**). Block standard averaging was implemented with in-house Python scripts.

For all simulations, the blocked standard error (BSE) reached plateaus for both metrics, indicating convergence. The simulation lengths were large multiples of the correlation times for end-to-end distance (PmE Δ Fru*f*, 6.3 ns; PmB Δ Fru*f* Δ Gly, 9.24 ns; Hid Δ X, 5 ns; PmE, 1.4 ns; PmB Δ Gly, 47.7 ns; PmB, 101.63 ns) and radius of gyration (PmE Δ Fru*f*, 3.9 ns; PmB Δ Fru*f* Δ Gly, 11.7 ns; Hid Δ X, 7.9 ns; PmE, 0.9 ns; PmB Δ Gly, 45.6 ns; PmB, 97.4 ns). Further, the numbers of independent samples were >> 1 for both the end-to-end distance (PmE Δ Fru*f*, 237.8; PmB Δ Fru*f* Δ Gly, 162.3; Hid Δ X, 299.2; PmE, 1067.7; PmB Δ Gly, 52.5; PmB, 24.6) and the radius of gyration (PmE Δ Fru*f*, 385.9; PmB Δ Fru*f* Δ Gly, 127.7; Hid Δ X, 189.4; PmE, 1710.2; PmB Δ Gly, 54.9; PmB, 25.7). Our designated equilibration time of 200 ns is therefore greater than the correlation time indicating that the properties of the system are not correlated or related on this timescale and that the simulation length should be sufficient.

3.4.3 Data analysis

Molecular conformations were visualized using VMD, with the PaperChain and Twister visualization algorithms used to highlight carbohydrate rings and chains, [45] as required.

Trajectories were extracted at 25 ps intervals with analysis performed on frames 250 ps apart. Metrics, such as end-to-end distances, were extracted from the simulation trajectories using Tcl scripting via VMD's Tk console. Data analyses were performed with in-house Python scripts and plots generated using Matplotlib. [46]

3.4.4 Chain flexibility

The end-to-end distance, *r*, was measured from C1 of ManNAcA at the non-reducing end, to C3/4 of GlcNAc at the reducing end of each chain, thus excluding the highly flexible terminal residues.

3.4.5 Glycosidic linkages

Given the limited flexibility of the carbohydrate ring, the primary source of flexibility in the repeating unit arises from the glycosidic linkages between neighboring monosaccharides. These glycosidic linkages are conveniently described by two dihedral angles, ϕ and ψ. As per our previous work, we define ϕ = H_1-C_1-O_1-C'_x and ψ = C_1-O_1-C'_x-H'_x with x representing the linkage position. [28] In the case of the Fru*f* linkages which are (2→x) linked, these definitions were adjusted to ϕ = C_1-C_2-O_2-C'_x and ψ = C_2-O_2-C'_x-H'_x.

3.4.6 Conformational analysis

The most common chain conformations for each simulation were determined by clustering the production trajectory frames into families and calculating the relative occupancies of each family. Clusters comprising less than 6 % of the production run (post equilibration) were discarded. Clustering was performed using the WMC PhysBio plug-in for VMD's built-in measure cluster command. [47] Prior to clustering, the molecules were aligned on the RU 3 and RU 4 residues excluding the Fru*f*/Gly residue(s) and any hydrogens on the backbone. Clustering was then performed as an RMSD fit to the ring and linkage atoms of the central four repeating units of the chains excluding non-ring atoms, the Fru*f*/Gly residue(s), and the highly flexible terminal RU 1 and RU 6. Five clusters were created with a cut-off of 3 Å.

Clustering analysis was also performed on the central two repeating units, RU 3 and RU 4. Alignment was performed using the ManNAcA residue of RU 4 (excluding hydrogens) and was followed by clustering using the ring and linkage atoms of RU 3 and RU 4 excluding non-ring atoms and Fru*f*/Gly residue(s) creating three clusters with a cutoff of 1.1 Å, before discarding clusters comprising less than 6 % of the trajectory.

3.4.7 Solvent accessible surface

Hydrophilic/hydrophobic regions of the molecular surface were analyzed using VMD's built in "measure sasa" command. The solvent accessible surface area (SASA) analysis was performed by probing first hydrophilic regions (comprising hydroxyl groups, carbonyl groups, ring oxygens, nitrogen, and linkage oxygens) and then hydrophobic/neutral regions (comprising methyl groups, CH_2 groups, ring carbons, and ring protons) of the molecule using a van der Waal's radius of 2.5 Å – larger than that of water to imitate a potential small binding molecule. The ratio of hydrophilic SASA to total SASA was then calculated to determine the percentages of the hydrophilic and hydrophobic surface area exposed to solvent.

3.5 Results

We compare the effect of differing backbone linkages and substituents on the conformation and dynamics of the six CPS chains. We begin with a comparison of CPS chain extension and flexibility, followed by the molecular conformations and finally analyze the potential binding epitopes exposed on the saccharide chains.

3.5.1 Chain extension and flexibility

The molecular extension and flexibility of a carbohydrate is commonly measured by the fluctuation in the end-to-end distance, *r*, of the chain, during a simulation (**Figure 3.3A**). Times series plots of *r* for the simulations of 6 RU of each of the six CPS reveal considerable differences in chain extension and dynamics across the antigens (**Figure 3.3B-G**, left column). This is consistent with the trend observed for the 3 RU chains (**Supplementary Figure S3.3**).

The three unbranched CPS chains (**Figure 3.3B-D**) have similar flexibilities (σ = 5 for all three 6 RU molecules), with a wide range of conformations (see the trajectory snapshots at 50 ns intervals shown above the graphs). However, the backbones show considerably different chain extensions, as is clear from the *r* distribution histograms (**Figure 3.3**, center column). PmE Δ Fru*f*, comprising all $\beta(1\rightarrow3)$ glycosidic linkages, is the least extended chain, with a modal *r* value of 34 Å (**Figure 3.3B**). Replacement of either of the two glycosidic linkages in the backbone with a $\beta(1\rightarrow4)$ linkage increases the length of the RU and thus the overall chain extension: Hid Δ X (with a 4-linked GlcNAc) has an increased modal *r* value of 38 Å (**Figure 3.3D**) and PmB Δ Fru*f* Δ Gly (with a 4-linked ManNAcA) is the most extended chain with a modal *r* value of 45 Å (**Figure 3.3C**).

Addition of Fru*f* and/or Gly side chains to the ManNAcA backbone residue has little effect on the CPS chain extension, the modal *r* values remain very similar to the unsubstituted backbones: PmE 37 Å (**Figure 3.3E**), PmB Δ Gly 44 Å (**Figure 3.3F**) and PmB 45 Å (**Figure 3.3G**). Interestingly, side chains considerably

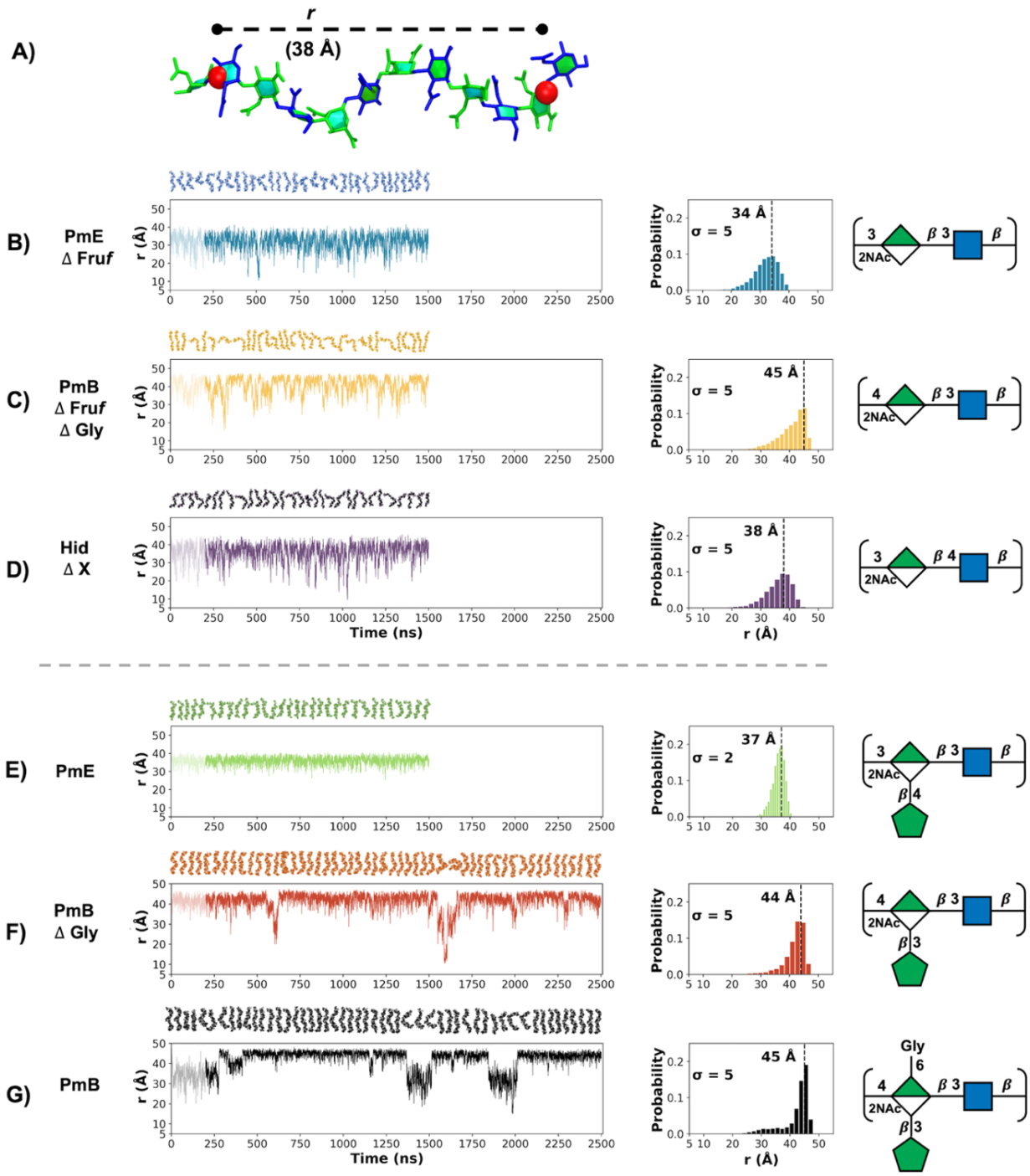


Figure 3.3: *Time series of the end-to-end distance, r, in the 6 RU CPS chains.* ***(A)*** *r, is measured (Å) from C3/C4 of GlcNAc at the reducing end to C1 of ManNAcA at the non-reducing end thus excluding the highly flexible terminal residues. Time series plots (left column) and corresponding histograms (center column) for the simulation trajectories are shown for:* ***(B)*** *PmE Δ Fruf,* ***(C)*** *PmB Δ Fruf Δ Gly,* ***(D)*** *Hid Δ X,* ***(E)*** *PmE,* ***(F)*** *PmB Δ Gly, and* ***(G)*** *PmB; the initial 200 ns (indicated by lighter color) are considered initial equilibration and the remaining trajectory is the production run. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine.*

Conformational snapshots at 50 ns intervals are shown above the time series plots; SNFG line structures are shown in the right column. The histograms are labeled with the standard deviations (σ) and modal peak r value.

decrease the flexibility of the $\beta(1\rightarrow3)$ linked PmE (σ = 2), but not the $\beta(1\rightarrow4)$ linked PmB molecules (σ = 5 for both). The decrease in PmE flexibility can be explained by increased steric hindrance: the addition of a side chain to the already highly substituted ManNAcA residue (2-NAc, 5-COOH substituted, and 3- or 4-linked to neighbor) results in all ring hydroxyls being substituted. Close comparison of the *r* time series plots in **Figure 3.3C** (PmB Δ Fru*f* Δ Gly) with **Figure 3.3F** (PmB Δ Gly) and G (PmB), shows a similar decreased flexibility of the chain on a ~250 ns time scale. However, on a 1 000 ns time scale, new populations of very low *r* values appear that maintain the overall chain flexibility on a level similar to the unsubstituted backbone. Interestingly, in the more substituted PmB these populations occur more frequently than in PmB Δ Gly.

The glycosidic linkages are the primary source of flexibility in the CPS chains. Side chains on ManNAcA reduce the overall flexibility of the chains through impeding rotation of the ManNAcA→GlcNAc glycosidic linkage through steric clashes between neighboring residues; rotation of the GlcNAc→ManNAcA is largely unaffected (**Supplementary Figure S3.5** and **Table S3.1**).

3.5.2 CPS chain conformations

For all the 6 RU CPS the primary conformations are well defined helices (**Figure 3.4**), which dominate for more than 50 % of simulation time in all cases (minor conformational clusters and the cluster occupancy are shown in the Supplementary Material, **Figures S3.6** and **S3.7**) which is consistent with the conformations observed in the 3 RU chains (**Supplementary Figure S3.2**). The helices, however, differ in their handedness according to the constituent glycosidic linkages: the $\beta(1\rightarrow3)$ linked PmE (**Figure 3.4A**) and the PmE Δ Fru*f* unsubstituted backbone (**Figure 3.4B**) form right-handed helices, whereas the molecules containing a $\beta(1\rightarrow4)$ glycosidic linkage — PmB Δ Fru*f* Δ Gly (**Figure 3.4C**), PmB Δ Gly

(**Figure 3.4D**), PmB (**Figure 3.4E**), and Hid Δ X (**Figure 3.4F**) — all form left-handed helices. Further, the PmE molecules are more compact than the PmB molecules.

In all cases, addition of Fru*f* and/or Gly side chains reduces the flexibility of the CPS backbone. For PmE, a Fru*f* side chain increases the occupancy of the primary helical conformation from 51 % (PmE Δ Fru*f*, **Figure 3.4A**) to 77 % (PmE, **Figure 3.4B**), and results in a more extended helix; the modal *r* distance increases from 34 Å to 37 Å (**Figure 3.3**). Similarly, the 53 % occupancy of the main helical conformation of PmB Δ Fru*f* Δ Gly (**Figure 3.4C**) increases to 80 % in PmB Δ Gly (**Figure 3.4D**), although in this case the helices have similar extensions. The addition of a Gly side chain in PmB (**Figure 3.4E**) does not alter the backbone conformation or flexibility of the primary helical conformation. Further, the Fru*f* (**Figure 3.4B**, **D** and **E**, magenta) or Gly (**Figure 3.4E**, orange) side chains are highly solvent exposed and thus present potential targets for antibody binding, while also shielding the backbone residues.

The alignment of the hydrophobic (NAc, grey) and hydrophilic (COOH, red) residue substitutions differs in the three backbones, resulting in a considerable difference in surface hydrophilicity (**Supplementary Figure S3.8**). For PmE Δ Fru*f* (**Figure 3.5A**) the NAc and COOH moieties are well separated on the backbone, whereas PmB Δ Fru*f* Δ Gly has frequent shielding of the groups of ManNAcA COOH by the NAc on an adjacent GlcNAc residue resulting in a considerably less hydrophilic surface. However, the Fru*f* side chains with their many exposed hydroxyl groups increase the hydrophilicity of the molecular surface (**Figure 3.5**, Quicksurf visualizations). This is particularly in the case of PmB (**Supplementary Figure S3.8**), which has the potential to impact antibody binding.

As discussed above, the simulations of PmB Δ Gly and PmB show minor populations with short *r* distances. These populations are associated with bent conformations of the CPS chain. The bent conformations align adjacent NAc groups (**Figure 3.5A E** and **F**), which may assist with conformational stabilization.

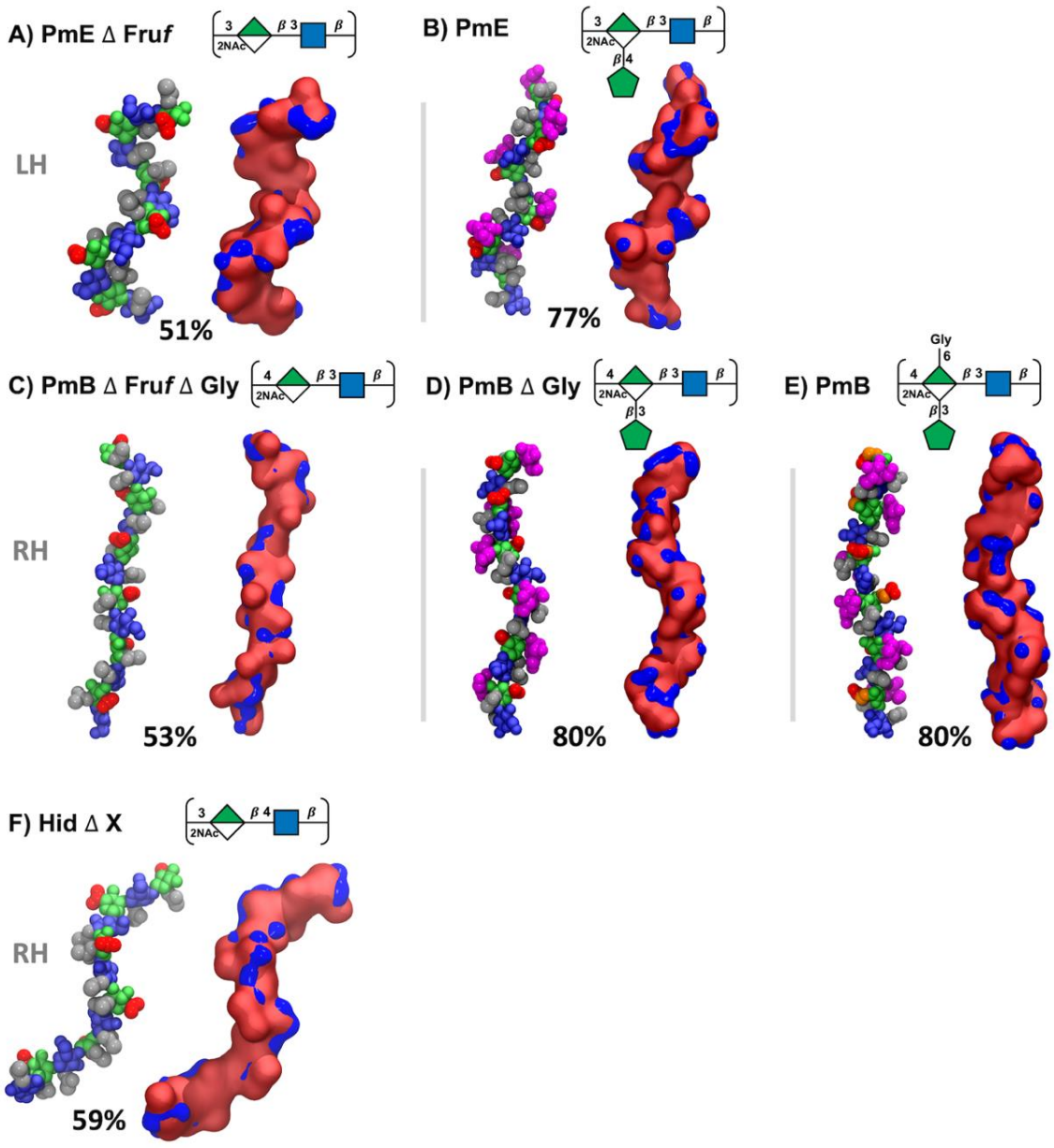


Figure 3.4: *Primary conformational families (seed frame for each cluster is shown) determined by cluster analysis of RU 2 to RU 5 of the CPS for* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmE,* ***(C)*** *PmB Δ Fruf Δ Gly,* ***(D)*** *PmB Δ Gly,* ***(E)*** *PmB, and* ***(F)*** *Hid Δ X. The handedness of the helix is indicated on the left of the figure by LH (left-handed) or RH (right-handed). X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. Conformations are visualized with the VDW (van der Waal; left) and Quicksurf representations (right). VDW representations were colored as for SNFG symbols: ManNAcA green, GlcNAc blue, NAc grey, COOH red, Fruf magenta, and Gly orange. For the Quicksurf representation, hydrophobic and neutral atoms were colored red, hydrophilic atoms blue.*

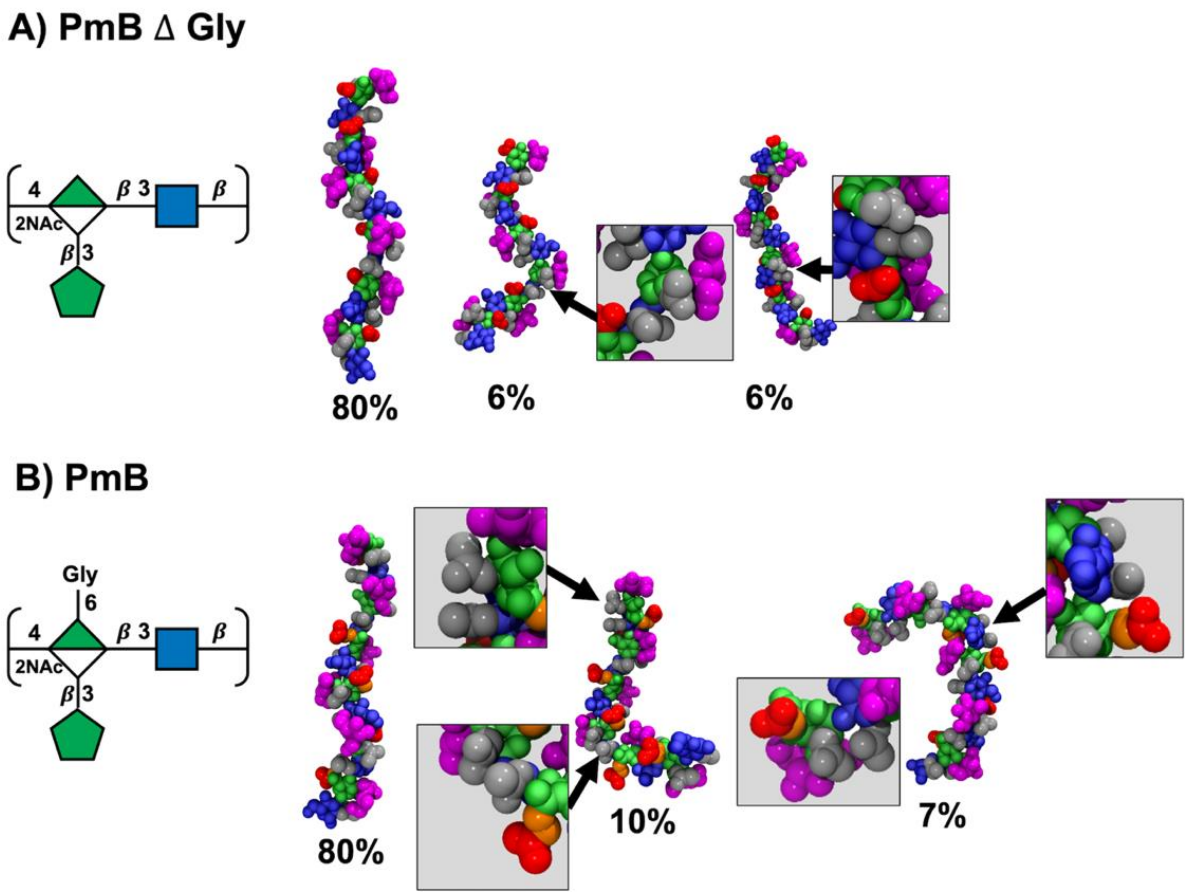


Figure ***3.5:*** *The main conformational families (seed frame for each cluster is shown) of* ***(A)*** *PmB Δ Gly and* ***(B)*** *PmB with NAc pairing of the non-primary clusters indicated in zoomed in boxes. The VDW (van der Waal) representation was used with colors as per SNFG: ManNAcA residues are shown in green, GlcNAc residues in blue, NAc groups in grey, COOH groups in red, Fruf in magenta, and Gly in orange.*

3.5.3 Epitopes

As the CPS regions bound by antibodies comprise one to seven residues, it is useful to compare the 6 RU CPS chains on this length scale. **Figure 3.6** focuses on the principal conformations of the two central repeat units (RU 3 and RU 4) for each of the six molecules; potential binding epitopes (Ep) are indicated by shaded boxes. There are key differences in the alignment of the side chains and residue substitutions, forming a range of different epitopes (despite a relatively conserved backbone). In the following analysis, we focus on substitutions on opposing sides of the CPS chain, which we term A and B.

For side A, the PmE Δ Fru*f* molecule (**Figure 3.6A**) has a hydrophobic epitope exposing three partially aligned N-acetyl groups, EpA-1. On side B, EpB-1 comprises the exposed COOH (red), of C6 on ManNAcA, and the N-acetyl on GlcNAc in the following RU. In PmE, the Fru*f* substitution considerably changes side A, forming the EpA-2 epitope with the three aligned NAc groups seen in EpA-1 now shielded by the Fru*f* residue. Similarly, on side B, the Fru*f* shields the exposed COOH forming the EpB-2 epitope (the NAc would also be shielded by the adjacent Fru*f* as in PmE the Fru*f* residues alternate faces). Conformationally, we see increased conformational definition as the occupancy increases from 56 % in PmE Δ Fru*f* to 83 % in PmE.

In PmB Δ Fru*f* Δ Gly (**Figure 3.6C**) side A of the chain is considerably altered (compared to PmE Δ Fru*f*), with one NAc replaced by a hydrophilic carboxylic acid group and a further NAc joining the group, forming EpA-3. Side B remains similar to the PmE molecules, though, with the ManANAcA N-acetyl replacing the GlcNAc N-Acetyl – forming EpB-1'. Fru*f* substitution in PmB Δ Gly (**Figure 3.6E**) completely disrupts the molecular surface presented on both sides of the CPS with shielding Fru*f* residues forming epitopes EpA-4 and EpB-3. Conformationally, we see the occupancy increase from 47 % to 51 %. Additional Gly substitution in PmB (**Figure 3.6F**) replaces COOH groups with bulkier Gly and thus further alters both sides of the chain, forming epitopes EpA-5 and EpB-4. The addition of Gly further increases the occupancy to 70 % and suggests that for this longer, less compact, helix an additional side group is required to achieve adequate shielding.

Interestingly, Hid Δ X (**Figure 3.6C**) has altered both sides of the chain, forming EpA-6 and EpB-5. On side A, the NAc pairs are sandwiched by COOH groups on either side with ¼-½ turn between NAc and COOH groups; on the mostly unsubstituted side B, the NAc of the ManNAcA is positioned where the COOH is in PmE and PmB. The central paired NAc in EpA-6 may be similar enough to EPA-1 to allow for antibody cross-reaction.

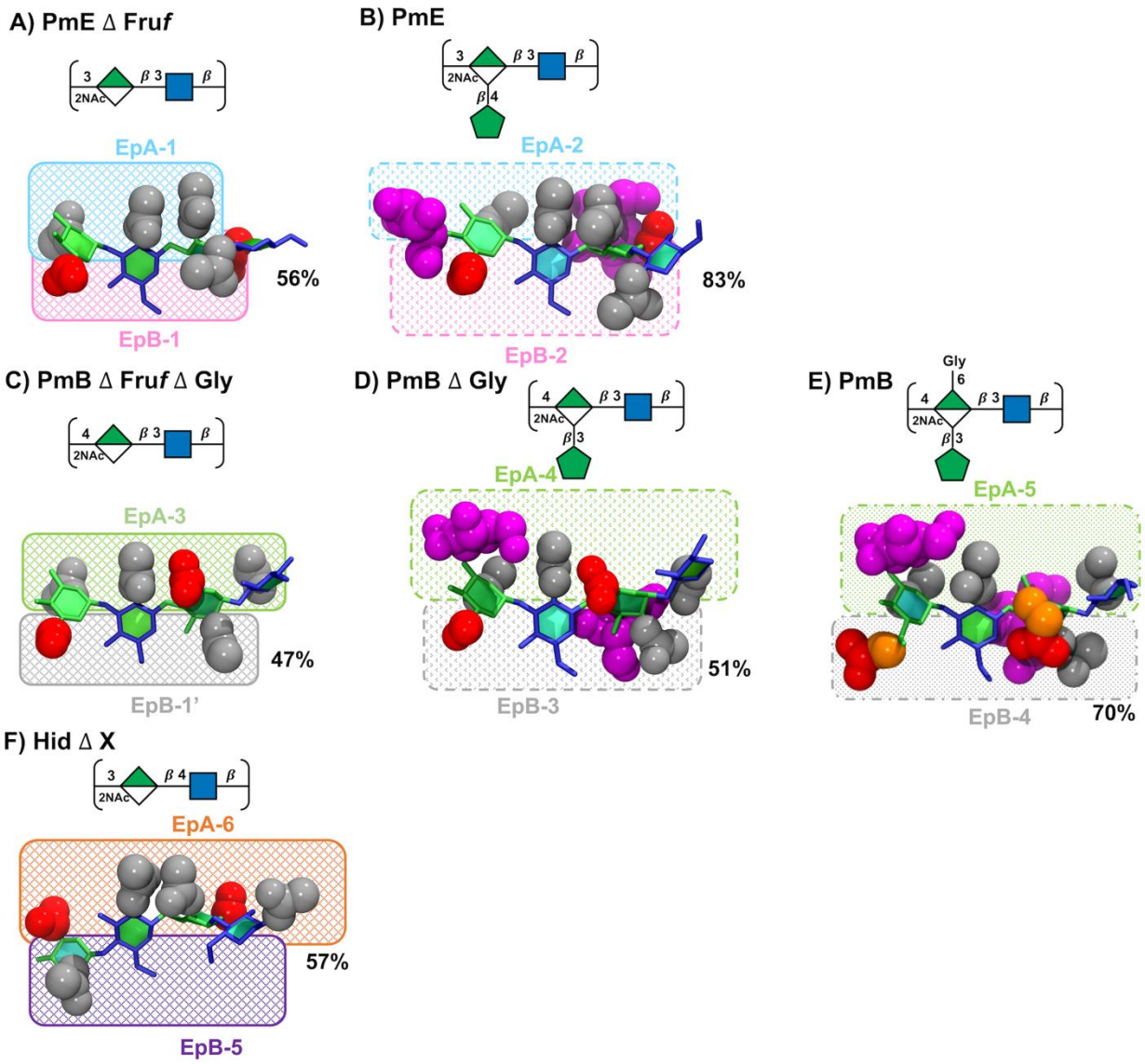


Figure 3.6: *Main conformations and associated percentages identified from clustering of RU 3 and RU 4 only with conformational epitopes (Ep) indicated by shaded boxes for:* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmE,* ***(C)*** *PmB Δ Fruf Δ Gly,* ***(D)*** *PmB Δ Gly,* ***(E)*** *PmB, and* ***(F)*** *Hid Δ X. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. The VDW (van der Waal) representation was used with colors as per SNFG: ManNAcA residues are shown in green, GlcNAc residues in blue, NAc groups in grey, COOH groups in red, Fruf in magenta, and Gly in orange.*

3.6 Discussion

This study of six structurally similar CPS provides further evidence that small structural differences in a carbohydrate chain can have a large impact on both molecular conformation and the epitopes exposed for antibody binding. Although *P. multocida* type E and B as well as the related *H. influenzae* type d and e CPS molecules all have well-defined helical molecular conformations, these studies predict that the presence of $\beta(1\rightarrow4)$ glycosidic linkages in the backbone changes both the handedness of the helix and increases the extension of the helices relative to chains with only $\beta(1\rightarrow3)$ linkages. Further, side chain substitution of the backbone with Fru*f* decreases the molecular flexibility, with a concomitant increase in conformational definition. This has the potential to increase the antigenicity of the CPS fragments, as a more conformationally defined molecule presents a better target for antibodies. However, side chain substitutions also considerably alter the binding epitopes, and potentially shield the ManNAcA-GlcNAc backbone from antibody binding with the longer, more accessible serotype B helices requiring more than one side group to achieve the same level of backbone shielding as those of serotype E. The ManNAcA and GlcNAc amino sugars are expected to be immunogenic. The ManNAcA residue is often found with other amino sugars and is a motif commonly expressed by pathogens. [48-50] Further, mannose binding proteins found in the lungs and serum of *Bovidae* (such as surfactant proteins, mannose binding proteins, and collectins) are highly specific for fucose, mannose, N-acetyl-D-mannosamine, N-acetyl-D-glucosamine, and glucose, playing an important role in immune activation and bacterial clearance. [51-53] Therefore, we expect that the highly exposed Fru*f* and Gly substituents may be used to disguise an immunogenic backbone.

Our simulations suggest very limited cross-reactivity between any of the six CPS we modeled, in line with the immunological data currently available. Early studies using whole-cell killed vaccines as well as CPS extracts demonstrated limited cross-protection between serotypes B and E. [24-26] Elucidation of the CPS structures containing the acid-labile Fru*f* [2] allows interpretation of these experiments as well as the systematic modeling studies presented here. The significant differences that we have found in the CPS backbone conformations and lack of common epitopes between the CPS predict limited cross-reactivity

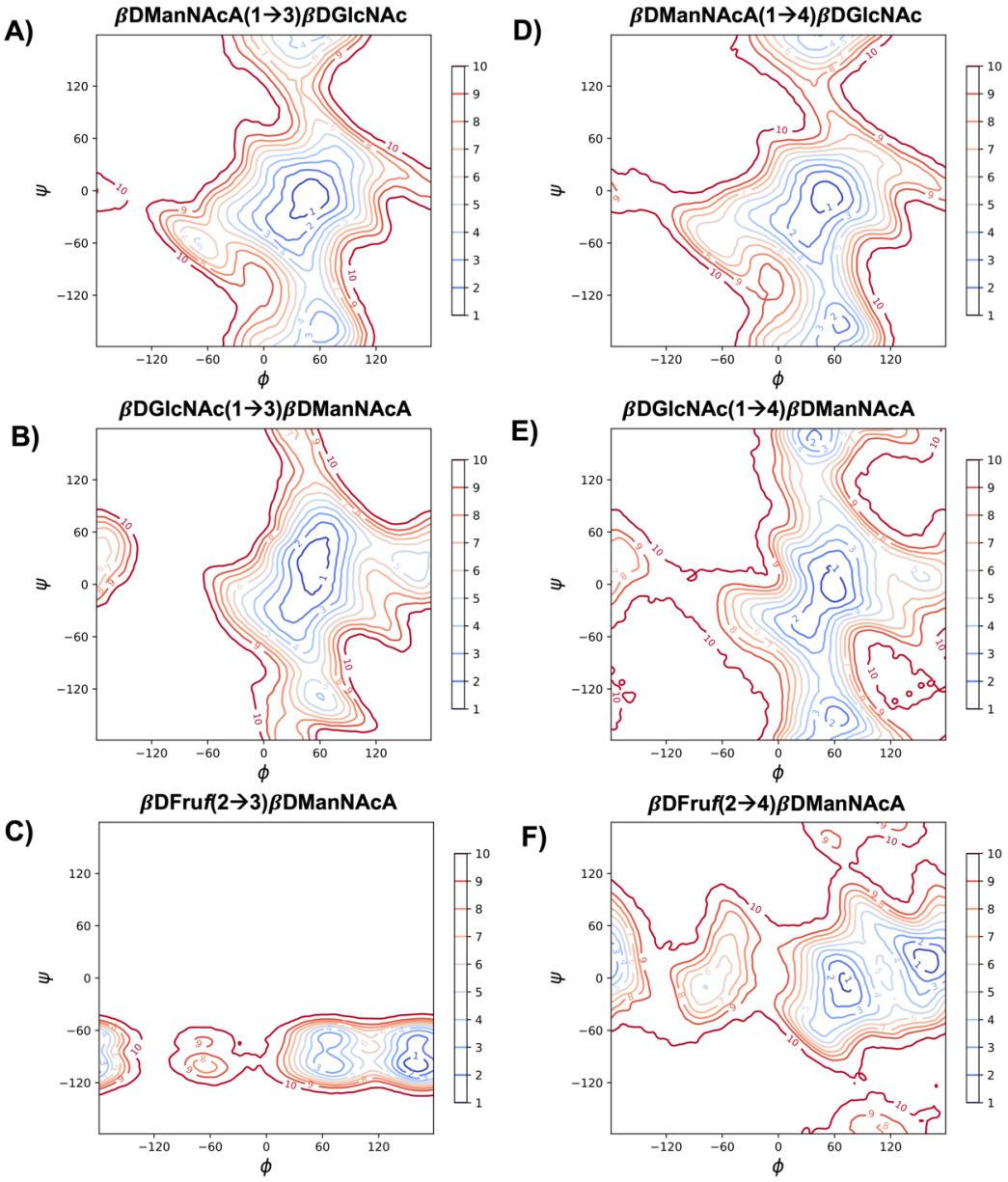
and cross-protection these two serotypes, therefore supporting the development of a bivalent CPS-based vaccine. A bivalent vaccine has been found to be effective against a type E challenge in cattle with passive mice protection against both type E and B, suggesting the cattle should also be protected against type B. [26]

For the closely related *H. influenzae* type e (Hie) and d (Hid) CPSs we also suggest a lack of cross-reactivity as the Hid Δ X CPS has very different conformations and potential binding epitopes from the Hie (PmB Δ Gly) and, most notably, the absence of the Fru*f* residue in Hid has an impact on the binding surface of the molecule as well as on the potential epitopes. This may also account for Hid being a less successful pathogen: the lack of Fru*f* side chain results in more exposure of the immunogenic backbone. Low levels of cross-reactivity between Hie and Hid may occur; with limited cross-protection. Although the variable amino acid substituents of Hid (L-alanine, L-serine, L-threonine) were not modeled here, we expect a similar result with the amino acids having little impact on conformation but shielding the backbone although potentially less effectively than Fru*f*.

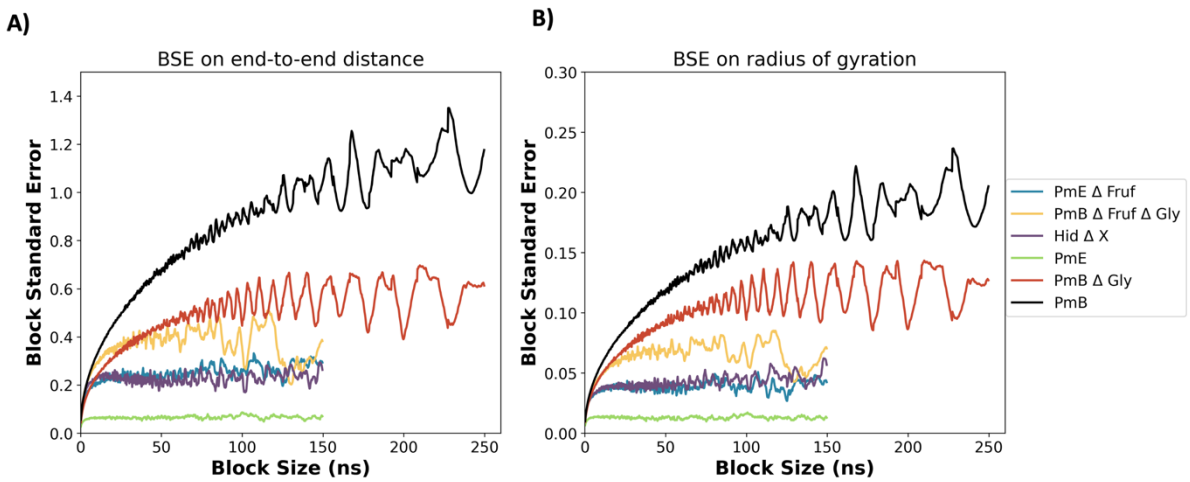
Our modeling investigation of *Streptococcus pneumoniae* serogroup 10 found that strongly cross-reactive serotypes share common epitopes. [27] Here we find no evidence for common epitopes between the six CPSs studied, which suggest little potential for cross-protection between them. This is supported by early immunological studies with PmE and PmB CPSs that showed low levels of limited cross-protection. However, direct structural information on the interactions of CPSs with antibodies is required for corroboration of our prediction that the lack of cross-protection between the CPS is due to an absence of shared epitopes. Unfortunately, molecular binding studies of CPSs with bactericidal antibodies are not currently available and are rarely feasible as they require significant resources to perform the challenging isolation and characterization of monoclonal antibodies against the specific antigen of interest. Furthermore, structural determination is required in order to generate crystal structure files that are necessary to perform in-silico molecular binding experiments. To date this has been performed for Group B Streptococcus and some meningococcal serotypes through a vaccine manufacturer partnership with several research groups. [54,55] In the absence of direct structural data, immunological studies using reciprocal

PmB and PmE CPS (or, preferably, the CPS conjugate vaccines and structural mutants modeled here) would assist in determining the importance of specific structural moieties/epitopes on the CPS for cross-protection. Such validation would be extremely valuable, as elucidation of the structural basis for cross-protection would greatly aid in the rational development of more effective CPS-based vaccines against *Pasteurella multocida*.

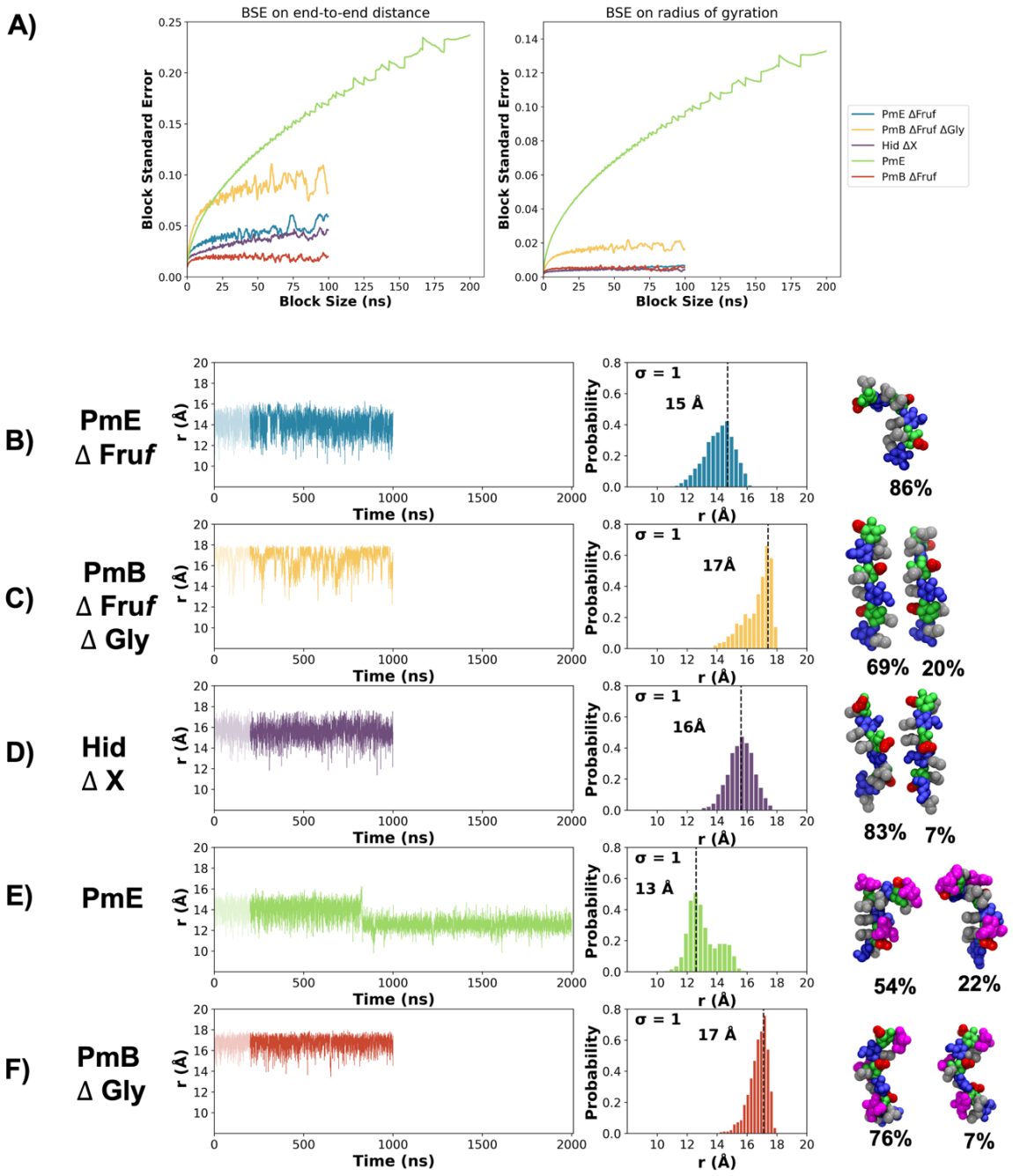
3.7 Supplementary Material



Supplementary Figure S3.1*: Contour plots for the PMF calculations of the disaccharide glycosidic linkages for the modeled molecules. The left column displays the 1→3 linked PMFs for:* ***(A)*** *ManNAcA(1→3)GlcNAc,* ***(B)*** *GlcNAc(1→3)ManNAcA,* ***(C)*** *Fruf(2→3)ManNAcA. The right column displays the 1→4 linked PMFs for:* ***(D)*** *ManNAcA(1→4)GlcNAc,* ***(E)*** *GlcNAc(1→4)ManNAcA,* ***(F)*** *Fruf(2→4)ManNAcA. Higher energy is represented by redder lines and lower energy by bluer lines.*

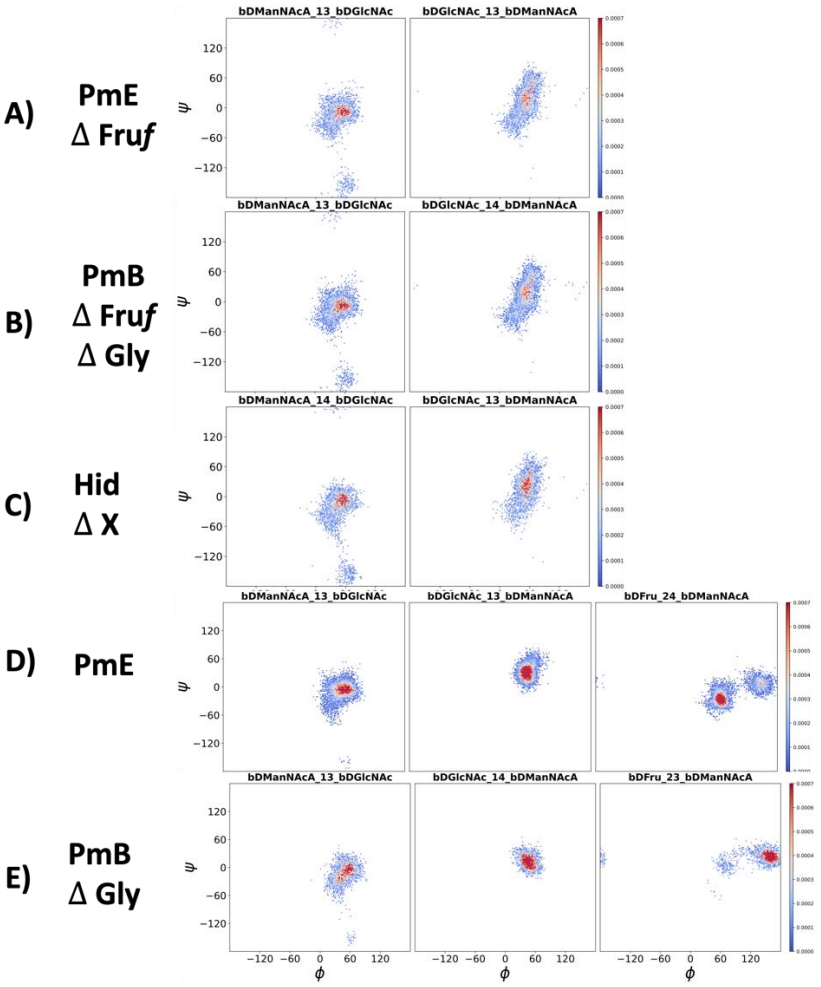


Supplementary Figure S3.2: *Block standard averaging analysis for 6 RU modeled P. multocida and H. influenzae CPS molecules.* ***(A)*** *block standard error (BSE) versus block size (ns) calculated on end-to-end distance.* ***(B)*** *BSE versus block size calculated on radius of gyration. For all molecules, the BSE visually reaches a plateau.*

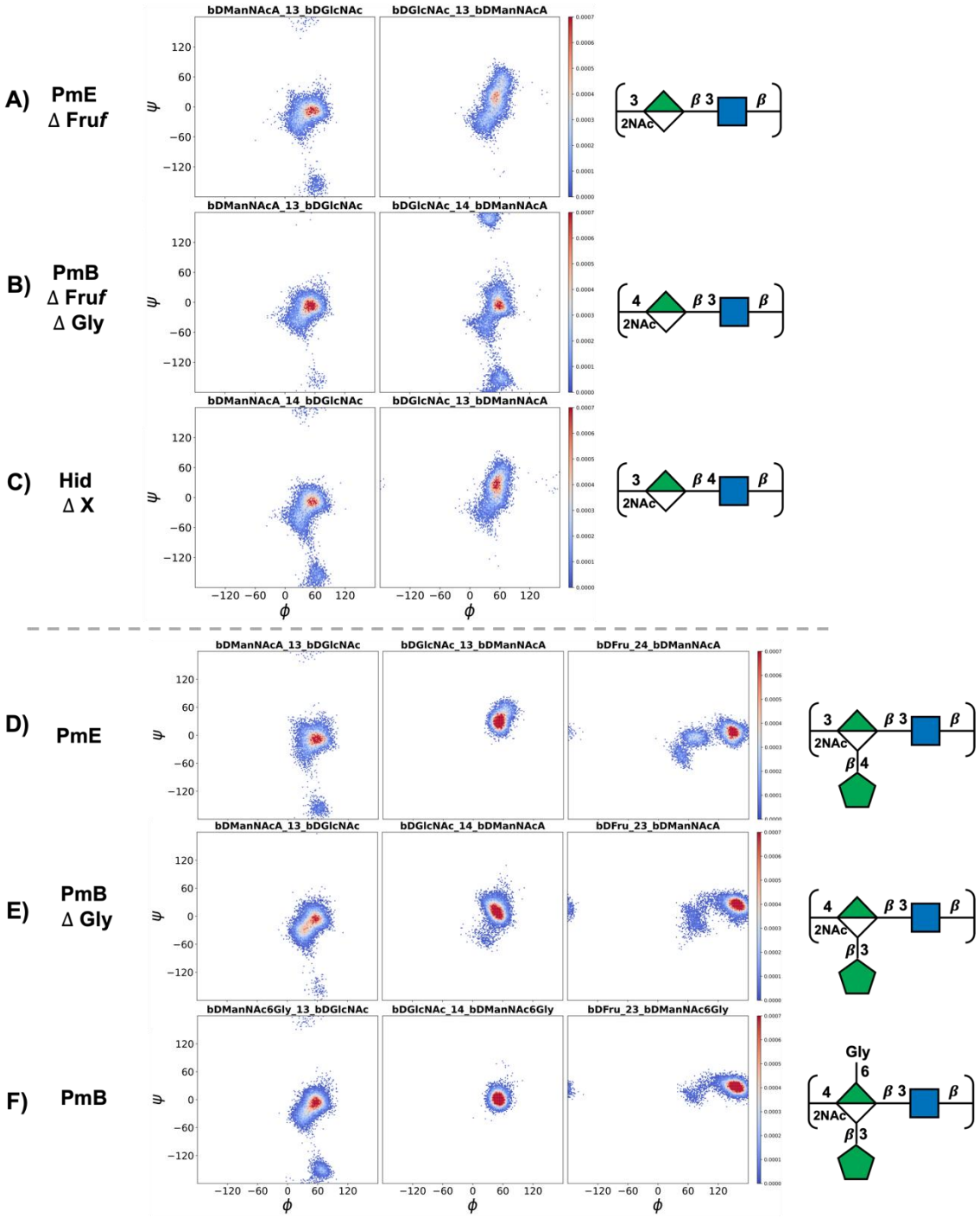


Supplementary Figure S3.3: (A) *Block standard averaging analysis for modeled 3 RU P. multocida and H. influenzae CPS molecules with block standard error (BSE) versus block size (ns) calculated on end-to-end distance in the left column and BSE versus block size calculated on radius of gyration in the right column. For all molecules, the BSE visually reaches a plateau. Time series of the end-to-end distance, r, in the 3 RU CPS chains with r defined as the distance (Å) from C3/C4 of GlcNAc of RU 1 to C1 of ManNAcA in RU 3. Time series plots (left column) and corresponding histograms (right column) for the simulation trajectories as well as main clusters are shown for:* ***(B)*** *PmE Δ Fruf,* ***(C)*** *PmB Δ Fruf,* ***(D)*** *Hid Δ X,* ***(E)*** *PmE, and* ***(F)*** *PmB Δ Gly. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. For each trajectory the initial*

200 ns (lighter coloring) are considered initial equilibration and the remaining trajectory is the production run. The histograms are labelled with the standard deviations (σ) and modal peak r value. For the clustering analysis, alignment was performed on the central (RU 2) backbone residues (excluding Fruf) and clustering was performed on the ring and backbone atoms excluding any side groups with five clusters and a cutoff of 1.5 Å. Clusters are shown in the VDW (van der Waal) representation using colors as per SNFG were used. [15-17] *ManNAcA residues are shown in green, GlcNAc residues in blue, NAC groups in grey, COOH groups in red, and Fruf in magenta. Notably PmE shows two distinct populations of r which correspond to different alignments of NAc groups (grey) in the molecule. The primary occupancy at 54 % resembles a minor cluster in the 6 RU molecule and is highly prevalent largely due to the short nature of the chain and the high flexibility of the terminal repeating units.*



Supplementary Figure S3.4: *Heatmap plots for the 3 RU molecules showing phi (ϕ) vs psi (ψ) dihedral angles for ManNAcA→GlcNAc and GlcNAc→ManNAcA glycosidic linkages of:* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmB Δ Fruf Δ Gly, and* ***(C)*** *Hid Δ X as well as for the ManNAcA→GlcNAc and GlcNAc→ManNAcA and Fruf→ManNAcA linkages of* ***(D)*** *PmE, and* ***(E)*** *PmB Δ Gly. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. The linkages were defined as we have done previously for these types of linkages.* [27,28] *Glycosidic linkages were defined as* $\phi = H_1\text{-}C_1\text{-}O_1\text{-}C'_x$ *and* $\psi = C_1\text{-}O_1\text{-}C'_x\text{-}H'_x$.

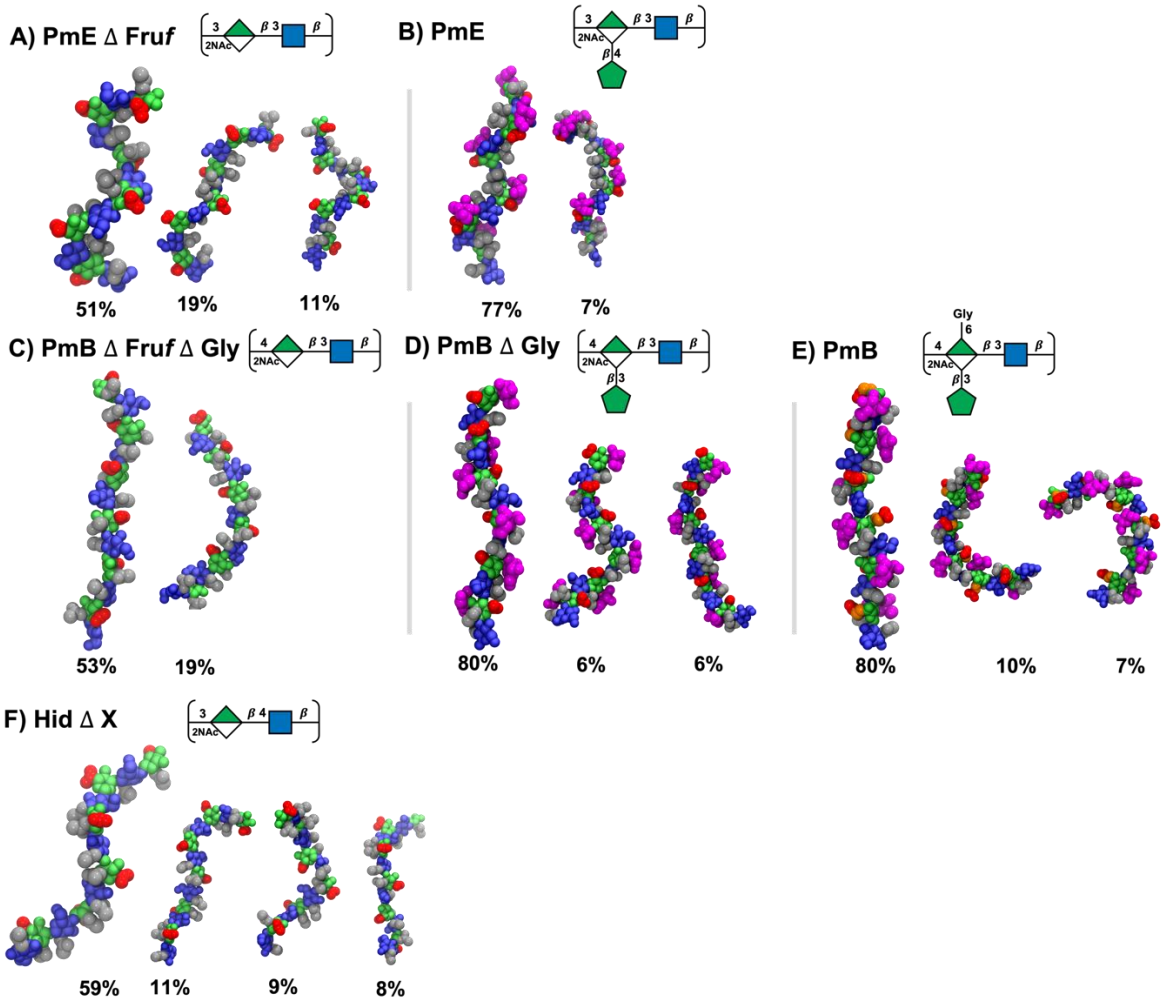


Supplementary Figure S3.5: *Heatmap plots for the modeled 6 RU molecules showing phi (ϕ) vs psi (ψ) dihedral angles for ManNAcA→GlcNAc and GlcNAc→ManNAcA glycosidic linkages of:* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmB Δ Fruf Δ Gly, and* ***(C)*** *Hid Δ X as well as for the ManNAcA→GlcNAc and GlcNAc→ManNAcA and Fruf→ManNAcA linkages of* ***(D)*** *PmE,* ***(E)*** *PmB Δ Gly, and* ***(F)*** *PmB. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. The linkages were defined as we have done previously for these types of linkages.* [27,28] *Glycosidic linkages were defined as* $\phi = H_1\text{-}C_1\text{-}O_1\text{-}C'_x$ *and* $\psi = C_1\text{-}O_1\text{-}C'_x\text{-}H'_x$*. For the*

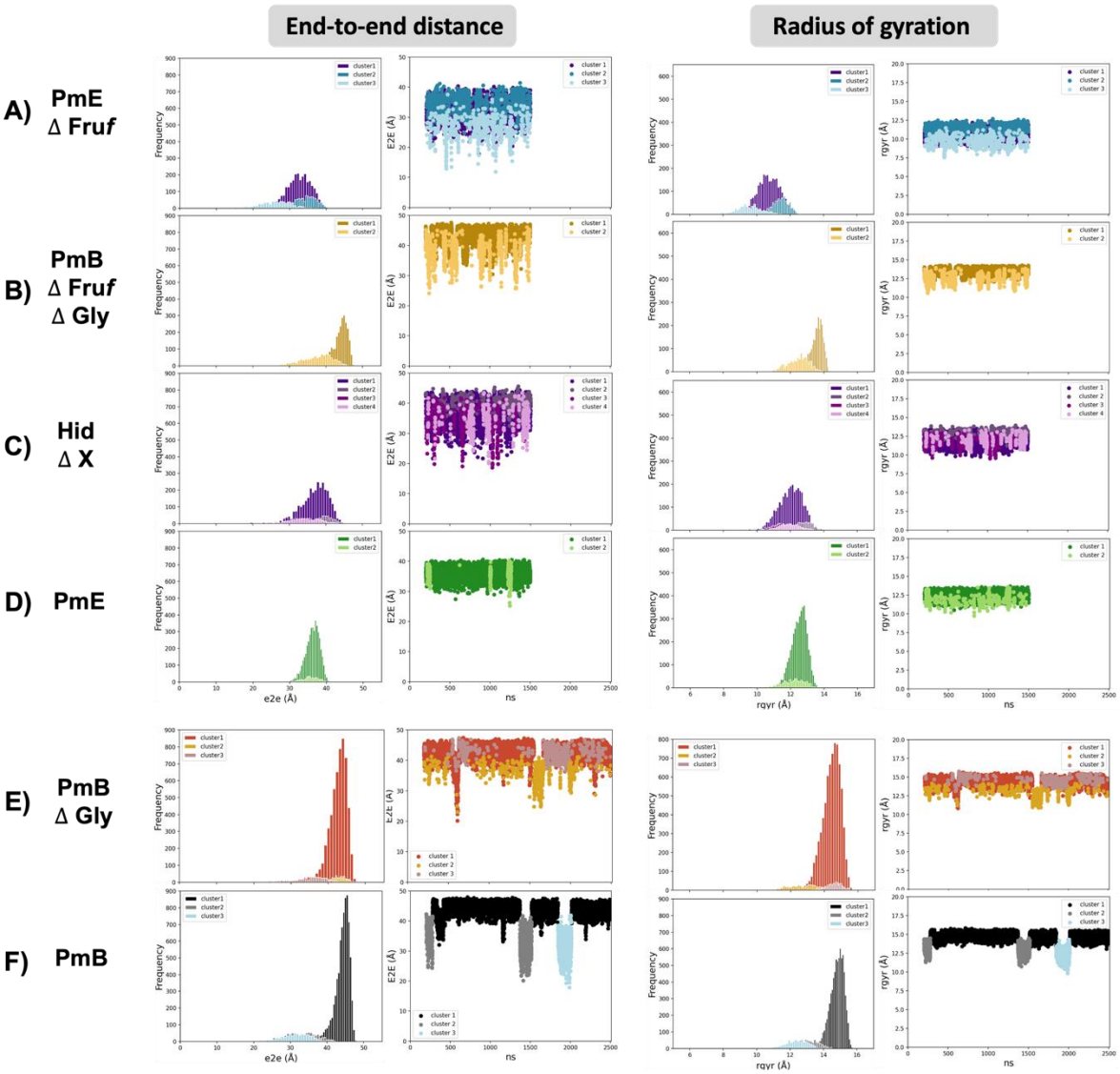
heatmaps, higher occupancy is indicated by redder areas and lower occupancy by bluer areas as shown in the legend on the right of the heatmaps. SNFG line structures are shown in the right column for each molecule. The backbone molecules (A, B, and C) have similar major occupancies (represented by redder areas on the heatmap) with the PmB Δ Fruf Δ Gly (4-linked ManNAcA) more flexible in the GlcNAc→ManNAcA linkage and Hid Δ X (4-linked GlcNAc) in the ManNAcA→GlcNAc linkage than PmE Δ Fruf (fully 3-linked backbone). The substituted molecules (D, E, and F) have similar primary occupancy, but are less disperse and flexible than the backbone molecules (redder and less disperse heatmaps) which is expected from the steric hindrance caused by greater substitution. Further, the Fruf→ManNAcA linkages are very similar in primary occupancy, but for PmE this linkage is more disperse and flexible than for PmB Δ Gly or PmB.

Table S3.1: *Tabulation of the phi, psi (ϕ,ψ) dihedral angles for each glycosidic linkage of the modeled molecules.*

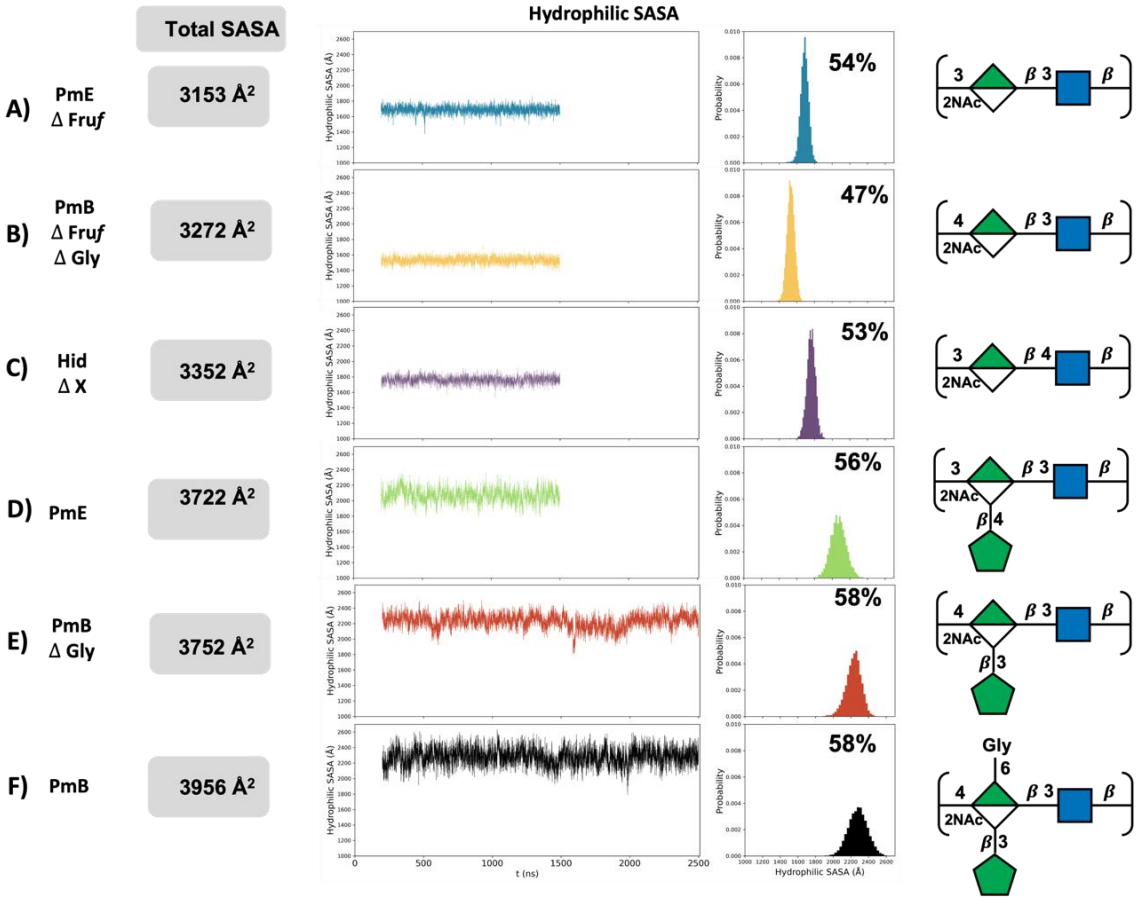
Molecule	**ManNAcA → GlcNAc (*ϕ*,*ψ*)**	**GlcNAc → ManNAcA (*ϕ*,*ψ*)**	**Fru*f* → ManNAcA (*ϕ*,*ψ*)**
PmE Δ Fru*f*	45, -10	45, 40	N/A
PmB Δ Fru*f* Δ Gly	45, -10	60, -10	N/A
Hid Δ X	45, -10	45, 60	N/A
PmE	50, -10	45, 30	150, 10
PmB Δ Gly	50, -5	40, 15	165, 25
PmB	50, -10	45, 0	160, 25



Supplementary Figure S3.6*: Main conformational families and associated percentages identified for the modeled 6 RU P. multocida CPS molecules with the terminal repeat units (RU 1 and RU 6) excluded from analysis.* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmE,* ***(C)*** *PmB Δ Fruf,* ***(D)*** *PmB Δ Fruf Δ Gly,* ***(E)*** *PmB, and* ***(F)*** *Hid Δ X. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. The VDW (van der Waal) representation was used with colors as per SNFG: ManNAcA residues are shown in green, GlcNAc residues in blue, NAc groups in grey, COOH groups in red, Fruf in magenta, and Gly in orange. Line structures are also shown above the clusters using SNFG colors.*



Supplementary Figure S3.7*: Distribution and time series plots of the end-to-end distance and radius of gyration versus time (ns) for the main conformational families of the modeled 6 RU molecules:* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmB Δ Fruf Δ Gly,* ***(C)*** *Hid Δ X,* ***(D)*** *PmE,* ***(E)*** *PmB Δ Gly, and* ***(F)*** *PmB. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine.*



Supplementary Figure S3.8*: Total solvent accessible surface area (SASA) to a 2.5 Å probe (left column) with time series plots of hydrophilic SASA (Å) vs time (ns) (middle column) and histograms showing probability distributions of hydrophilic SASA (Å) (right column) for the following 6 RU molecules:* ***(A)*** *PmE Δ Fruf,* ***(B)*** *PmB Δ Fruf Δ Gly,* ***(C)*** *Hid Δ X,* ***(D)*** *PmE,* ***(E)****PmB Δ Gly, and* ***(F)*** *PmB. X represents a variable amino acid moiety: L-alanine, L-serine, or L-threonine. Average percentage of the surface area that is hydrophilic is indicated on the histograms in the right column. The backbones have similar total SASA increasing in the order: PmE Δ Fruf < PmB Δ Fruf Δ Gly < Hid Δ X. The hydrophilic SASA, however, has a different order: PmB Δ Fruf Δ Gly < Hid Δ X < PmE Δ Fruf. Thus, adding Fruf increases both the total SASA and the hydrophilic SASA for PmE (54 % to 56 %) and PmB Δ Gly (47 % to 58 %) with the greater increase for the PmB molecule potentially due to greater extension in this molecule. The subsequent addition of Gly to PmB again increases the total surface area, but the relative hydrophilic surface stays the same (58 %).*

3.8 References

1. N. I. Richardson, M. M. Kuttel, F. St Michael, C. Cairns, A. D. Cox and N. Ravenscroft, Cross-reactivity of Haemophilus influenzae type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies. *Glycoconj. J.*, 2021, **38**, 735–746.

2. F. St Michael, C. M. Cairns, P. Fleming, E. V. Vinogradov, J. D. Boyce, et al, The capsular polysaccharides of Pasteurella multocida serotypes B and E: Structural, genetic and serological comparisons. *Glycobiology*, 2021, **31**, 307-314.

3. S. F. Capik, Dvm, H. K. Moberly, and R. L. Larson, Systematic review of vaccine efficacy against Mannheimia haemolytica, Pasteurella multocida, and Histophilus somni in North American cattle. *Bov. pract.*, 2021, **55**, 125-133.

4. B. A. Wilson and M. Ho, Pasteurella multocida: from Zoonosis to Cellular Microbiology. *Clin Microbiol Rev*, 2013, **26**, 631-655.

5. J. Lubroth, M. M. Rweyemamu, G. Viljoen, A. Diallo, B. Dungu and W. Amanfu, Veterinary vaccines and their use in developing countries. *Rev. - Off. Int. Epizoot.*, 2007, **26**, 179-201.

6. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of the capsular polysaccharide elaborated by Haemophilus influenzae type d. *Carbohydr. Res.*, 1981, **97**, 285-291.

7. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of two capsular polysaccharides elaborated by different strains of *Haemophilus influenzae* type e. *Carbohydr. Res.*, 1981, **88**, 77-84.

8. I. W. Wilkie, M. Harper, J. D. Boyce, et al, in *Pasteurella multocida*, ed. K. Aktories, J. Orth and B. Adler, Springer, Berlin, Heidelberg, 2012, pp. 1-22.

9. M. Harper, J. D. Boyce and B. Adler, Pasteurella multocida pathogenesis:125 years after Pasteur. *FEMS Microbiol. Lett.*, 2006, **265**, 1-10.

10. A. Benikrane and M. C. L. De Alwis, Haemorrhagic septicaemia, its significance, prevention and control in Asia. *Vet. Med. Czech*, 2002, **47**, 234–240.

11. G. R. Carter, in *Methods in microbiology*, ed. T. Bergan, Academic Press, New York, NY, 1984, pp. 247-258.

12. K. L. Heddleston, J. E. Gallagher and P. A. Rebers, Fowl cholera: gel diffusion precipitin test for serotyping Pasteurella multocida from avian species. *Avian Dis.*, 1972, **16**, 925-936.

13. K. Aktories, J. H. C. Orth and B. Adler, in *Current Topics in Microbiology and Immunology*, ed. K. Aktories, J. Orth and B. Adler, Springer Berlin Heidelberg, Germany, 2012, pp. 39-51.

14. J. D. Boyce and B. Adler, The capsule is a virulence determinant in the pathogenesis of Pasteurella multocida M1404 (B: 2). *Infect. Immun.*, 2000, **68**, 3463-3468.

15. R. S. Haltiwanger, Symbol nomenclature for glycans (SNFG). *Glycobiology*, 2016, **26**, 217.

16. A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, et al, Symbol nomenclature for graphical representations of glycans. *Glycobiology*, 2015, **25**, 1323-1324.

17. S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, et al, Updates to the symbol nomenclature for glycans guidelines. *Glycobiology*, 2019, **29**, 620-624.

18. M. P. E. Slack, Long Term Impact of Conjugate Vaccines on Haemophilus influenzae Meningitis: Narrative Review. *Microorganisms*, 2021, **9**, 886.

19. F. Tsui, R. Schneerson and W. Egan, Structural studies of the Haemophilus influenzae type e capsular polysaccharide. *Carbohydr. Res.*, 1981, **88**, 85-92.

20. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of two capsular polysaccharides elaborated by different strains of haemophilus influenzae type e. *Carbohydr. Res.*, 1981, **88**, 77-84.

21. F. Tsui, R. Schneerson, R. A. Boykins, A. B. Karpas and W. Egan, Structural and immunological studies of the Haemophilus influenzae type d capsular polysaccharide. *Carbohydr. Res.*, 1981, **97**, 293-306.

22. P. Branefors-Helander, L. Kenne, B. Lindberg, K. Petersson and P. Unger, Structural studies of the capsular polysaccharide elaborated by Haemophilus influenzae type d. *Carbohydr. Res.*, 1981, **97**, 285-291.

23. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

24. C. W. Penn and L. K. Nagy, Capsular and somatic antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.*, 1974, **16**, 251-259.

25. C. W. Penn and L. K. Nagy, Isolation of a protective, non-toxic capsular antigen from Pasteurella multocida, types B and E. *Res. Vet. Sci.*, 1976, **20**, 90-96.

26. L. K. Nagy and C. W. Penn, Protection of cattle against experimental haemorrhagic septicaemia by the capsular antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.*, 1976, **20**, 249-253.

27. N. I. Richardson, M. M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity. *Front. Mol. Biosci.*, 2022, **9**, 961532.

28. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines*, 2021, **9**, 432.

29. E. A. Kabat, The nature of an antigenic determinant. *J. Immunol.*, 1966, **97**, 1-11.

30. M. M. Kuttel, J. Ståhle and G. Widmalm, CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.,* 2016, **37**, 2098-2105.

31. M. Kuttel, Y. Mao, G. Widmalm and M. Lundborg, *in Proceedings of 2011 IEEE Seventh International Conference on eScience,* ed. Anonymous, 2011.

32. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, et al, Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.,* 2008, **29**, 2543-2564.

33. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Comput. Chem.,* 2009, **5**, 2353-2370.

34. O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, et al, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *J. Chem. Theory Comput.,* 2011, **7**, 3162-3180.

35. W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.,* 1996, **14**, 33-38.

36. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.,* 1983, **79**, 926-935.

37. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al, Scalable molecular dynamics with NAMD. *J. Comput. Chem.,* 2005, **26**, 1781-1802.

38. J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten, Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.,* 2007, **28**, 2618-2640.

39. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.,* 1993, **98**, 10089-10092.

40. W. F. Van Gunsteren and H. J. Berendsen, A leap-frog algorithm for stochastic dynamics. *Mol. Simul.,* 1988, **1**, 173-185.

41. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.,* 1995, **103**, 4613-4621.

42. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.,* 1985, **31**, 1695.

43. S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems. *Mol. Phys.,* 1983, **50**, 1055-1076.

44. A. Grossfield and D. M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.,* 2009, **5**, 23-48.

45. S. Cross, M. M. Kuttel, J. E. Stone and J. E. Gain, Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.,* 2009, **28**, 131-139.

46. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, 2007, **9**, 90-95.

47. Gracia Luis, *WMC PhysBio Clustering*, Weill Cornell Medical College, 2012.

48. G. Acker and C. Kammerer, Localization of enterobacterial common antigen immunoreactivity in the ribosomal cytoplasm of Escherichia coli cells cryosubstituted and embedded at low temperature. *J. Bacteriol.*, 1990, **172**, 1106-1113.

49. C. Adlam, J. M. Knights, A. Mugridge, J. C. Lindon, P. Baker, et al, Purification, characterization and immunological properties of the serotype-specific capsular polysaccharide of Pasteurella haemolytica (serotype A1) organisms. *Microbiology*, 1984, **130**, 2415-2426.

50. T. Yoneyama, Y. Araki and E. Ito, The primary structure of teichuronic acid in Bacillus subtilis AHU 1031. *Eur. J. Biochem.*, 1984, **141**, 83-89.

51. N. Kawasaki, T. Kawasaki and I. Yamashina, Mannan-Binding Protein and Conglutinin in Bovine Serumz. *J. Biochem.*, 1985, **98**, 1309-1320.

52. U. Holmskov, P. Holt, K. B. Reid, A. C. Willis, B. Teisner and J. C. Jensenius, Purification and characterization of bovine mannan-binding protein. *Glycobiology*, 1993, **3**, 147-153.

53. O. Neth, D. L. Jack, A. W. Dodds, H. Holzel, N. J. Klein and M. W. Turner, Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infect. Immun.*, 2000, **68**, 688-693.

54. F. Carboni, R. Adamo, M. Fabbrini, R. De Ricco, V. Cattaneo, et al, Structure of a protective epitope of group B Streptococcus type III capsular polysaccharide. 2017, **114**, 5017-5022.

55. F. Nato, J. C. Mazie, J. M. Fournier, B. Slizewicz, N. Sagot, et al, Production of polyclonal and monoclonal antibodies against group A, B, and C capsular polysaccharides of Neisseria meningitidis and preparation of latex reagents. *J. Clin. Microbiol.*, 1991, **29**, 1447-1452.

4 Chapter 4

4.1 Preface

The following article, "Conformational and Immunogenicity Studies of the *Shigella flexneri* Serogroup 6 O-Antigen: The Effect of O-Acetylation", describes a computational modeling study of the *Shigella flexneri* serotype 6 (Sf 6) lipopolysaccharide O-antigen (O-Ag). [1] Sf 6 is a leading cause of *S. flexneri* diarrheal disease and mortality making the Sf 6 O-Ag a proposed target for inclusion in a multivalent vaccine to ensure broad protection against *Shigella*. [2]

This stepwise modeling study explored the conformational aspects of the Sf 6 O-Ag which has a different backbone structure and composition from that shared by the other *S. flexneri* serotypes. The Sf 6 O-Ag was compared to the common, Sf Y, O-Ag as well as that of a non-biological, neutral Sf 6 O-Ag and of O-acetylated Sf 6's as partial, non-stoichiometric O-acetylation is present. [3,4]

Our simulations showed flexibility in all the modeled molecules - in common with serogroups Y, 2, 3, and 5 modeled previously. [5,6] Conformational differences between Sf Y and Sf 6 both in helical handedness (as seen in Chapter 3) and in helical extension were observed. The change in helical handedness was as a result of the change of a linkage in the backbone from a 3-linked, flexible equatorial configuration in serotype Y to a more constrained, 4-linked axial configuration in Sf 6 (the same motif was seen in Chapter 3).

Our study also considers the effect of charge on the backbone and molecular surface as well as the impact of O-acetylation (known to affect antigenicity and immunogenicity in other antigens) on the antigenicity of Sf 6. Both have limited effect on conformation but do affect molecular surface hydrophilicity slightly.

Overall, this study indicated significant conformational differences between the Sf Y and Sf 6 suggesting low cross-reactivity between Sf 6 and Sf Y and supporting the need to include Sf 6 in an effective vaccine. These results are further supported by an in vivo study in mice which was performed by our collaborators

and serves as experimental validation for the modeling work.

Conformational and Immunogenicity Studies of the *Shigella flexneri* Serogroup 6 O-Antigen: The Effect of O-Acetylation

Nicole Inge Richardson[a], Neil Ravenscroft[a], Vanessa Arato[b], Davide Oldrini[b], Francesca Micoli[b], and Michelle M. Kuttel*[c].

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] GSK Vaccines Institute for Global Health (GVGH) S.r.l., via Florentina 1, 53100 Siena, Italy

[c] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Correspondence: Michelle Kuttel: mkuttel@cs.uct.ac.za

N.I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M.M. Kuttel, Conformational and Immunogenicity Studies of the *Shigella flexneri* Serogroup 6 O-Antigen: The Effect of O-Acetylation, *Vaccines*, 2021, **9**, 432**.**

4.2 Abstract

The pathogenic bacterium *Shigella* is a leading cause of diarrheal disease and mortality, disproportionately affecting young children in low-income countries. The increasing prevalence of antibiotic resistance in *Shigella* necessitates an effective vaccine, for which the bacterial lipopolysaccharide O-antigen is the primary target. *S. flexneri* serotype 6 has been proposed as a multivalent vaccine component to ensure broad protection against *Shigell*a. We have previously explored the conformations of *S. flexneri* O-antigens from serogroups Y, 2, 3, and 5 that share a common saccharide backbone (serotype Y). Here we consider serotype 6, which is of particular interest because of an altered backbone repeat unit with non-stoichiometric O-acetylation, the antigenic and immunogenic importance of which have yet to be established. Our simulations show significant conformational changes in serotype 6 relative to the serotype Y backbone. We further find that O-acetylation has little effect on conformation and hence may not be essential for the antigenicity of serotype 6. This is corroborated by an in vivo study in mice, using Generalized Modules for Membrane Antigens (GMMA) as O-antigen delivery systems, that shows that O-acetylation does not have an impact on the immune response elicited by the *S. flexneri* serotype 6 O-antigen.

4.3 Introduction

Diarrheal disease is the eighth most common cause of death world-wide, with the highest mortality in infants and geriatrics. [7-9] Diarrheal disease disproportionately affects low-income regions: approximately 90 % occurs in south Asia and sub-Saharan Africa. [8,9] After rotavirus, the pathogenic Gram-negative bacterium *Shigella* is the primary cause of diarrheal mortality, so the increasing prevalence and antibiotic resistance of *Shigella* is a cause for concern. [2,8-13] As a vaccine has been licensed for rotavirus, *Shigella* is a focus of current vaccine development. [2,8]

The O-antigen (O-Ag) on the bacterial cell surface lipopolysaccharide is the primary target of the host immune response against *Shigella* and the focus of current vaccine design. [2,10,14] On the basis of O-Ag composition, the *Shigella* species is divided into the four subgroups *Shigella dysenteriae, Shigella flexneri, Shigella sonnei*, and *Shigella boydii.* [2,3] *S. flexneri* is the most common subgroup to cause disease (66 %), followed by *S. sonnei* (24 %), *S. dysenteriae* (5 %), and S. boydii (5 %).[14] For *S. flexneri*, five serotypes—2a, 6, 3a, 2b, and 1b—account for almost 90 % of disease. [2,14] Based on analysis of structural similarities, it has been proposed that a tetravalent vaccine containing *S. flexneri* serotypes 2a, 3a, 6, and *S. sonnei*, would provide direct protection against 75 % of *S. flexneri* disease, with potential for broader coverage against non-vaccine serotypes (from shared group antigens) as high as 90 %. [2,10,14]

S. flexneri is differentiated into serogroups based on type O-factor groups (I; II; III; IV; V; VI; VII) and further into serotypes based on group O-factors (3,4; 6; 7, 8; 9; 10; IV-1) that are determined by phage-mediated glucosylation, phosphorylation, and O-acetylation of the O-Ag. [3,15] Apart from serogroup 6, all of the *S. flexneri* serotypes share a common backbone repeating unit (serotype Y (Sf Y), **Figure 4.1A**) [3,15] comprising four residues: three rhamnoses (Rha) and a single N-acetylglucosamine (GlcNAc). The serogroup 6 (Sf 6) backbone repeating unit consists of two Rha residues, a galacturonic acid (GalA), and an N-acetylgalactosamine (GalNAc) residue (**Figure 4.1C**) [3,15] and is acidic due to the presence of the charged GalA residue.

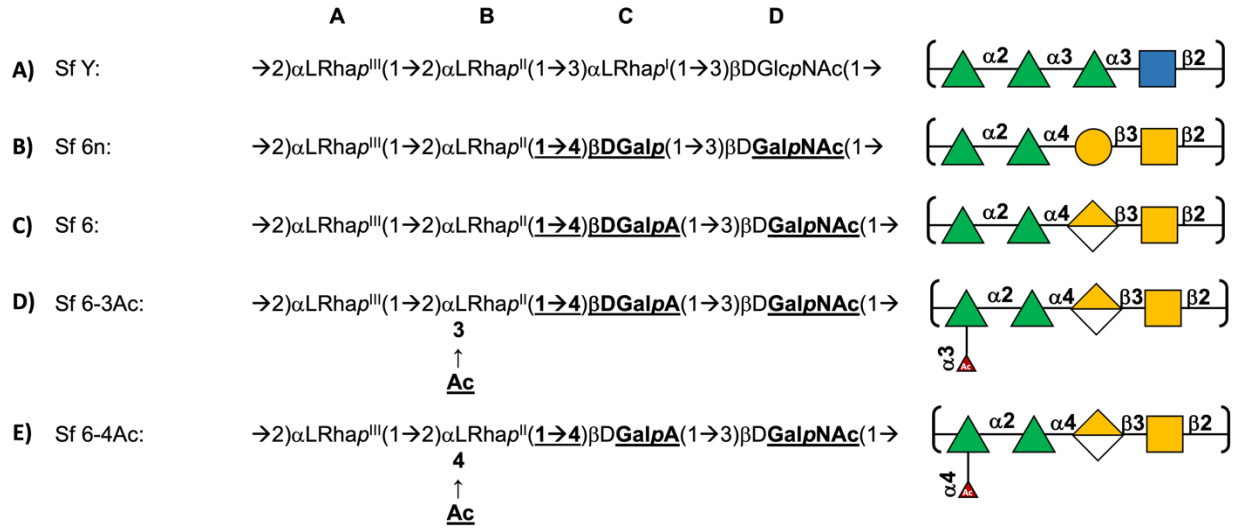


Figure 4.1: *Primary line structures (differences between serotypes shown in bold) and schematic diagrams of the O-antigen repeating units of S. flexneri modelled in this work:* ***(A)*** *serotype Y (Sf Y),* ***(B)*** *non-biological serotype 6n (Sf 6n) with Gal instead of GalA,* ***(C)*** *serotype 6 (Sf 6),* ***(D)*** *serotype 6 with 100 % O-acetylation at O3 of RhaIII (Sf 6-3Ac), and* ***(E)*** *serotype 6 with 100 % O-acetylation at O4 of RhaIII (Sf 6-4Ac). Diagrams use the Symbol Nomenclature for Glycans (SNFG)* [16,17] *with green triangle for Rha, blue square for GlcNAc, blue circle for Glc, yellow circle for Gal, yellow half diamond for GalA, yellow square for GalNAc, and red triangle for O-acetyl.*

Sf 6 has only one structural modification characterized: a phage-mediated, random and non-stoichiometric O3/O4 acetylation of RhaIII (residue A, **Figure 4.1D-E**). [3,4] Modification on O3 confers group O-factor 9 onto Sf 6 and is also observed in serotypes 1b, 2a, 5a, 7a, Y1, and Y2. [3,5] An important question in the context of the O-Ag as a vaccine target is whether O-acetylation is important for antigenicity and immunogenicity; in the case of *S. flexneri* serotype 2a (Sf 2a), O-factor 9 has been shown to not add to broad antibody recognition. [18]

When it is present, cross-protection between antigens allows for reduction of the number of costly vaccine components. [19] Cross-reactivity between serotypes is expected when a shared epitope is present and, although Sf 6 has a different backbone from the non-serotype 6 backbones, cross reactivity between Sf 2a and Sf 6 is expected, due to the shared O-factor 9 epitope. In clinical trials, vaccination with an O-acetylated Sf 2a conjugate in humans (children and adults) induced protective antibodies against Sf 6. [20] The reverse

was also seen in two cases where children (unvaccinated with Sf 2a) who developed Sf 6 disease subsequently developed protective antibodies against Sf 2a. [20] However, data from mouse models showed no cross-reactivity between Sf 6 and Sf 2a. [20]

Molecular modeling of antigen conformations can provide insight into the likelihood of cross-protection between *S. flexneri* O-Ags. [5,6] We have previously modelled the conformations of *S. flexneri* serogroup Y, 2, 3, and 5 O-antigens. [5,6] The molecular dynamics simulations revealed that *S. flexneri* serotypes are highly flexible, with a wide distribution of conformations. Substitutions of the backbone residues (glucosylation and/or O-acetylation) were found to limit the flexibility and distribution of conformations to varying degrees and we proposed three guiding heuristics to describe and predict the effect of substitutions. [6] In particular, of relevance to Sf 6 is that we found that substitution with O-factor 9 restricts the O-Ags to predominantly helical conformations. Conformational differences highlighted by this work support the inclusion of both serotypes 2a and 3a in a potential vaccine. We now focus on Sf 6 (type O-factor VI; group O-factor 9), the second most prevalent cause of *S. flexneri* disease in low- and middle-income countries. [14]

In this work we performed molecular dynamics (MD) simulations of the five structures listed in **Figure 4.1**: the common Sf Y backbone (**Figure 4.1A**); a neutral (non-biological) serotype 6n in which GalA is replaced with Gal (Sf 6n; **Figure 4.1B**); Sf 6 (**Figure 4.1C**); as well as Sf 6 with 100 % O-acetylation at either O3 of RhaIII (Sf 6-3Ac, **Figure 4.1D**) or O4 of RhaIII (Sf 6-4Ac, **Figure 4.1E**). The specific aims of this study are to compare: (a) the conformation and dynamics of the unique Sf 6 backbone with the common Sf Y backbone; (b) the conformation of Sf 6 with the neutral Sf 6n, to establish the effect of charge on conformation; and (c) the effect of O-acetylation at O3 or O4 of RhaIII (residue A) on Sf 6 conformation.

To corroborate our modeling results on O-acetylation, we performed immunological studies in mice with the aim of understanding whether the presence of the O-acetyl group has an impact on *S. flexneri* 6 O-Ag immunogenicity. We genetically manipulated wild type Sf 6 by deleting the *oacC* gene which codes for the O-acetyltransferase C enzyme that is responsible for the O-acetylation of the O-Ag on RhaIII (residue A).

Mice were then immunized using Sf 6 Generalized Modules for Membrane Antigens (GMMA) [21] (a proposed alternative delivery system for *Shigella* O-Ags) [22,23] obtained from either the wild-type or the *oacC* knock-out *S. flexneri* 6 strains [24] and the immune responses compared.

4.4 Materials and Methods

4.4.1 Molecular simulations

We used our established systematic approach to the modeling of polysaccharides, as follows. [6,25] To identify the preferred orientation of each glycosidic linkage, two-dimensional ϕ,ψ potential mean force (PMF) calculations were performed for the glycosidic linkages of all the disaccharide fragments of the O-Ag repeating units. The glycosidic linkage orientations are generally defined by two dihedral angles, ϕ= H_1-C_1-O_1-C_x' and ψ = C_1-O_1-C_x'-H_x', which are equivalent ϕ and ψ in IUPAC nomenclature. [5]

The preferred glycosidic linkage orientations identified by the PMFs were then used to build three repeating unit (3 RU) O-Ag chains for initial Molecular Dynamics (MD) simulations in aqueous solution (data not shown), which were then extended to six repeating unit (6 RU) chains. Chain length is an important consideration when modeling O-Ags, as a short chain may have insufficient molecular flexibility while long chains become extremely time consuming and too computationally expensive to model. *S. flexneri* exhibits a range of O-Ag chain lengths, however, 3 RU is considered sufficient to represent O-Ag conformation (based on our previous work with *S. flexneri* and studies in mice where as few as 1.5 RU were immunogenic). [5,6,21,26,27] Furthermore, antibodies have been shown to only bind small fragments of the O-Ag between one and seven residues in length corresponding to 1-2 RU in the case of *S. flexneri*. [28]

Following initial system equilibration, molecular dynamics runs of 1-2 μs were performed and data analyses were performed on these production runs.

4.4.1.1 PMF calculations

Disaccharide structures were built using CarbBuilder and visualized with Visual Molecular Dynamics (VMD) software. [29-31] PMF calculations were performed using the Metadynamics package incorporated into the Nanoscale Molecular Dynamics (NAMD) software. [32] The ϕ and ψ dihedrals used as collective variables to establish preferred conformations for rotation about the ϕ and ψ dihedrals of the glycosidic

linkages in the disaccharide units.

For all neutral disaccharides, calculations were performed in vacuum (gas phase). For all charged disaccharides, calculations were performed in solution, using the TIP3P [33] model of water with a sodium counter-ion to ensure system neutrality (a requirement for the calculations of electrostatic interactions to converge). Systems were solvated and ionized using VMD's built in solvation and ionization tools.

Each disaccharide system was run at 300 K for 1 000 ns (vacuum simulations) or until the biasing energy reached at least 10 kCal.mol-1 (solution simulations). Once the runs were complete, internal scripts were used to extract global and local minimum energies and to generate contour plots of φ versus ψ.

4.4.1.2 Molecular dynamics

Simulations were run using NAMD (version 2.13) with CUDA extensions for utilizing graphics processors for parallel computing. [34] The CHARMM36 additive force field was chosen for the simulations. [35,36]

The data for Sf Y from our previous work [5] was extended by 1 000 ns and included in this study. Starting structures of the Sf 6 O-Ags (**Figure 4.1**) were built with CarbBuilder using the global minima for the glycosidic linkages determined from the PMF calculations. The starting structures were subsequently minimized using NAMD for 10 000 steps at 310 K.

Minimized structures were solvated in cubic TIP3P [33] water boxes of 90 Å per side. The charged Sf 6, Sf 6-3Ac and Sf 6-4Ac systems underwent an ionization step to add six sodium (Na+) counter ions to neutralize the system. Initial minimization and heating protocols comprised 5 K incremental temperature reassignments beginning at 10 K up to 310 K with 5 000 steps of NAMD minimization and 2 000 steps of MD at each temperature reassignment.

Periodic boundary conditions equivalent to the cubic box size were employed for the solvated simulation with wrapping on. Long range electrostatics were implemented with Particle Mesh Ewald summation (PME) on a 90 Å grid size. [37] 1-3 pairs were excluded from non-bonded interactions, 1-4 interactions were

scaled by a factor of 1, and a dielectric constant of 1 was used for the system. Smoothing functions were applied to both the electrostatics and van der Waals forces with switching and cut-off distances of 15 Å and 12 Å respectively.

A Leap-Frog Verlet integrator was used to integrate the equations of motion over a step size of 1 fs. A distance of 18 Å was used as the cut-off for inclusion in the pair list for calculation of non-bonded forces. The short-range non-bonded interactions were calculated every 1 fs, full electrostatics calculations were performed every 2 fs, and atoms were reassigned every 10 fs. [38]

Simulations were sampled under isothermal-isobaric (nPT) ensemble. Langevin dynamics were used to control the temperature with a damping coefficient of 5/ps. Nosé-Hoover Langevin piston dynamics were used as a barostat to maintain a target pressure of 1 atm. Variable system volume was used with a piston period of 100 fs and decay of 50 fs. Post equilibration (200 ns), simulations underwent production runs of 1 800 ns for Sf Y and Sf 6, 800 ns for Sf 6n, and 900 ns Sf 6-3Ac and Sf 6-4Ac as different simulation lengths were required for the different models to appear converged.

4.4.1.3 Simulation convergence

We addressed the convergence using the method of block standard averaging, applied to two measurables: molecular end-to-end distance and radius of gyration (see **Supplementary Figure S4.1**), as previously described and implemented the method using in-house Python scripts. [6,39]

For all serotypes modeled, the blocked standard error (BSE) can be seen to have reached plateaus. The simulation lengths were large multiples of the correlation times (Sf Y, 50 ns; Sf 6n, 24 ns; Sf 6, 6 ns; Sf 6-3Ac, 11 ns; Sf 6-4Ac,18 ns) and numbers of independent samples (Sf Y, 40; Sf 6n, 83; Sf 6, 180; Sf 6-3Ac, 102; Sf 6-4Ac, 60) were >>1. Furthermore, the equilibration time of 200 ns was >> the correlation time indicating that the properties of the system are not correlated or related on this timescale and that the simulation length should be sufficient. Thus, the simulations appear converged and 200 ns equilibration is sufficient. For Sf Y and Sf 6, longer simulation times were required for convergence (both 2 000 ns) than

for Sf 6n (1 000 ns), Sf 6-3Ac (1 100 ns), and Sf 6-4Ac (1 100 ns).

4.4.1.4 Data analysis

Output trajectories were extracted every 25 ps and analyzed at 250 ps intervals. Inter-atomic distances and dihedral angles were measured using VMD's TK console and graphical user interface (GUI) with data analyses performed using in-house Python scripts and plots generated using Matplotlib. [40]

The end-to-end distance, r, was measured from C2 of RhaII (residue B) at the non-reducing end, to C1 of RhaI/GalA (residue C) at the reducing end, thus excluding the highly flexible terminal residues.

The dihedral angles for each glycosidic linkage were measured as a combination of the dihedrals from the central repeating units, 3 and 4 thus providing a sample of the most central angles.

Molecular conformations were visualized using VMD and clustering of production run trajectories were performed using the WMC PhysBio GUI for VMD's built-in cluster command. [41] Prior to clustering, the molecules were aligned on the ring and linkage atoms of the least flexible central repeating unit, RU 3. Clustering was performed on the ring and linkage atoms RUs 2, 3, 4, and 5, avoiding the highly flexible terminal repeating units 1 and 6. A cut-off of 5.5 Å was set and clusters < 6 % were discarded.

Hydrophilic/hydrophobic regions of the molecular surface were analyzed using VMD's built in "measure sasa" command. The solvent accessible surface area (SASA) analysis was performed by probing first hydrophilic regions (comprising hydroxyl groups, carbonyl groups, amine groups, ring oxygens, and linkage oxygens) and then hydrophobic/neutral regions (comprising methyl groups, CH_2 groups, ring carbons, and ring protons) of the molecule using a van der Waal's radius of 1.4 Å - analogous to that of water. The ratio of hydrophilic to hydrophobic/neutral regions was then calculated to determine the percentage hydrophilic surface area available for potential antibody binding.

When necessary, carbohydrate rings were visualized using the PaperChain visualization algorithm and the

hydrophilic and hydrophobic surfaces were visualized using the Quicksurf visualization algorithm. [42,43]

4.4.2 Immunological studies

We compared the immune responses in vaccinated mice and their functional activity against both Sf 6 and de-O-acetylated Sf 6 strains. We genetically manipulated wild type Sf 6 by deleting the *oacC* gene which codes for the O-acetyltransferase C enzyme that is responsible for the O-acetylation of the O-Ag on RhaIII (residue A). Mice were then immunized using Sf 6 GMMA [21] obtained from either the wild-type or the *oacC* knock-out *S. flexneri* 6 strains. [24] GMMA are Outer Membrane Vesicles (OMV) naturally released from Gram-negative bacteria mutated to increase OMV yield, proposed as an alternative delivery system for *Shigella* O-Ags. [22,23]

4.4.2.1 Bacterial strains, mutant generation and growth conditions

A *S. flexneri* 6 wild type strain was obtained from the Wellcome Trust Sanger Institute and Public Health England. [44] Strain Sf6_Sh10.8537 was selected for the generation of deletion mutants. To generate the mutants, the kanamycin resistance gene *aph* was used to replace the *tolR* and the *oacC* genes. The resistance cassette replacement constructs were amplified from the pKD4 vector using the following primers: Fw ATGTTTGAAATTGATAGCCTATTATTAATAACATCCGTGATAATCTTGTCGTGTAGGCTGGA-GCTGCTTC and Rv GGTTTGTTTTGTTATATTAATGAAAGGTAGTTCAATTAATTTAAATGTTA-CATATGAATATCCTCCTTAG. PCR products were used to transform recombination-prone *S. flexneri* 6 recipient cells carrying pKD46 as described previously. [45] The *oacC* gene was also deleted in a non-*ΔtolR* background (not overblebbling) for use as a target strain in SBA experiments.

All bacterial strains were grown at 30 °C in liquid Luria–Bertani (LB) medium in a rotary shaker for 16 hours. For GMMA production, overnight cultures were diluted in HTMC medium (15 g/L glycerol, 30 g/L yeast extract, 0.5 g/L $MgSO_4$, 5 g/L KH_2PO_4, and 20 g/L K_2HPO_4) to an optical density at 600 nm (OD600) of 0.3 and grown at 30 °C in a rotary shaker for 8 hours using baffled flasks with a liquid to air

volume ratio of 1:5.

4.4.2.2 GMMA production and characterization

After growth, bacteria were pelleted through centrifugation at 5000× g for 45 minutes at 4 °C. Cell-free supernatants were ultracentrifuged at 175,000× g for 2 hours at 4 °C; the resulting pellet, containing GMMA, was washed with phosphate-buffered saline (PBS), further ultra-centrifuged at 175,000× g for 2 hours at 4 °C and finally resuspended in PBS.

GMMA purity was assessed by HPLC–SEC analysis; [46] total protein content was estimated by microBCA (Thermo Scientific, Waltham, MA, USA); O-Ag sugar content was quantified by determination of methyl pentoses (6-deoxyhexoses) with the Dische colorimetric method. [47]

O-Ag extraction and purification from GMMA was performed as previously described. [46]

Nuclear magnetic resonance (NMR) spectroscopy was used to confirm O-Ag identity and to calculate the degree of O-acetylation. [15] All NMR spectra were acquired at 50 °C with an AEON AVANCE III 600 MHz spectrometer (Bruker, Billerica, MA, USA) equipped with a high-precision temperature controller using a 5 mm QCI CryoProbe as previously described. [21]

4.4.2.3 Immunogenicity studies in mice

Animal studies were performed at the GSK Animal Care Facility under the animal project 526/2020-PR 26/05/2020, approved by the Italian Ministry of Health. Five-week-old female wild-type mice were immunized intraperitoneally with 200 μL of vaccine at days 0 and 28. Sera were collected at day 42. Different O-Ag doses were tested in the range 0.005-0.5 mg.

Individual mouse sera were tested for anti-O-Ag total IgG by enzyme-linked immunosorbent assay (ELISA), as previously described. [48] *S. flexneri* 6 Group 4 Capsule (G4C), [21] sharing the same O-Ag RU, at a concentration of 5 μg/mL in carbonate buffer pH 9.6, was used as a coating antigen. Single sera were

tested against both a wild-type and an *oacC* knock-out *S. flexneri* 6 strain in a serum bactericidal assay (SBA) based on luminescent readout as described previously. [21,49]

Results of the assay were expressed as the IC50: the reciprocal serum dilution that produced a 50 % reduction of luminescence, which corresponds to 50 % growth inhibition of the bacteria present in the assay. [50,51] GraphPad Prism 7 software was used for curve fitting and IC50 determination. Titers below the minimum measurable signal were assigned a titer of 50, corresponding to half of the first dilution of sera tested.

Statistical analysis was performed using GraphPad Prism 7. Dose-response relationships were evaluated through Spearman's rank correlation. The parallelism of dose-response curves was assessed by the parallel line method: when the slopes of the curves for O-Ac and non-OAc O-Ags obtained by log-transforming ELISA or SBA results vs log transformed antigen doses were not significantly different from each other, comparison of the Y-intercepts was performed.

4.5 Results

Using the simulation data for Sf Y, Sf 6n, Sf 6 and the O-acetylated Sf 6-3Ac and Sf 6-4Ac, we first compare the chain flexibilities of the O-Ags, then consider the molecular conformations, the characteristics of the molecular surface and finally contrast the minimal binding epitopes. In the light of these results, we then consider the effect of O-acetylation on Sf 6 immunogenicity in mice.

4.5.1 Chain extension and flexibility

Time series plots of the molecular end-to-end distance, r, over the course of the MD simulations provide a simple comparison of the molecular flexibility and chain extension of the O-Ags. We defined *r* (illustrated in **Figure 4.2A**) as the distance from C2 of RhaII (residue B) in RU 1 to C1 of RhaI (residue C) in RU 6 (for Sf Y), or the equivalent C1 of GalA in RU 6 (for Sf 6's). Comparison of the *r* time series plots (**Figure 4.2** left column) and corresponding histograms (**Figure 4.2** right column) reveals significant differences between the Sf Y and Sf 6 O-Ags. Sf Y (**Figure 4.2B**) is the most flexible, with the greatest variance (σ = 15) and range (10 Å to 70 Å) of r. Further, *r* has a bimodal distribution for Sf Y (with peaks at 25 Å and 52 Å) that is markedly different from the unimodal, right skewed distributions of the Sf 6's: the Sf Y backbone has a significant population of compact conformations that is absent in the Sf 6's. In contrast, the distribution of *r* across the four variations of Sf 6 is remarkably similar: the graphs show that the chains are predominantly extended and that the addition of charge or O-acetylation to the Sf 6 backbone has only a minor effect on the O-Ag chain extension and flexibility. The uncharged Sf 6n (**Figure 4.2C**) is the most flexible Sf 6 O-Ag with the largest range of *r* (15 Å to 70 Å, σ = 12). Sf 6 (**Figure 4.2D**) has a similar *r* distribution (10 Å to 70 Å, σ = 11), while the O-acetylated Sf 6-3Ac (**Figure 4.2E**) and Sf 6-4Ac (**Figure 4.2F**) are the most extended chains and show some reduction in flexibility (σ = 10). Overall, flexibility of the O-Ags decreases in the order Sf Y >> Sf 6n > Sf 6 > Sf 6-4Ac > Sf 6-3Ac.

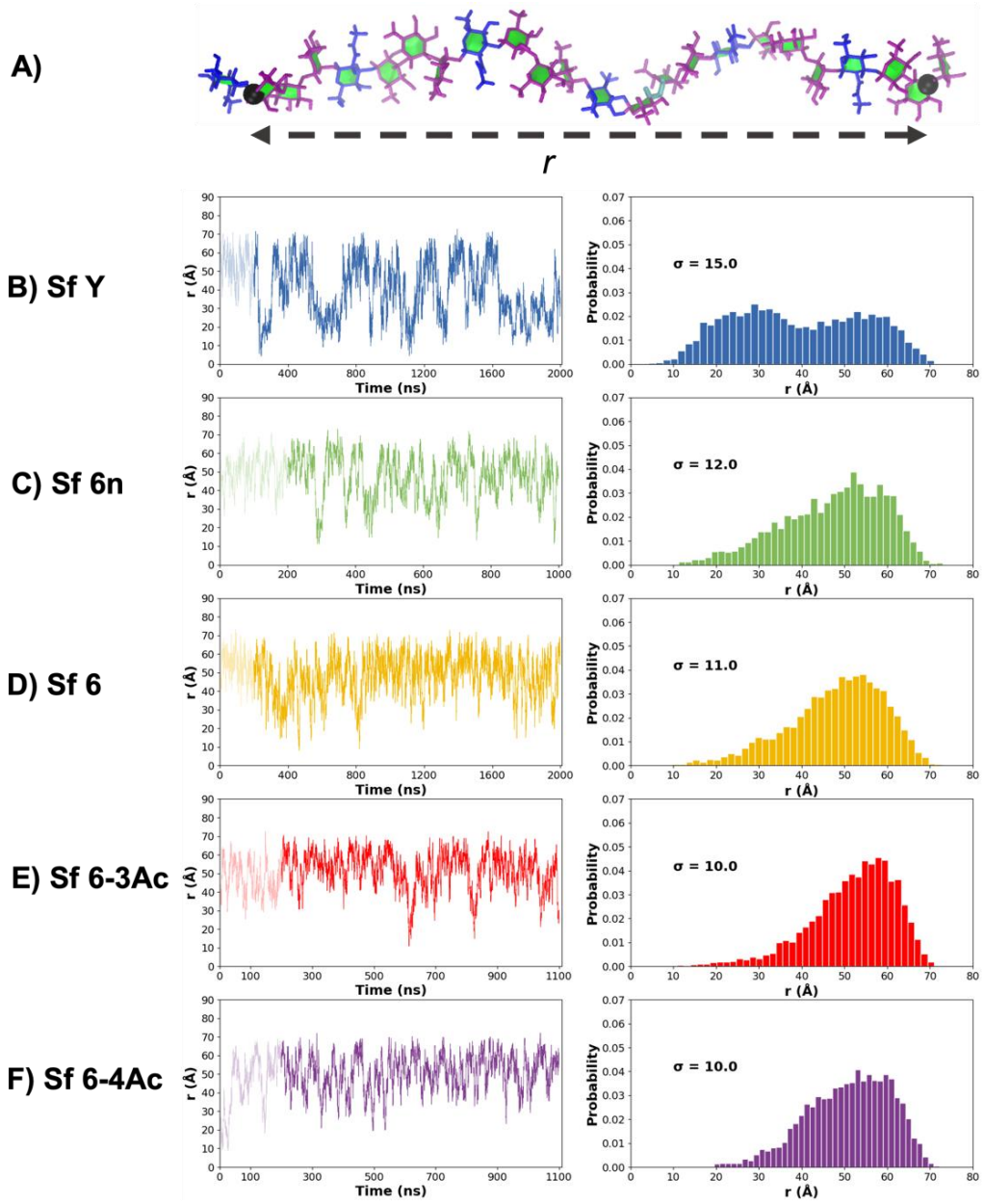


Figure 4.2*: *End-to-end distance, r, is defined as shown in* ***(A) *with end-to-end distance time series analysis as well as histograms for* ***(B)*** *Sf Y,* ***(C)*** *Sf 6n,* ***(D)*** *Sf 6,* ***(E)*** *Sf 6-3Ac, and* ***(F)*** *Sf 6-4Ac. The standard deviation of each histogram, σ, is shown.*

4.5.2 Molecular conformations

4.5.3 Molecular surface

While similar in conformation, the O-Ags differ in charges and substituents. Comparison of the hydrophilic

and hydrophobic regions of the molecular surface in the O-Ags highlights differences between the O-Ags that may affect antibody binding.

Time series and histograms of the hydrophilic surface area of each frame of the trajectory are plotted in **Figure 4.4**, where hydrophilic surface area refers to the percentage of the surface area that is hydrophilic relative to total solvent accessible surface area. Sf Y (**Figure 4.4A**) is markedly more hydrophobic than Sf 6n (**Figure 4.4B**) and Sf 6 (**Figure 4.4C**), and also shows a greater variation in surface hydrophilicity. This is because the extended conformations of the backbone (**Figure 4.4F**) expose more hydrophobic patches (blue) than the collapsed conformations of Sf Y (**Figure 4.4G**). The hydrophobicity of Sf Y relative to the Sf 6's can be rationalized by considering the surfaces of the individual constituent monosaccharides of the backbone (**Supplementary Figure S4.4**) – Gal is markedly more hydrophilic than Glc or Rha. Overall, the hydrophilicity of the exposed O-Ag surface decreases in the order: Sf 6 (55 %) > Sf 6n (52 %) > Sf 6-3Ac (51 %) > Sf 6-4Ac (50 %) > Sf Y (46 %).

Unsurprisingly, the charged Sf 6 is more hydrophilic than the neutral O-Ag (Sf 6n). It is also clear that O-acetylation decreases the hydrophilicity of the chain: with Sf 6-3Ac (**Figure 4.4D**) showing a smaller decrease (4 %) than Sf 6-4Ac (5 %, **Figure 4.4E**) which has the O-methyl more exposed on the chain.

4.5.4 Minimal binding epitope

As antibodies bind to O-Ag regions comprising one to seven residues, [28] it is useful to consider the differences in the 6 RU O-Ag's on this length scale. **Figure 4.5** compares the main conformation and the surface hydrophilicity of a central four residue (BCDA) segment across the Sf Y and Sf 6 O-Ags. It is clear that Sf Y (**Figure 4.5A**) has a very different conformation from the SF 6's. The Sf Y repeating unit is more condensed, residue B is rotated 180 °, and the residue B methyl (Me) group interacts with the N-Acetyl on residue D. The surface representations for Sf Y also show smaller regions of hydrophilicity (blue) compared to the Sf 6's.

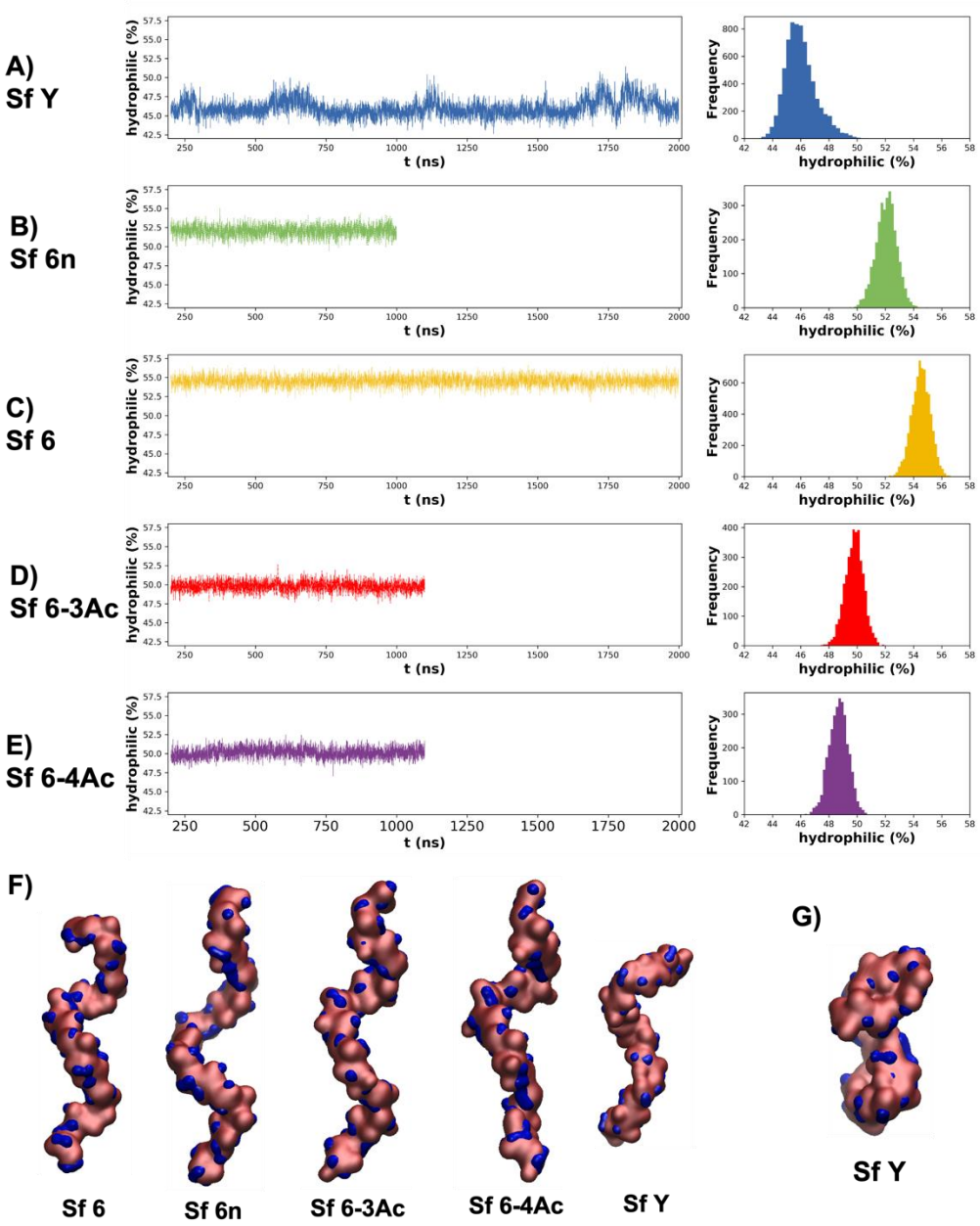


Figure 4.4: *Time series and histogram plots of relative % hydrophilic surface area for:* ***(A)*** *Sf Y,* ***(B)*** *Sf 6n,* ***(C)*** *Sf 6,* ***(D)*** *Sf 6-3Ac,* ***(E)*** *Sf 6-4Ac. VMD Quicksurf representation of the hydrophilic (blue) and combined neutral/hydrophobic (pink) surfaces of the O-antigens are shown in* ***(F)*** *ordered from highest (Sf 6) to lowest (Sf Y) % hydrophilic surface.* ***(G)*** *VMD Quicksurf representation of a collapsed Sf Y conformation from around. Note that for Sf Y and Sf 6, longer simulation times were required to reach convergence (both 2 000 ns) than for Sf 6n (1 000 ns), Sf 6-3Ac (1 100 ns), and Sf 6-4Ac (1 100 ns).*

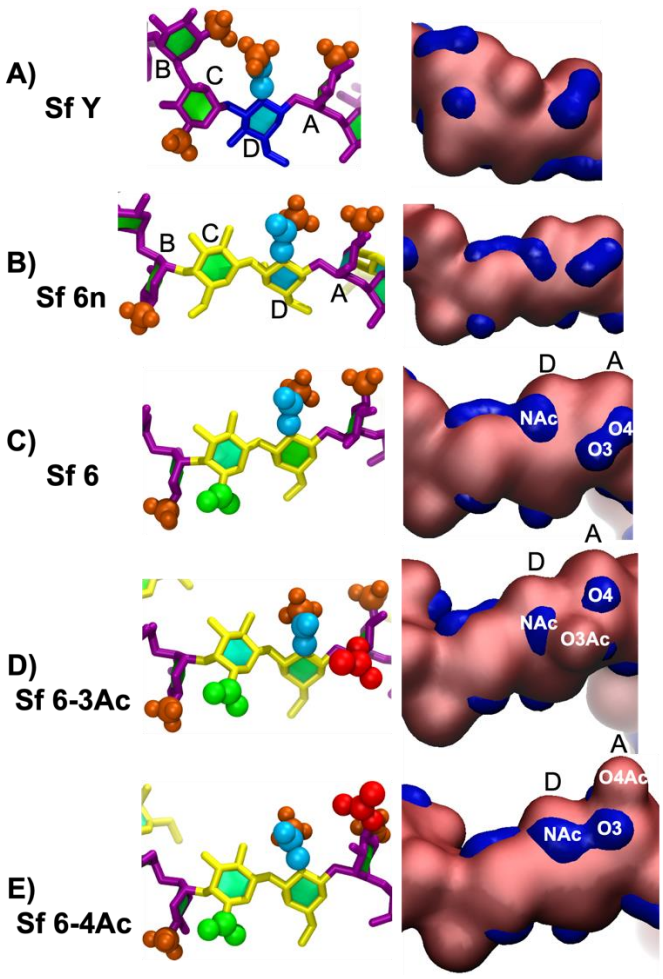


Figure 4.5: *Close up of BCDA residues of RU 3/4 in the licorice/paperchain/VDW and Quicksurf representations of the primary clusters of:* ***(A)*** *Sf Y,* ***(B)*** *Sf 6n,* ***(C)*** *Sf 6,* ***(D)*** *Sf 6-3Ac,* ***(E)*** *Sf 6-4Ac O-antigens. The licorice/paperchain/VDW colors: purple – Rha; dark blue – Glc; yellow – Gal; green – COOH; red – O-acetyl; orange – Me; and cyan-electron withdrawing portion of NAc. The Quicksurf colors: blue – hydrophilic surface; pink – hydrophobic & neutral surface.*

The Sf 6's (**Figure 4.5B-D**) all have similar conformations of the central BCD fragment; the methyl (Me) on residue B interacts with the hydroxyl (OH) or carboxylic acid (COOH) group on C6 of residue C. The substitutions on the Sf 6 backbone do not affect the orientation of this fragment significantly, but they do affect the DA fragment and the binding surface. Relative to Sf 6 (**Figure 4.5C**), O-acetylation at O3 in Sf 6 3Ac (**Figure 4.5D**) or O4 in Sf 6-4Ac (**Figure 4.5E**) results in a disruption of the hydrophilic regions (blue) of residue A (O3 and O4 region) and residue D (NAc region). For the residue A region, the substituted position (O3 or O4) becomes hydrophobic, reducing the size of the local hydrophilic region and, in the case

of Sf 6-4Ac, shifts upwards due to a change in orientation of the O-acetyl to reduce steric hindrance. The residue D region (associated with NAc) shifts towards residue A in Sf 6-3Ac due to NAc – O3Ac interactions and in Sf 6-4Ac, merges with the O3 hydrophilic region.

Overall, the differences between Sf Y and Sf 6 orientations and hydrophilic surfaces indicate the likely antigen binding sites are significantly different. Furthermore, this four-residue region is very similar in the Sf 6's, suggesting that O-acetylation does not have a significant effect on the potential antigen binding site, and hence immunogenicity, of the Sf 6 O-Ag.

4.5.5 The impact of O-acetylation on the immunogenicity of Sf 6 GMMA in mice

To investigate the impact of O-acetylation on the immune response induced by the Sf 6 O-Ags (and verify the predictions made by modeling) the GMMA-producing strain Sf 6 Sh10.8537 *ΔtolR* was further mutated to abolish RhaIII O-acetylation by removing the *oacC* gene - coding for the O-acyltransferase C enzyme - responsible for the backbone modification that confers the presence of O-factor 9. Both GMMA we free of soluble proteins and DNA as detected by HPLC-SEC analysis. The O-Ag to protein weight ratio was 0.4 in the wild type GMMA and 0.23 in *ΔoacC* GMMA. The O-Ag chains on both GMMA had three main populations at average sizes of 174 kDa (G4C), 22 kDa and 1.7 kDa, [21] respectively. 1H NMR analysis of the O-Ag extracted from the *ΔoacC* GMMA confirmed the absence of O-acetylation on RhaIII, while the O-Ag from the wild type GMMA was O-acetylated at positions 3 and 4 of RhaIII (73.3 % and 17.9 %, respectively), as shown in Supplementary **Figure S4.5**.

The characterization of the resulting GMMA by 1H NMR analysis (**Supplementary Figure S4.5**) confirmed the absence of O-acetylation on RhaIII. Moreover, the % O-acetylation at positions 3 and 4 of RhaIII of the native O-Ag were calculated as 73.26 % and 17.95 %, respectively.

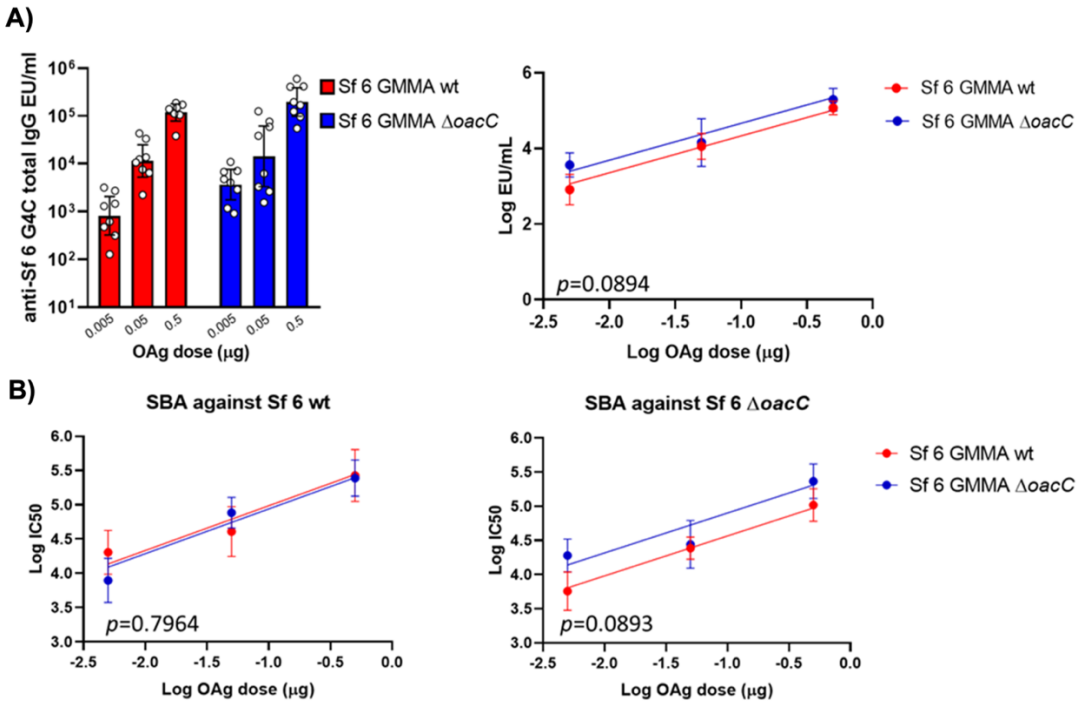


Figure 4.6: *Immunogenicity in mice of S. flexneri 6 GMMA differing in O-Ag O-acetylation on Rha^{III}. Eight CD1 mice per group were immunized intraperitoneally with three doses of O-Ag (0.005, 0.05, 0.5 μg). S. flexneri 6 Group 4 Capsule (G4C), encompassing the same RU of the O-Ag, was used as the ELISA coating antigen.* ***(A)*** *Summary graph of anti-G4C specific IgG geometric mean units (bars) and individual antibody levels (dots) on the left; statistical comparison of the resulting dose response curves on the right;* ***(B)*** *Summary graphs of SBA IC50 titers against S. flexneri 6 wild type (on the left) and ΔoacC (on the right) strains. ELISA and SBA data were analyzed using the parallel line approach. Each curve represents log-transformed doses on the abscissa and the log-transformed ELISA units or SBA titers on the ordinate. The parallelism of the lines was tested by comparison of the slopes, which resulted in no significant differences. Subsequently, the Y-intercept of the curves were compared, and the p-values are reported in the graphs.*

The immunogenicity of GMMA displaying O-Ag with or without O-acetylation on RhaIII was tested in a dose ranging study in mice, comparing GMMA constructs at the same O-Ag dose. Analysis of sera collected two weeks after the second immunization at day 28 (day 42) showed that wild type and *ΔoacC* GMMA did not induce significantly different anti-O-Ag total IgG response in the dose range tested (**Figure 6A**).

SBA analysis against both wild type and *ΔoacC S. flexneri* serotype 6 strains confirmed the results obtained in ELISA, indicating additionally, that the functionality of the antibodies induced was not affected by O-Ag O-acetylation (**Figure 6B**).

4.6 Discussion

Our simulations show that the Sf Y and Sf 6 O-Ags are very flexible, in common with serogroups 2, 3, and 5 modeled previously. However, while Sf Y adopts both elongated and collapsed structures, including right-handed helices, the Sf 6 O-Ags are predominantly in an extended left-handed helical conformation. This conformational change is a result of the change of the BC linkage from a flexible equatorial configuration in Sf Y to a more constrained axial configuration in Sf 6. Due the significant conformational differences between the Sf Y backbone and Sf 6, cross-protection between Sf 6 and Sf Y is not expected to occur.

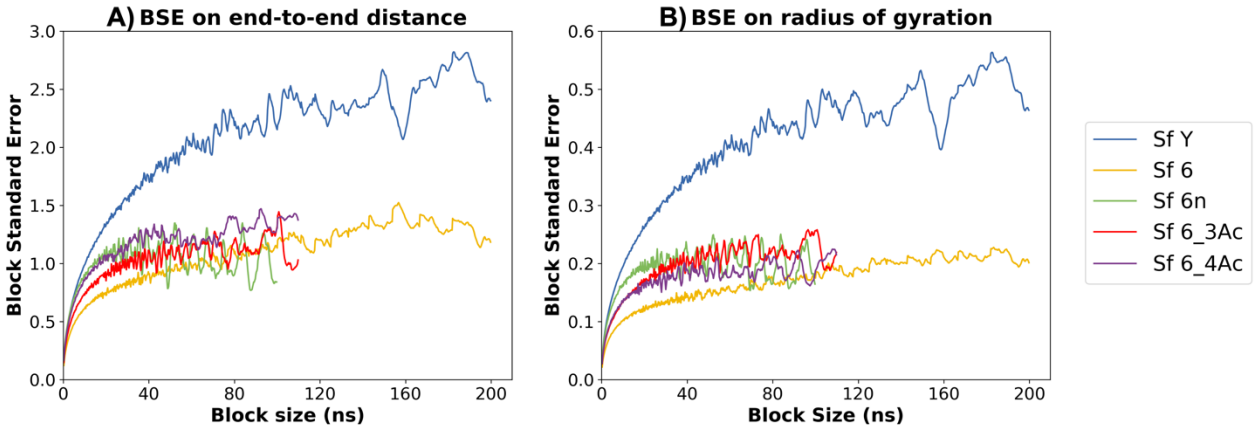
O-acetylation is known to alter carbohydrate chemical and physical properties, such as molecular conformation and hydrophobicity, thereby affecting the antigenicity and immunogenicity of antigens of relevance for vaccine design. [52] Recent reviews on *Shigella* vaccine development [27] and the role of O-acetylation [52] confirm its contribution to the functional immune response for some licensed bacterial polysaccharide-based vaccines (*Salmonella typhi* Vi and *Neisseria meningitidis* serogroup A), but not for others, indicating that the importance of O-acetylation must be established on a case-to-case basis. For *Shigella*, O-acetylation is recognized as a source of additional antigenic diversity and has been characterized by group O-factors. [24] In the case of Sf 6, the presence of RhaIII3/4Ac introduces group O-factor 9 and three subtypes have been recognized on the basis of levels of O-acetylation (I and II) and its absence (III). [52] Further, the presence of an O-acetylated rhamnobiose (αLRhaIII3/4Ac(1→3)αLRhaII) has been suggested as the structural basis for the observed cross-reactivity between Sf 2a and SF6. [20,52]

Our simulations show that O-acetylation of Sf 6 at O3 or O4 of RhaIII (residue A) does not alter the backbone conformation significantly, suggesting that O-acetylation may not be essential for antigenicity. However, O-acetylation results in a significant decrease in the hydrophilicity of the O-Ag in Sf 6 compared to de-O-acetylated Sf 6 based on solvent accessible surface area calculations. This may affect the binding affinity of Sf 6 compared to de-O-acetylated Sf 6, however, the biological significance of this parameter is yet to be established.

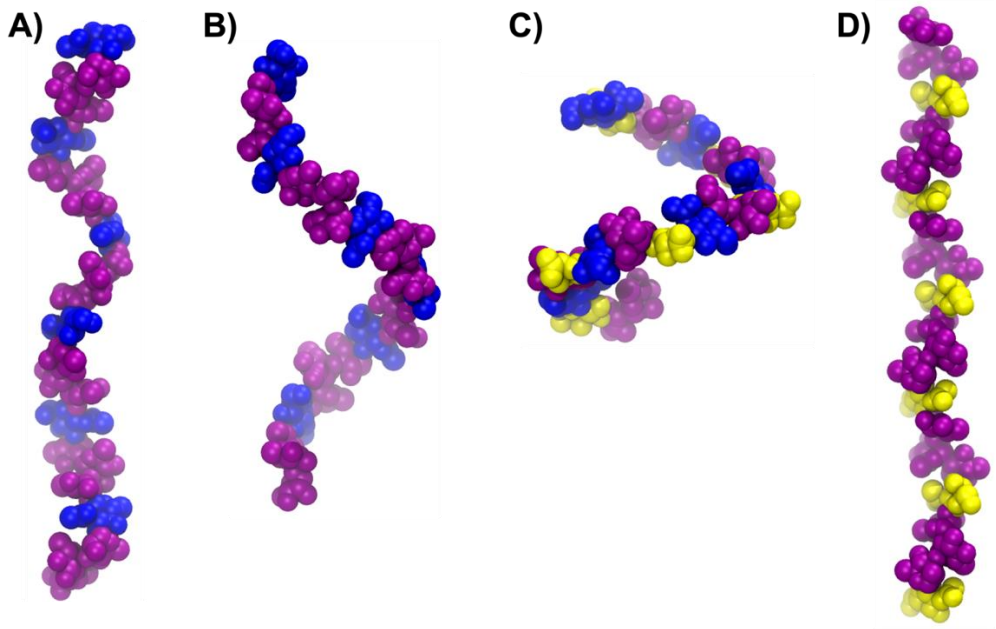
Indeed, our comparison of Sf 6 O-Ags with and without O-acetylation on RhaIII (delivered on GMMA) in mice confirmed that O-acetylation does not play a major role in the ability of the O-Ag to induce antibodies able to recognize the O-acetylated RU and to kill O-acetylated or de-O-acetylated *S. flexneri* 6 bacteria. The results obtained in vivo corroborate the modeling analysis, showing no impact of O-acetylation on immunogenicity. This is in line with the marginal difference observed in the O-Ag conformation. These results are also in agreement with our previous studies on CRM197 glycoconjugates in which an O-acetylated and de-O-acetylated CRM197 glycoconjugate induced similar IgG and SBA titers. [21]

Further validation of the importance of the O-acetylation and cross reactivity of *S. flexneri* serotype 6 in other animal models (and preferably humans) is necessary to definitively confirm the role of O-acetylation in immunogenicity and provide insight into appropriate, representative animal models for *Shigella* sp vaccine development.

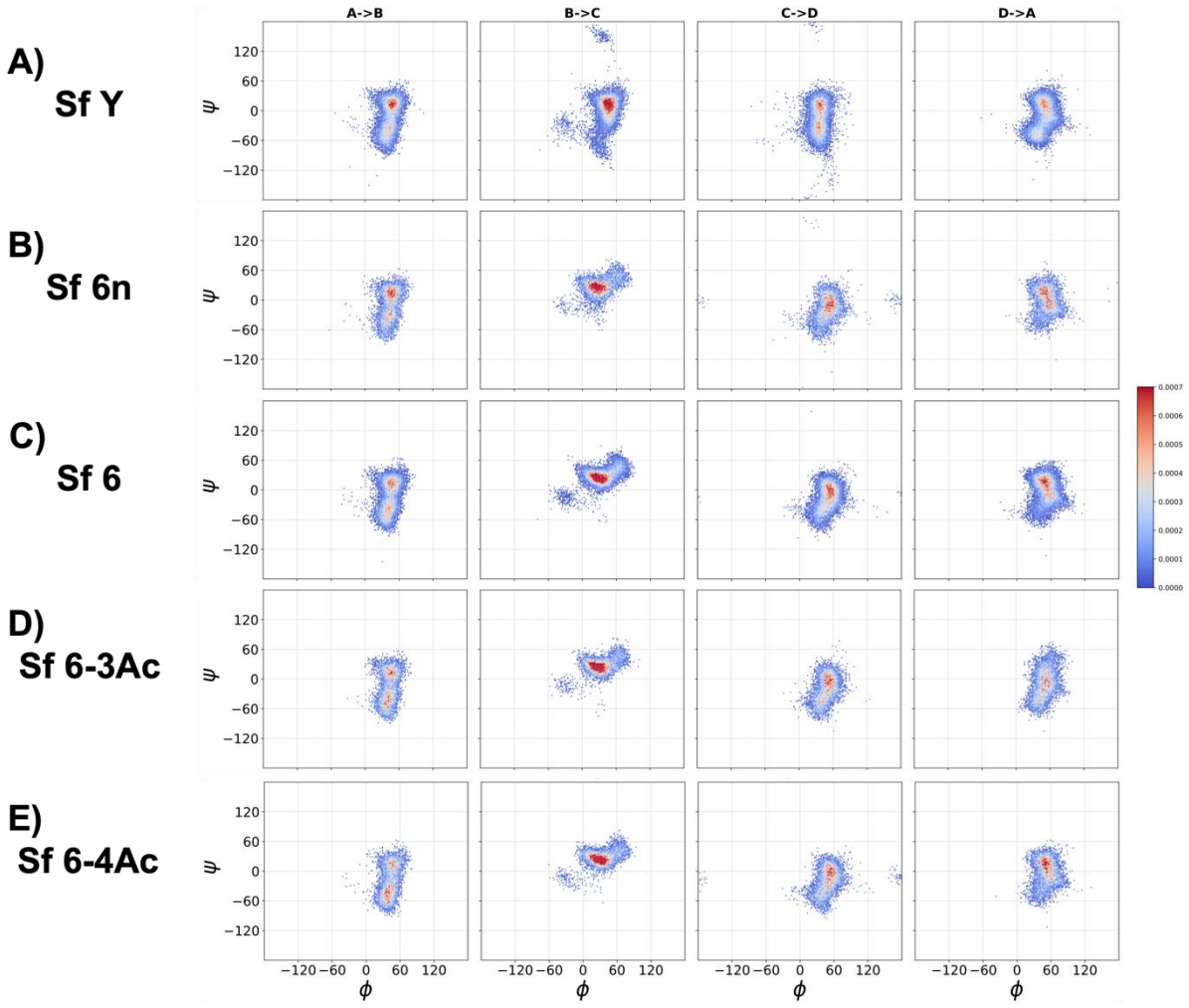
4.7 Supplementary Material



Supplementary Figure S4.1: *Block standard averaging analysis calculated for the time series of* ***(A)*** *end-to-end distance and* ***(B)*** *radius of gyration of the 6 RU O-antigen chains. The blocked standard error (BSE) reaches a plateau with increased block size (ns) for all of the O-antigens indicating convergence. Further analysis using the approximate BSE reveals correlation times < 55 ns (much less than the simulations times of 1 000-2 000 ns) and numbers of independent samples >> 1 which support the convergence of the simulation.*



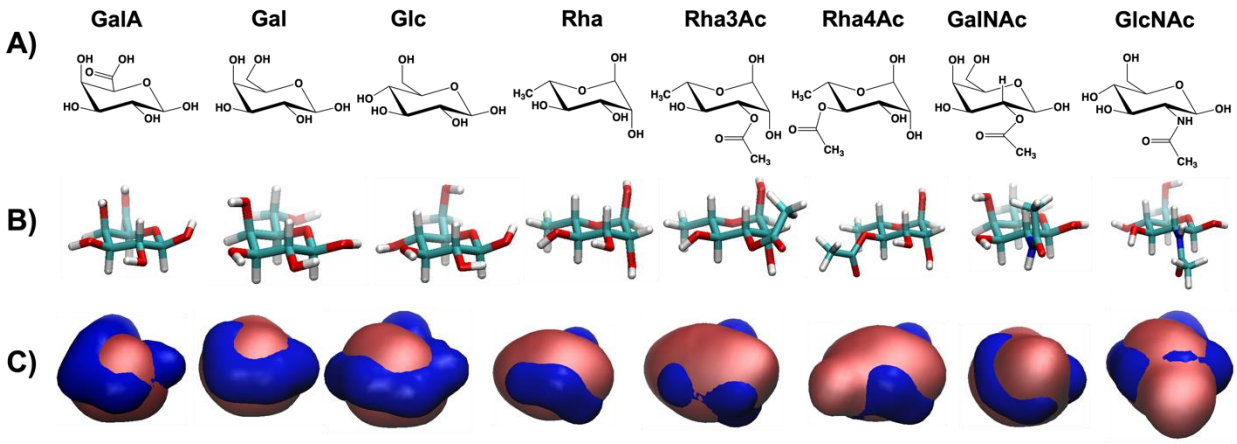
Supplementary Figure S4.2: *6 repeating unit static structures of the Sf Y repeating unit, each with one of the structural modifications seen in serotype 6's backbone:* ***(A)*** *Sf Y backbone with no replacement,* ***(B)*** *replacing the BC (αLRha1→3αLRha) linkage with an (αLRha1→4αLRha) linkage,* ***(C)*** *replacing residue C, rhamnose, with galactose (linkages unchanged),* ***(D)*** *replacing residue D, GlcNAc, with GalNAc (linkages unchanged). Only structure* ***(B)*** *changes from a right-handed to left-handed helix indicating the change in linkage position (which involves changing from an equatorial to axial linkage) causes the change in handedness of the helix.*



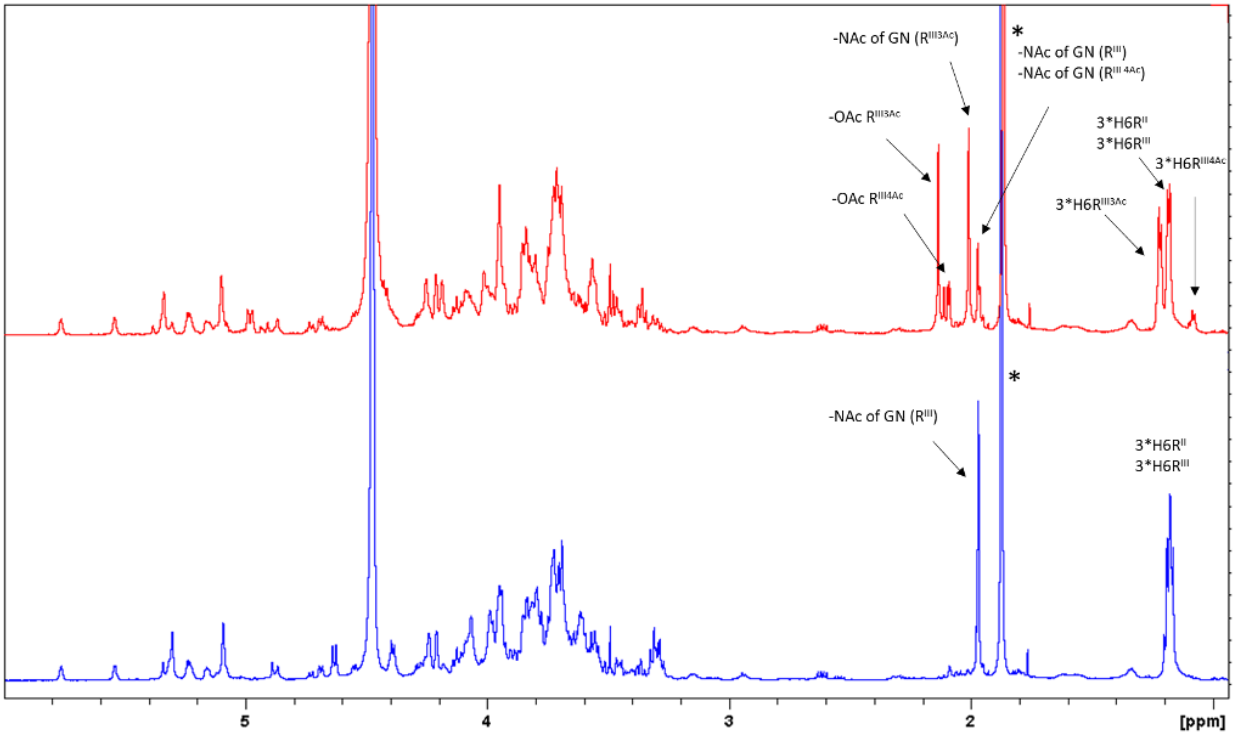
Supplementary Figure S4.3: *Heat map representation of scatter plots for the phi (ϕ), psi (ψ) dihedral angles of each glycosidic linkage for:* ***(A)*** *Sf Y,* ***(B)*** *Sf 6n,* ***(c)*** *Sf 6,* ***(D)*** *Sf 6-3Ac,* ***(E)*** *Sf 6-4Ac. Changes in chain conformation are primarily caused by changes in the orientations of the constituent glycosidic linkages due to the carbohydrate rings being quite constrained. Points from both of the central repeating units (RU 3 and RU 4) were included to broadly sample the backbone behavior. The color scale indicates the relevant occupancy of the dihedral angles during the simulation with red being high occupancy, and blue low occupancy.*

Table S4.1: *Tabulation of primary and secondary occupied dihedral angles of glycosidic linkages for the modeled O-antigens.*

O-antigen	Glycosidic Linkage (of repeating unit ABCD)						
	AB		*BC*	*CD*		*DA*	
	Primary (ϕ, ψ)	*Secondary* (ϕ, ψ)	*Primary* (ϕ, ψ)	*Primary* (ϕ, ψ)	*Secondary* (ϕ, ψ)	*Primary* (ϕ, ψ)	*Secondary* (ϕ, ψ)
Sf Y	50, 15	40, -45	45, 10	40, 10	40, -45	50, 10	40, -45
Sf 6n	45, 15	40, -30	25, 25	55, -10	N/A	50, 15	60, -10
Sf 6	45, 15	40, -45	30, 20	55, 0	N/A	50, 15	N/A
Sf 6-3Ac	45, 15	40, -45	25, 25	55, -5	N/A	55, -5	N/A
Sf 6-4Ac	45, 15	35, -45	25, 25	55, -5	N/A	55, 15	N/A



Supplementary Figure S4.4: (A) *line structures of S. flexneri RU constituent monosaccharides,* ***(B)*** *CPK models of the monosaccharides,* ***(C)*** *Quicksurf representation of monosaccharides with hydrophilic regions in blue and hydrophobic/neutral regions in pink. Monosaccharides are arranged in order of decreasing % hydrophilic surface: GalA, 74 %; Gal, 68 %; Glc, 62 %; Rha, 60.4 %; Rha3Ac, 60.2 %; Rha4Ac, 59.3 %; GalNAc, 59 %; GlcNAc, 58 %.*



Supplementary Figure S4.5: *^{1}H NMR spectra (from 1 to 6 ppm) of S. flexneri 6 O-Ags extracted from GMMA produced by the wild type (red) and ΔoacC mutant (blue) strains, confirming expected O-acetylation patterns. Diagnostic signals are assigned in the spectra. *=acetate buffer. For the annotations, GN = GlcNAc, R = Rha.*

4.8 References

1. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines*, 2021, **9**, 432.

2. K. L. Kotloff, M. S. Riddle, J. A. Platts-Mills, P. Pavlinac and A. K. M. Zaidi, Shigellosis. *Lancet*, 2018, **391**, 801-812.

3. Y. Knirel, Q. Sun, S. Senchenkova, A. Perepelov, A. Shashkov and J. Xu, O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms. *Biochemistry*, 2015, **80**, 901-914.

4. P. Chassagne, C. Fontana, C. Guerreiro, C. Gauthier, A. Phalipon, et al, Structural studies of the O-acetyl containing O-antigen from a Shigella flexneri serotype 6 strain and synthesis of oligosaccharide fragments thereof. *Eur. J. Org. Chem.*, 2013, **2013**, 4085-4106.

5. J. Hlozek, N. Ravenscroft and M. M. Kuttel, Effects of Glucosylation and O-Acetylation on the Conformation of Shigella flexneri Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B*, 2020, **124**, 2806-2814.

6. J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, Molecular Modeling of the Shigella flexneri Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a. *Vaccines*, 2020, **8**, 643.

7. M. Naghavi, A. A. Abajobir, C. Abbafati, K. M. Abbas, F. Abd-Allah, et al, Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, 2017, **390**, 1151-1210.

8. I. A. Khalil, C. Troeger, B. F. Blacker, P. C. Rao, A. Brown, et al, Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.*, 2018, **18**, 1229-1240.

9. C. Troeger, M. Forouzanfar, P. C. Rao, I. Khalil, A. Brown, et al, Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.*, 2017, **17**, 909-948.

10. K. L. Kotloff, J. A. Platts-Mills, D. Nasrin, A. Roose, W. C. Blackwelder and M. M. Levine, Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine*, 2017, **35**, 6783-6789.

11. D. E. Atherly, K. D. C. Lewis, J. Tate, U. D. Parashar and R. D. Rheingans, Projected health and economic impact of rotavirus vaccination in GAVI-eligible countries: 2011–2030. *Vaccine*, 2012, **30**, A7-A14.

12. U.S Department of Health and Human Services, CDC, Antibiotic Resistance Threats in the United States, 2019, https://www.cdc.gov/drugresistance/Biggest-Threats.html, (Accessed:09 March 2021).

13. S. Shrivastava, P. Shrivastava and J. Ramasamy, World health organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *J. Med. Soc.,* 2018, **32**, 76-77.

14. S. Livio, N. A. Strockbine, S. Panchalingam, S. M. Tennant, E. M. Barry, et al, Shigella Isolates From the Global Enteric Multicenter Study Inform Vaccine Development. *Clin. Infect. Dis.,* 2014, **59**, 933-941.

15. A. V. Perepelov, M. E. Shekht, B. Liu, S. D. Shevelev, V. A. Ledov, et al, Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol. Med. Mic.,* 2012, **66**, 201-210.

16. R. S. Haltiwanger, Symbol nomenclature for glycans (SNFG). *Glycobiology,* 2016, **26**, 217.

17. S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, et al, Updates to the symbol nomenclature for glycans guidelines. *Glycobiology,* 2019, **29**, 620-624.

18. C. Gauthier, P. Chassagne, F. Theillet, C. Guerreiro, F. Thouron, et al, Non-stoichiometric O-acetylation of Shigella flexneri 2a O-specific polysaccharide: synthesis and antigenicity. *Org. Biomol. Chem.,* 2014, **12**, 4218-4232.

19. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

20. N. Farzam, R. Ramon-Saraf, Y. Banet-Levi, L. Lerner-Geva, S. Ashkenazi, et al, Vaccination with Shigella flexneri 2a conjugate induces type 2a and cross-reactive type 6 antibodies in humans but not in mice. *Vaccine,* 2017, **35**, 4990-4996.

21. M. M. Raso, G. Gasperini, R. Alfini, F. Schiavo, M. G. Aruta, et al, GMMA and Glycoconjugate Approaches Compared in Mice for the Development of a Vaccine against Shigella flexneri Serotype 6. *Vaccines,* 2020, **8**, 160.

22. C. Gerke, A. M. Colucci, C. Giannelli, S. Sanzone, C. G. Vitali, et al, Production of a Shigella sonnei vaccine based on generalized modules for membrane antigens (GMMA), 1790GAHB.*; PLoS One,* 2015, **10**, e0134478.

23. F. Micoli and C. A. MacLennan, Outer membrane vesicle vaccines. *Semin. Immunol.,* 2020, **50**, 101433.

24. Y. A. Knirel, J. Wang, X. Luo, S. N. Senchenkova, R. Lan, et al, Genetic and structural identification of an O-acyltransferase gene (oacC) responsible for the 3/4-O-acetylation on rhamnose III in Shigella flexneri serotype 6. *BMC Microbiol.,* 2014, **14**, 1-10.

25. J. Hlozek, M. M. Kuttel and N. Ravenscroft, Conformations of Neisseria meningitidis serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.,* 2018, **465**, 44-51.

26. F. Theillet, C. Simenel, C. Guerreiro, A. Phalipon, L. A. Mulard and M. Delepierre, Effects of backbone substitutions on the conformational behavior of *Shigella flexneri* O-antigens: implications for vaccine strategy. *Glycobiology,* 2011, **21**, 109-121.

27. L. Barel and L. A. Mulard, Classical and novel strategies to develop a Shigella glycoconjugate vaccine: from concept to efficacy in human. *Hum. Vaccin. Immunother.,* 2019, **15**, 1338-1356.

28. E. A. Kabat, The nature of an antigenic determinant. *J. Immunol.,* 1966, **97**, 1-11.

29. M. M. Kuttel, J. Ståhle and G. Widmalm, CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.,* 2016, **37**, 2098-2105.

30. M. Kuttel, Y. Mao, G. Widmalm and M. Lundborg, *in Proceedings of 2011 IEEE Seventh International Conference on eScience,* ed. Anonymous, 2011.

31. W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.,* 1996, **14**, 33-38.

32. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al, Scalable molecular dynamics with NAMD. *J. Comput. Chem.,* 2005, **26**, 1781-1802.

33. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.,* 1983, **79**, 926-935.

34. J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten, Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.,* 2007, **28**, 2618-2640.

35. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, et al, Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.,* 2008, **29**, 2543-2564.

36. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Comput. Chem.,* 2009, **5**, 2353-2370.

37. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.,* 1993, **98**, 10089-10092.

38. W. F. Van Gunsteren and H. J. Berendsen, A leap-frog algorithm for stochastic dynamics. *Mol. Simul.,* 1988, **1**, 173-185.

39. A. Grossfield and D. M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.,* 2009, **5**, 23-48.

40. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.,* 2007, **9**, 90-95.

41. Gracia Luis, *WMC PhysBio Clustering*, Weill Cornell Medical College, 2012.

42. S. Cross, M. M. Kuttel, J. E. Stone and J. E. Gain, Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.,* 2009, **28**, 131-139.

43. M. Krone, J. E. Stone, T. Ertl and K. Schulten, *in Proceedings of Eurographics Conference on Visualization (EuroVis)* ed. M. Meyer and T. Weinkauf, Vienna Austria, 2012.

44. O. Rossi, K. S. Baker, A. Phalipon, F. Weill, F. Citiulo, et al, Draft genomes of Shigella strains used by the STOPENTERICS consortium. *Gut Pathog.,* 2015, **7**, 1-6.

45. K. A. Datsenko and B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.,* 2000, **97**, 6640-6645.

46. G. De Benedetto, R. Alfini, P. Cescutti, M. Caboni, L. Lanzilao, et al, Characterization of O-antigen delivered by Generalized Modules for Membrane Antigens (GMMA) vaccine candidates against nontyphoidal Salmonella. *Vaccine,* 2017, **35**, 419-426.

47. Z. Dische and L. B. Shettles, A specific color reaction of methylpentoses and a spectrophotometric micromethod for their determination. *J. Biol. Chem.,* 1948, **175**, 595-603.

48. L. Lanzilao, G. Stefanetti, A. Saul, C. A. MacLennan, F. Micoli and S. Rondini, Strain selection for generation of O-antigen-based glycoconjugate vaccines against invasive nontyphoidal Salmonella disease.*; PLoS One,* 2015, **10**, e0139847.

49. F. Necchi, A. Saul and S. Rondini, Setup of luminescence-based serum bactericidal assay against Salmonella Paratyphi A. *J. Immunol. Methods,* 2018, **461**, 117-121.

50. F. Necchi, A. Saul and S. Rondini, Development of a high-throughput method to evaluate serum bactericidal activity using bacterial ATP measurement as survival readout. *PLoS One,* 2017, **12**, e0172163.

51. O. Rossi, E. Molesti, A. Saul, C. Giannelli, F. Micoli and F. Necchi, Intra-laboratory evaluation of luminescence based high-throughput serum bactericidal assay (L-SBA) to determine bactericidal activity of human sera against Shigella. *High-Throughput;* 2020, **9**, 14.

52. L. A. Mulard, Bacterial polysaccharides as major surface antigens: interest in O-acetyl substitutions. *Carbohydr. Chem.,* 2017, **43**, 71-103.

5 Chapter 5

5.1 Preface

The following article "Modeling of Pneumococcal Serogroup 10 Capsular Polysaccharide Molecular Conformations Provides Insight into Epitopes and Observed Cross-Reactivity" describes a modeling study of the *Streptococcus pneumoniae* serogroup 10 capsular polysaccharides (CPSs) which exhibit complex asymmetry and high levels of mistyping. [1-3]

The *S. pneumoniae* serogroup 10 CPSs contain a ribitol phosphate moiety (a motif seen in *Haemophilus influenzae* - Chapter 2) which contributes significant flexibility to the backbone. [4-6] The serogroup 10 CPSs also have side groups which play an important role in innate immune evasion. As for *H. influenzae* in Chapter 2, on a whole molecule level, the conformational analysis was not very informative.

This work required us to look more closely at potential binding epitopes rather than whole molecule conformations alone to rationalize the complex asymmetrical cross-reactivity observed. We were also able to provide further insights into potential antigen-antibody interactions through the overlaying of our potential binding epitopes with known galactofuranose recognizing intelectins (carbohydrate binding proteins of the immune system).

These additional analyses allowed for the rationalization of the observed epidemiology as well as the complex, asymmetrical cross-reactivity within the serogroup. This is useful to the design of next generation pneumococcal vaccines as the inclusion of a single serotype would be expected to provide sufficient protection against the remaining serotypes.

Modeling of Pneumococcal Serogroup 10 Capsular Polysaccharide Molecular Conformations Provides Insight into Epitopes and Observed Cross-Reactivity

Nicole I. Richardson[a], Michelle M. Kuttel[b], Neil Ravenscroft[a*]

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Correspondence: Neil Ravenscroft: neil.ravenscroft@uct.ac.za

N.I. Richardson, M.M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity, *Front. Mol. Biosci.*, 2022, **9**, 961532.

5.2 Abstract

Streptococcus pneumoniae is an encapsulated gram-negative bacterium and a significant human pathogen. The capsular polysaccharide (CPS) is essential for virulence and a target antigen for vaccines. Although widespread introduction of pneumococcal conjugate vaccines (PCVs) has significantly reduced disease, the prevalence of non-vaccine serotypes has increased. On the basis of the CPS, *S. pneumoniae* serogroup 10 comprises four main serotypes 10A, 10B, 10C and 10F; as well as the recently identified 10D. As it is the most prevalent, serotype 10A CPS has been included as a vaccine antigen in the next generation PCVs. Here we use molecular modeling to provide conformational rationales for the complex cross-reactivity reported between serotypes 10A, 10B, 10C, and 10F anti-sera. Although the highly mobile phosphodiester linkages produce very flexible CPS, shorter segments are conformationally defined, with exposed β-D-galactofuranose (βDGal*f*) side chains that are potential antibody binding sites. We identify four distinct conformational epitopes for the immunodominant βDGal*f* that assist in rationalizing the complex asymmetric cross-reactivity relationships. In particular, we find that strongly cross-reactive serotypes share common epitopes. Further, we show that human intelectin-1 has the potential to bind the exposed exocyclic 1,2-diol of the terminal βDGal*f* in each serotype; the relative accessibility of 3- or 6-linked βDGal*f* may play a role in the strength of the innate immune response and hence serotype disease prevalence. In conclusion, our modeling study and relevant serological studies support the inclusion of serotype 10A in a vaccine to best protect against serogroup 10 disease.

5.3 Introduction

Streptococcus pneumoniae is an encapsulated gram-negative bacterium and an important human pathogen responsible for significant disease and mortality, especially in children under five. [4] The capsular polysaccharide (CPS) is essential for virulence and an important target antigen for vaccines. [4,7] Although 100 different serotypes have been identified, most disease is caused by a subset of pathogenic serotypes. [8] Vaccines against the CPS of *S. pneumoniae* have been developed either as a pneumococcal polysaccharide vaccine (PPV) or a pneumococcal conjugate vaccine (PCV) - polysaccharide conjugated to a carrier protein. [4] Unlike polysaccharide vaccines, the glycoconjugate vaccines induce a T-cell dependent immune response and are effective in young children; the age group highly susceptible to infections [9] This has led to the licensure of PCVs targeting, initially, seven (PCV7, serotypes 4, 6B, 9 V, 14, 18C, 19F and 23F), then ten (PCV10, PCV7 plus serotypes 1, 5 and 7F), and 13 (PCV13, PCV10 plus serotypes 3, 6A and 19A) serotypes.

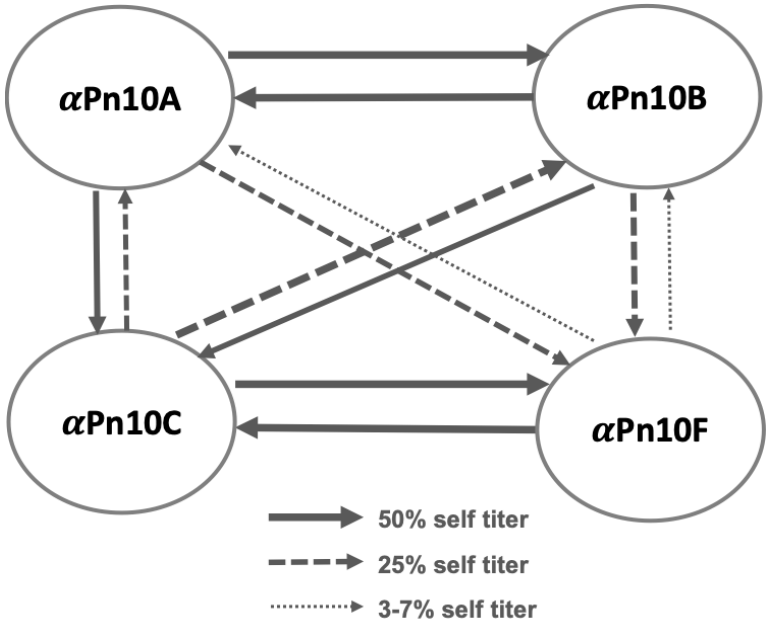


Figure 5.1: *S. pneumoniae serogroup 10 rabbit antisera cross-reactivity trends showing the cross-reactivity of antisera (α) raised against serogroup 10 CPSs (α Pn10A, α Pn10B, α Pn10C, and α Pn10F).* [3] *Self-cross-reactivity titer was set as 100 % with 50 % self-titer, 25 % self-titer, and 3 %–7 % self-titer represented by solid arrows, dashed arrows, and dotted arrows, respectively.*

Widespread introduction of PCVs has significantly reduced the burden of invasive pneumococcal disease (IPD) due to vaccine serotypes [10] However, increased prevalence of non-vaccine serotypes (due to serotype replacement as well as different geographical and socio-economic vaccination programs and prevalence of serotypes) means that IPD is still a leading cause of lower respiratory infection morbidity and mortality [11-13] This necessitated the development of the third generation of PCVs, with higher serotype valency and protection against the emerging serotypes. [14] A PCV15 vaccine (PCV13 plus serotypes 22F and 33F) and a PCV20 vaccine (PCV15 plus serotypes 8, 10A, 11A, 12F and 15B) have been licensed and higher valency vaccines such PCV24 (PCV20 plus serotypes 2, 9N, 17F and 20) and PCV30 are in development. [14-16]

The epidemiology of serogroup 10 varies quite significantly across different geographical regions, with Pn10A considered the most prevalent, followed by Pn10B, with Pn10C and Pn10F occasionally isolated but not considered a major burden of disease. [11,17-20] The prevalent serotype 10A is present in PPV23 and targeted for inclusion in the third generation PCVs: in PCV20 and PCV24. [6,15,21,22]

The literature reports considerable cross-reactivity within *S. pneumoniae* serogroup 10: between serotypes 10A (Pn10A), 10B (Pn10B), 10C (Pn10C), and 10F (Pn10F), summarized in **Figure 5.1**. [3,6] Specifically, serotype-specific rabbit antisera (α Pn) raised against one serotype recognized the CPS of the other serotypes, albeit with a lower affinity. [3] These data show some puzzling asymmetry, as follows. Strong reciprocal cross-reactivity was observed between Pn10A and Pn10B; as well as between Pn10C and Pn10F. Then, while Pn10A and Pn10B show strong cross-reactivity with Pn10C, the reciprocal cross-reactivity of Pn10C with these serotypes is weaker. Similarly, while Pn10A and Pn10B show moderate cross-reactivity with Pn10F, the reciprocal cross-reactivity of Pn10F is even weaker. The reasons for this asymmetry are unclear. Little information on the cross-reactivity and prevalence of the recently identified serotype 10D (Pn10D) is available. [8]

The repeating unit (RU) structures of the four main capsular polysaccharides in serogroup 10 (**Table 5.1** and **Figure 5.2**) contain the sugars β-D-galactofuranose (βDGal*f*), β-D-galactopyranose (βDGal), N-acetyl-

β-D-galactosamine (βDGalNAc), α-D-galactopyranose (αDGal), and D-ribitol-5-phosphate (Rib-ol-5P). [4-6] Pn10A and Pn10C share a common backbone as do Pn10B and Pn10F; side chain substitutions create four structurally distinct CPSs. Note that Pn10D has a considerably different backbone from the rest of serogroup 10 [8] and so was not included in this study.

Table 5.1: *Line structures of S. pneumoniae repeating units for serogroup 10 CPSs, side groups in bold.*

Pn10A:	[→5) βDGal*f*(1→3) βDGal*p*(1→4) **[βDGal*f*(1→3)] [βDGal*p*(1→6)]** βDGal*p*NAc(1→3) αDGal*p*(1→2) DRib-ol-5P(P→]
Pn10B:	[→5) βDGal*f*(1→3) βDGal*p*(1→4) **[βDGal*f*(1→3)]** βDGal*p*NAc(1→3) αDGal*p*(1→4) DRib-ol-5P(P→]
Pn10C:	[→5) βDGal*f*(1→3) βDGal*p*(1→4) **[βDGal*f*(1→6)]** βDGal*p*NAc(1→3) αDGal*p*(1→2) DRib-ol-5P(P→]
Pn10F:	[→5) βDGal*f*(1→3) βDGal*p*(1→4) **[βDGal*f*(1→6)]** βDGal*p*NAc(1→3) αDGal*p*(1→4) DRib-ol-5P(P→]
Pn10D:	[→6) αDGlc*p*(1→3) αDGlc*p*(1→4) **[βDGal*f*(1→3)] [βDGal*p*(1→6)]** βDGal*p*NAc(1→3) αDGal*p*(1→1) DRib-ol-5P(P→]

The backbones for Pn10A (**Figure 5.2A**) and Pn10C (**Figure 5.2C**) have a αDGal(1→2)Rib-ol-5P linkage Pn10B (**Figure 5.2B**) and Pn10F (**Figure 5.2D**) have a αDGal(1→4)Rib-ol-5P linkage. Furthermore, the substitutions on the branching βDGalNAc residue differ, as follows. Pn10B, Pn10C, and Pn10F are all singly substituted with βDGal*f*: at position three in Pn10B and at position six in both Pn10C and Pn10F. Pn10A has a double substitution on βDGalNAc: a βDGal*f* residue at position three and a βDGal residue at position six.

All serogroup 10 serotypes contain the β-D-galactofuranose monosaccharide - βDGal*f*. Absent in mammals, this sugar occurs in a range of bacterial pathogens (such as *Mycobacterium tuberculosis*, *Klebsiella pneumoniae* and *S. pneumoniae)* and is often critical for virulence and/or viability of the organism [24] and has been identified as an immunodominant epitope. [25] Human intelectin-1 (hIntL-1) is a carbohydrate binding protein in the human innate immune system that recognizes the exocyclic terminal 1,2-diol (ETD) epitope which is present in galactofuranose, but not galactopyranose. [26-28] This ETD motif is bound by

hIntL-1 with high specificity in molecules such as D-*glycero*-D-*talo*-oct-2-ulosonic acid (KO), 3-deoxy-D-*manno*-oct-2-ulopyranosonic acid (KDO), *β*DGal*f*, and glycerol-1-phosphate. [26,27] As binding by hIntL-1 is associated with increased phagocytosis of the pathogens, pathogens have evolved to evade the immune system by modifying the ETD motif (such as masking with O-acetylation) thus preventing hIntL-1 recognition. [29,30]

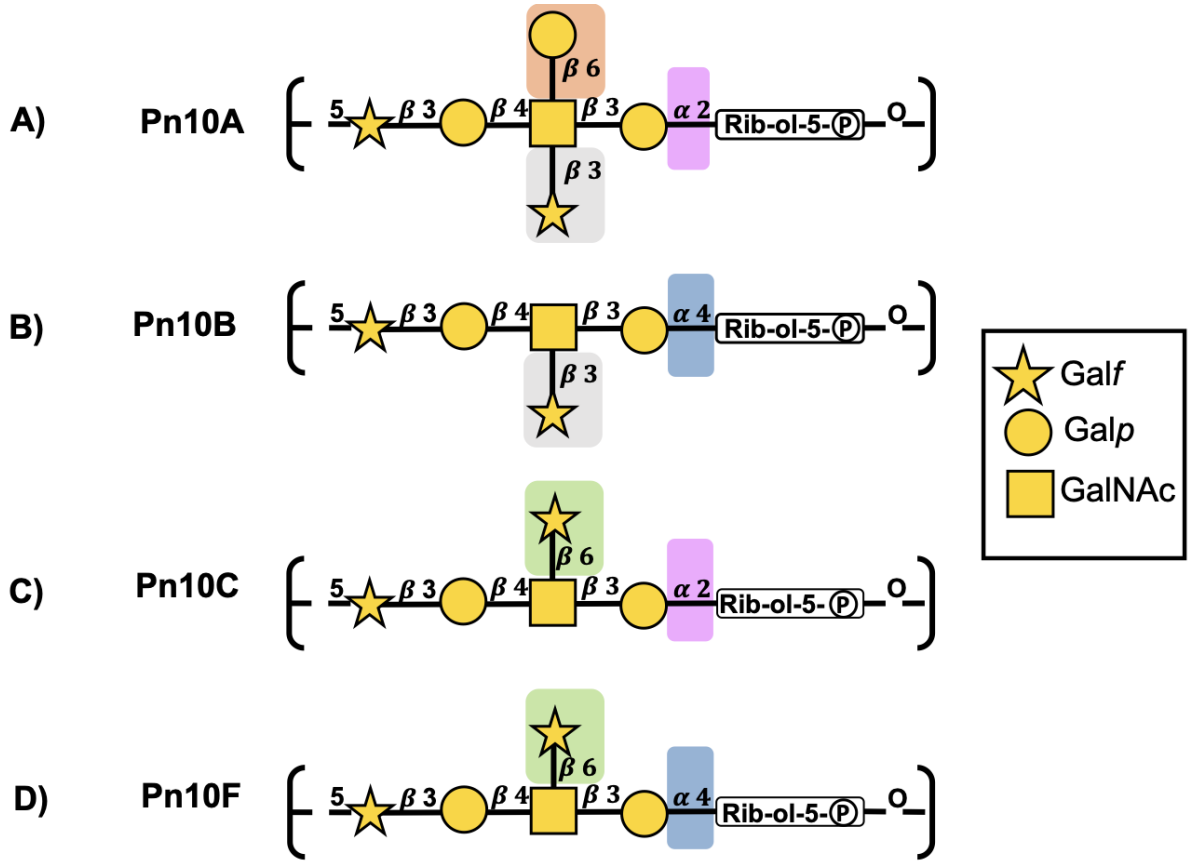


Figure 5.2: *S. pneumoniae serogroup 10 CPS repeating unit structures for* ***(A)*** *Pn10A,* ***(B)*** *Pn10B,* ***(C)*** *Pn10C, and* ***(D)*** *Pn10F represented with the SNFG (Symbol Nomenclature for Glycans) system.* [23] *Similarities and differences in repeating units are indicated by background color shading.*

Molecular modeling is a complementary technique that can be used to correlate the conformational features of carbohydrate antigens with data on serotype cross-protection produced by immunological studies (such as serological evaluations, animal trials, human trials, or epidemiological studies). [31-36] Recently, we have explored the importance of the presentation of structural features or epitopes on bacterial CPS for serotype

cross-protection. [37,38] Here we apply our established molecular modeling methodology to the four primary *S. pneumoniae* serogroup 10 CPSs (Pn10A, Pn10B, Pn10C and Pn10F) to establish their conformation and identify potential cross-protective epitopes. We aim to rationalize the observed cross-reactivity data and thus inform the design of the next generation of vaccines.

5.4 Materials and Methods

We used our established systematic approach to the modeling of polysaccharides, as previously described, [39,40] to build three repeating unit (3 RU) chains followed by 6 RU chains of Pn10A, Pn10B, Pn10C, and Pn10F in aqueous solution for initial molecular dynamics (MD) simulations.

Chain length is an important consideration when modeling CPSs, as a short chain may have insufficient molecular flexibility, while long chains are more computationally expensive to model. On the basis of our previous work, we consider a 6 RU chain to be representative of the behavior of the longer polysaccharide. Further, antibodies bind small fragments of the CPS between one and seven residues in length [41] corresponding to 1 RU (or a fraction thereof) in the case of *S. pneumoniae* making a 6 RU model sufficient to explore antibody binding epitopes.

Following initial system equilibration, MD simulations were performed of 3 RU and 6 RU chains of each of the CPSs. Data analyses were performed on these production runs, as described below.

5.4.1 Molecular dynamics

The 3 RU and 6 RU chains were built using in-house CarbBuilder software (version 1.2.42) [42,43] and visualized with the Visual Molecular Dynamics (VMD) software. [44] Starting structures for each molecule were built using low energy glycosidic linkages from potential mean force calculations, as per our previous work [40] and the psfgen tool was used by CarbBuilder to create protein structure (PSF) files for simulation with the CHARMM36 additive force field used to model the carbohydrates. [45,46] The starting structures were subsequently minimized with the Nanoscale Molecular Dynamics (NAMD) program (version 2.13) for 10 000 steps at 300 K. Minimized structures were solvated using VMD's built in solvation and ionization tools to add TIP3P [47] cubic water boxes of 84 Å per side for 3 RU systems and 140 Å per side for 6 RU systems. Systems were then neutralized with one sodium (Na^+) counter ion per repeating unit (three ions for the 3 RU, six for the 6 RU). Initial minimization and heating protocols comprised 5 K incremental

temperature reassignments beginning at 10 K up to 300 K with 500 steps of NAMD minimization and 8 000 steps of MD at each temperature reassignment. Solvated and ionized structures (PDB and PSF files) for each 6 RU system are available as Supplementary Material.

Simulations of 3 RU were run using NAMD (version 2.13) [48] with CUDA extensions for graphics processor acceleration; [49] simulations of 6 RU were run using NAMD (version 3.0) with CUDA extensions for graphics processor acceleration and a GPU resident computational mode.

Periodic boundary conditions equivalent to the cubic box size were employed for the solvated simulation with wrapping on. Long range electrostatics were implemented with the Particle Mesh Ewald summation grid spacing set to 1. [50] Atoms were not held fixed, and the initial center of mass motion was off. The 1-3 pairs were excluded from non-bonded interactions, 1-4 interactions were not scaled, and the dielectric constant was set to 1. Smoothing functions were applied to both the electrostatics and van der Waals forces with switching and cut-off distances of 10 Å and 12 Å, respectively.

A Leap-Frog Verlet integrator was used to integrate the equations of motion over a step size of 1 fs. A distance of 15 Å was used as the cut-off for inclusion in the pair list for calculation of non-bonded forces. The short-range non-bonded interactions were calculated every 1 fs, full electrostatics calculations were performed every 2 fs, and atoms were reassigned every 10 fs. [51]

Simulations were sampled under isothermal-isobaric (nPT) ensemble. Langevin dynamics [52] were used to control the temperature with a damping coefficient of 5/ps. Nosé-Hoover Langevin piston dynamics were used as a barostat to maintain a target pressure of 1 atm. [53,54] Variable system volume was used with a piston period of 100 fs and decay of 50 fs. Simulations of 3 000 ns were performed for the 6 RU systems (3 RU systems were run to 1 000 ns) comprising 200 ns of equilibration and 2 800 ns of production run as was required for convergence.

5.4.2 Convergence

We addressed convergence using block standard averaging [55] applied to two metrics: end-to-end distance and radius of gyration (**Supplementary Figure S5.1**). Block standard averaging was implemented with in-house Python scripts.

For all simulations, the blocked standard error (BSE) reached plateaus for both metrics, indicating convergence. The simulation lengths were large multiples of the correlation times for end-to-end distance (Pn10A, 74 ns; Pn10B, 110 ns; Pn10C, 69 ns; Pn10F, 99 ns) and radius of gyration (Pn10A, 70 ns; Pn10B, 108 ns; Pn10C, 69 ns; Pn10F, 120 ns). Further, the numbers of independent samples were >>1 for both the end-to-end distance (Pn10A, 40; Pn10B, 27; Pn10C, 45; Pn10F, 30) and the radius of gyration (Pn10A, 43; Pn10B, 28; Pn10C, 44; Pn10F, 25). Our designated equilibration time of 200 ns is therefore greater than the correlation time indicating that the properties of the system are not correlated or related on this timescale and that the simulation length should be sufficient.

5.4.3 Data analysis

Molecular conformations were visualized using VMD, with the PaperChain and Twister visualization algorithms used to highlight carbohydrate rings and chains [56] as required.

Trajectories were extracted at 25 ps intervals with analysis of performed on frames 250 ps apart. Metrics such as end-to-end distances were extracted from the simulation trajectories using Tcl scripting via VMD's Tk console. Data analyses performed with in-house Python scripts and plots generated using Matplotlib. [57]

5.4.3.1 Chain flexibility

The end-to-end distance, *r*, was measured from C1 of αDGal at the non-reducing end, to C1 of βDGal at the reducing end thus excluding the highly flexible terminal residues.

5.4.3.2 Conformational analysis

The most common chain conformations for each simulation were determined by clustering the production trajectory frames into families and calculating the relative occupancies of each family. Clusters comprising less than 6 % of the production run (post equilibration) were discarded. Clustering was performed using the WMC PhysBio plug-in for VMD's built-in measure cluster command. [58] Prior to clustering, the molecules were aligned on the RU 4 backbone excluding the Rib-ol-5P residue and any hydrogens on the backbone. Clustering was then performed as an RMSD fit to the ring and linkage atoms of the central four repeating units of the chains, excluding the highly flexible terminal RU 1 and RU 6. Eight clusters were created with a cut-off of 8 Å.

Clustering analysis was also performed on a single repeating unit, RU 4, and on the *β*DGalNAc – *β*DGal*f* linkage of RU 4. For the RU 4 analysis, molecules were aligned on the *β*DGalNAc residue of RU 4, excluding hydrogens. Clustering was then performed on the ring and linkage atoms of RU 4, creating five clusters with a cutoff of 1.5 Å, before discarding clusters comprising less than 10 % of the trajectory.

Similarly, for the *β*DGalNAc – *β*DGal*f* linkage, molecules were aligned on the *β*DGalNAc residue of RU 4 and clustering was then performed on the *β*DGalNAc and *β*DGal*f* ring and linkage atoms. Four clusters were created with a cut-off of 1.5 Å and clusters comprising less than 6 % of the production run (post-equilibration) were discarded.

5.4.3.3 hIntL-1 binding

The protein data bank (PDB) file describing the structure of human intelectin-1 (hIntL-1) bound to allyl-*β*-galactofuranose (allyl *β*Gal*f*) (PDB ID: 4WMY) was obtained from the official RSCB PDB. [28] Overlay of our CPS molecules with the hIntL-1 protein binding site were created by aligning the O4, C5, O5, C6, and O6 atoms of the central side group *β*DGal*f* residue from each CPS molecules with that of the allyl *β*DGal*f* in the hIntL-1 binding site. We then identified frames where the molecular conformation aligns with the hIntL-1 binding site free from protein-CPS intersection with good alignment of the galactofuranose rings

and exocyclic terminal-1,2-diol (ETD) moieties. For the 3 RU molecules, we aligned βDGal*f* from RU 2 and for the 6RU molecules, βDGal*f* from RU 3.

5.5 Results

We begin with a comparison of the flexibility of the serogroup 10 CPSs. This is followed by analysis of the CPS molecular conformation, exposed potential binding epitopes and, finally, an exploration of the potential for CPS binding to the hIntL-1 human innate immune protein.

5.5.1 Chain flexibility and conformation

The fluctuation in molecular end-to-end distance, *r*, over the length of the simulation is commonly used as a measure of chain extension and flexibility in carbohydrates. Here we define *r* to exclude the mobile terminal residues (**Figure 5.3A**). Times series plots of *r* for each CPS (**Figure 5.3B-E**, left column) reveal that all four molecules are highly flexible, fluctuating between a wide range of *r* values. Trajectory snapshots at 50 ns intervals (shown above the graphs) illustrate the considerable diversity in the molecular conformations over all the simulations. This flexibility is an expected consequence of a linear alditol (Rib-ol-5P) with a phosphodiester linkage in the CPS backbone: modeling of *Haemophilus influenzae* serotypes a and b CPS, which also contain this moiety, showed similarly flexible molecules. [38]

Comparison of the *r* distributions (**Figure 5.3B-E**, right column) reveals some broad differences between the serotypes: Pn10A and Pn10F have bimodal distributions, whereas Pn10B and Pn10C are closer to unimodal, narrower and somewhat shifted to smaller values of *r*. Interestingly, it is the most structurally dissimilar pairs of molecules – Pn10A and Pn10F (**Figure 5.2B**,**E**); Pn10B and Pn10C (**Figure 5.2C**,**D**) – that are most similar with respect to the *r* distributions. While all four molecules are flexible, these data suggest an order of flexibility of Pn10A ~ Pn10F > Pn10B > Pn10C and, relatively speaking, a slightly increased incidence of extended conformations for Pn10A and Pn10F.

The extreme flexibility of the CPS backbone means that there are no dominant chain conformations for any of the four carbohydrates (**Figure 5.4**). For RU 2 to RU 5 of the CPS backbone (i.e. 20 backbone residues) we found very few significant conformational families that occupy more than 6 % of the simulation. Further,

those we identified accounted for relatively small portions of the simulation (less than 25 %). This indicates that these flexible chains behave as random coils, in common with other flexible carbohydrates that we have modeled. [38,40]

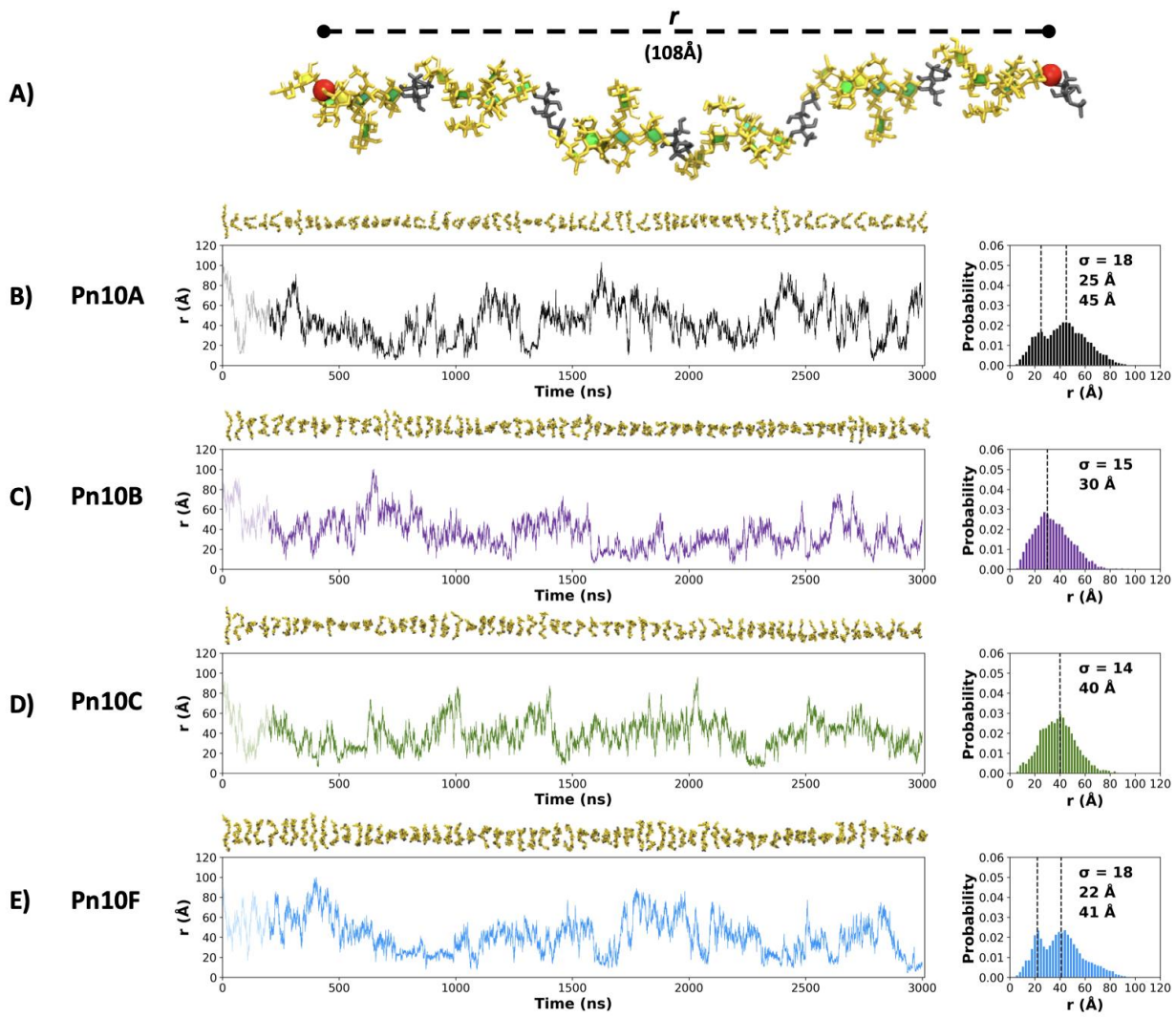


Figure 5.3: (A) *The end-to-end distance, r, is indicated on the 6 RU Pn10A molecule. Time series graphs (left column) of r and corresponding histograms (right column) for the 3 000 ns simulation trajectories are shown for:* ***(B)*** *Pn10A,* ***(C)*** *Pn10B,* ***(D)*** *Pn10C, and* ***(E)*** *Pn10F. Conformational snapshots at 50 ns intervals are shown above the time series plots. The histograms are labeled with the standard deviations (σ) and modal peak r value(s).*

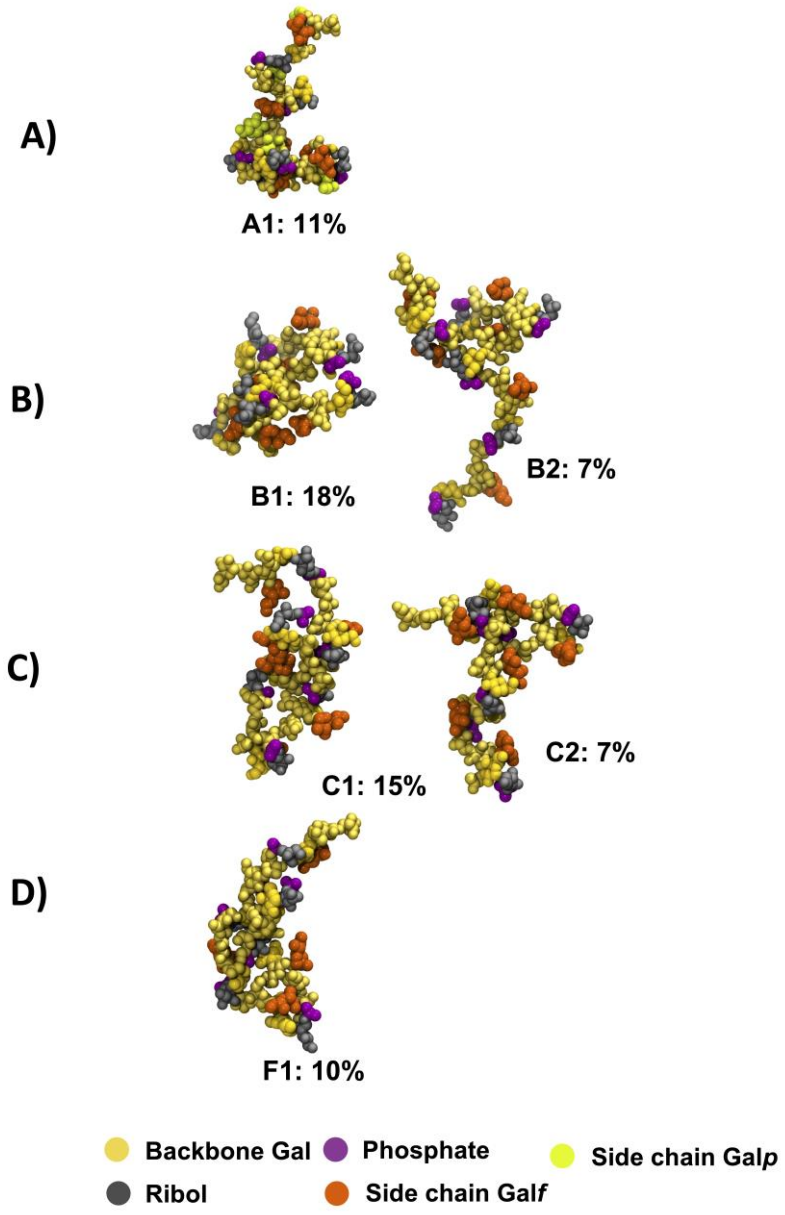


Figure 5.4: *Main conformational families (seed frame for each cluster is shown) identified for the 6 RU S. pneumoniae serotype CPS, with the terminal repeat units (RU 1 and RU 6) excluded from analysis:* ***(A)*** *Pn10A,* ***(B)*** *Pn10B,* ***(C)*** *Pn10C, and* ***(D)*** *Pn10F. Backbone galactose residues are shown in yellow, ribitol residues shown in grey, phosphate in purple, side group galactofuranose residues in orange, and side group galactopyranose residues in light green. The primary clusters occurred regularly throughout the trajectory. The secondary cluster of Pn10B occurred regularly in the second half of the simulation along with the primary cluster; for Pn10C the secondary cluster occurred only for ~400 ns around the 1 500 ns timestep.*

We therefore conclude that none of these flexible CPS show evidence for a conformational epitope. However, most of the CPS chain flexibility originates from the backbone phosphodiester linkage to the Rib-ol-5P in the CPS backbone, with the other glycosidic linkages being relatively constrained (**Supplementary Figure S5.2**), potentially forming stable epitopes. Further, in all four molecules, the side chain *β*DGal*f* and *β*DGal residues are solvent exposed (**Figure 5.4**) and are therefore accessible for antibody binding. As antibodies bind to carbohydrate epitopes that typically comprise one to seven residues, [41] it proved more useful to compare the 6 RU CPS we modeled on this smaller length scale.

5.5.2 Epitopes

The principal conformations for the single central repeating unit (RU 4) in each of the CPS chains are shown in **Figure 5.5**. With the highly mobile phosphodiester linkage excluded, we find dominant conformations with high occupancies ranging from 30 % to 80 % for the primary cluster and 13 % to 24 % for the secondary cluster; more than 50 % of the simulation trajectory falls within these dominant clusters. For Pn10A (**Figure 5.5A**) the 6-linked *β*DGal side chain (yellow) protrudes from the backbone and is located close to the phosphate group, providing a conformational rational for the dominance of *β*DGal as an antibody binding epitope for Pn10A. [59] In contrast, the immunodominant *β*DGal*f* side chain in Pn10A is 3-linked and aligned with the backbone and thus less exposed for binding. This is also the case for the 3-linked *β*DGal*f* side chain in Pn10B (**Figure 5.5B**), which is in a similar orientation to Pn10A. However, Pn10C and Pn10F (**Figure 5.5C** and **D**) have a 6-linked *β*DGal*f* side chain, which is considerably more exposed and thus a more promising epitope. This 6-linked side chain is also more flexible than the 3-linked *β*DGal*f* in Pn10A and Pn10B; it is oriented in different directions across the dominant conformational clusters in Pn10C and Pn10F. As the *β*DGal*f* residue is conserved across the serogroup, this immunodominant epitope could form the basis of the observed serotype cross-reactivity. In particular, for *β*DGal*f*, the exocyclic 1,2-terminal diol (ETD, colored green in **Figure 5.5**) is a potential epitope for CPS binding of the hIntL-1 human innate immune protein. A second ETD from the 4,5-linked ribitol of Rib-ol-5P is present in Pn10B and Pn10F, shown in cyan in **Figure 5.5B** and **D**.

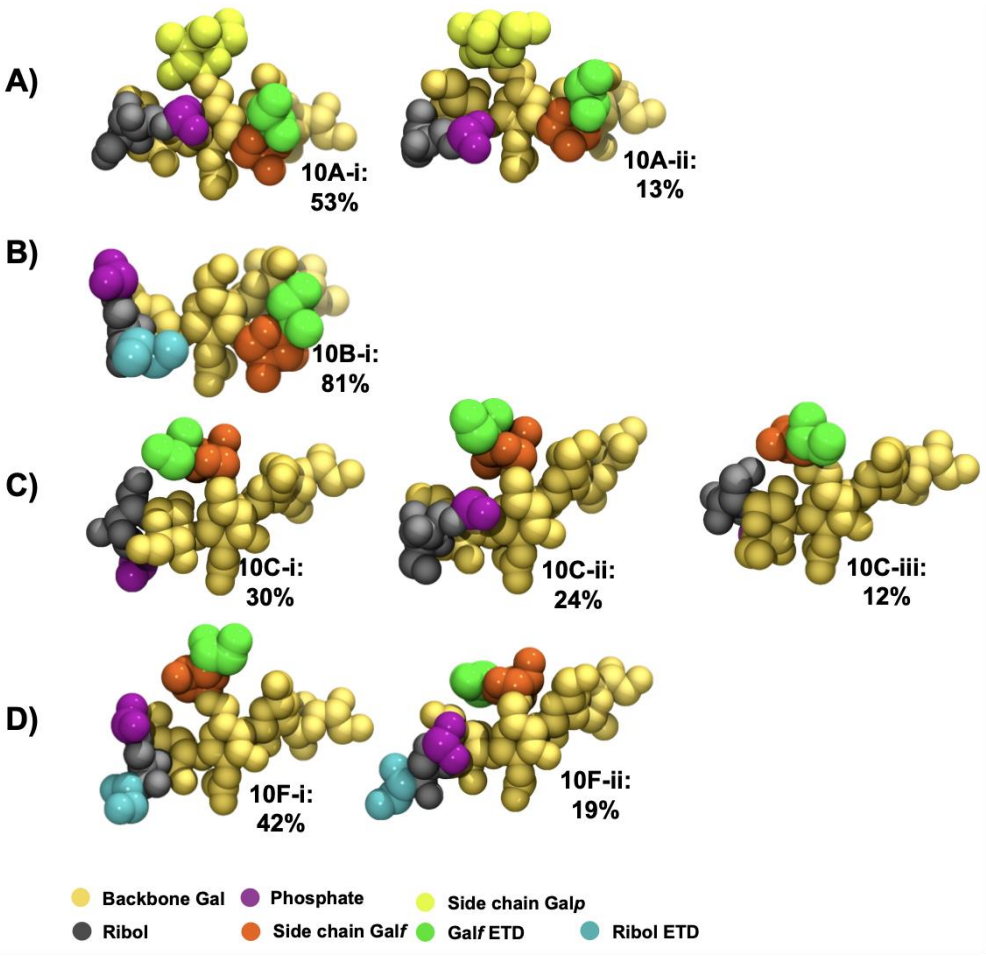


Figure 5.5: *Main conformational families (seed frame for each cluster is shown identified for RU 4 in the backbone of 6 RU S. pneumoniae serotype CPS:* ***(A)*** *Pn10A,* ***(B)*** *Pn10B,* ***(C)*** *Pn10C, and* ***(D)*** *Pn10F. Backbone galactose residues shown in yellow, Rib-ol-5P residues shown in grey, phosphate in purple, side group βGalf residues in orange, side group βGal residues in light green, ETD portion of βGalf in green, and ETD portion of Rib-ol-5P residues in cyan.*

To explore the range of conformations of the immunodominant *β*DGal*f* epitope in finer detail, we focus on the conformations of the *β*DGalNAc – *β*DGal*f* disaccharide in the CPS chain in **Figure 5.6**. The primary (blue) and secondary (red) conformational families are super-imposed in **Figure 5.6A** for all four serotypes. It is clear that 3-linked *β*DGal*f* (Pn10A and Pn10B) side chains have a single dominant conformation (> 90 % occupancy), whereas Pn10C and Pn10F, with 6-linked *β*DGal*f*, do not. Pn10C has two main conformations (56 % and 20 % occupancy, respectively) while Pn10F has three (37 %, 20 %, and 17 % occupancy, respectively). Comparison of representative structures for each conformational family suggests

four distinct conformational epitopes (EP) of the *β*DGalNAc – *β*DGal*f* disaccharide across the serotypes (**Figure 5.6B**). Pn10A and Pn10B share the same dominant conformational epitope of 3-linked *β*DGal*f*, which we term EP1. Pn10C and Pn10F also share a main conformation for their 6-linked *β*DGal*f*, which we term EP2, as well as a secondary conformation (EP3). Pn10F also has a unique conformation of this disaccharide, which we term EP4. Further - considering the *β*DGal*f* and ETD orientations relative to the backbone - EP1 and EP3 are similar while EP2 is related to EP3 by a 180° rotation; EP4 is a distinct epitope.

These conformational epitopes identified for *β*DGal*f* help to rationalize the trends in cross-reactivity observed for serogroup 10 (**Figure 5.1**), as follows. Firstly, the strong reciprocal cross-reactivity of Pn10A with Pn10B can be attributed to the shared EP1 epitope (**Figure 5.6B**, first panel). Similarly, the shared epitopes EP2 and EP3 explain the strong reciprocal cross-reactivity between Pn10C and Pn10F (**Figure 5.6B**, second and third panels). Then, the cross-reactivity of Pn10A/Pn10B with Pn10C/Pn10F can be explained by the conformational similarities between EP1 and EP2/EP3 (**Figure 5.6B**, first three panels). Further, for Pn10A/Pn10B the cross reactivity is stronger with Pn10C than with Pn10F, because the combined prevalence of the EP2 and EP3 epitopes is considerably more in Pn10C (76 %) than in Pn10F (57 %). Likewise, the weaker reciprocal cross-reactivity of Pn10C with these serotypes can be explained by the lower prevalence of the EP2 and EP3 conformational epitopes in Pn10C – fewer antibodies generated against these epitopes would hence, be less likely to recognize EP1 dominant in Pn10A/Pn10B. Similarly, the still weaker reciprocal cross-reactivity of Pn10F can be explained by the even lower expression of EP2 and EP3 as well as an additional conformational epitope, EP4, not shared by any of the other serotypes.

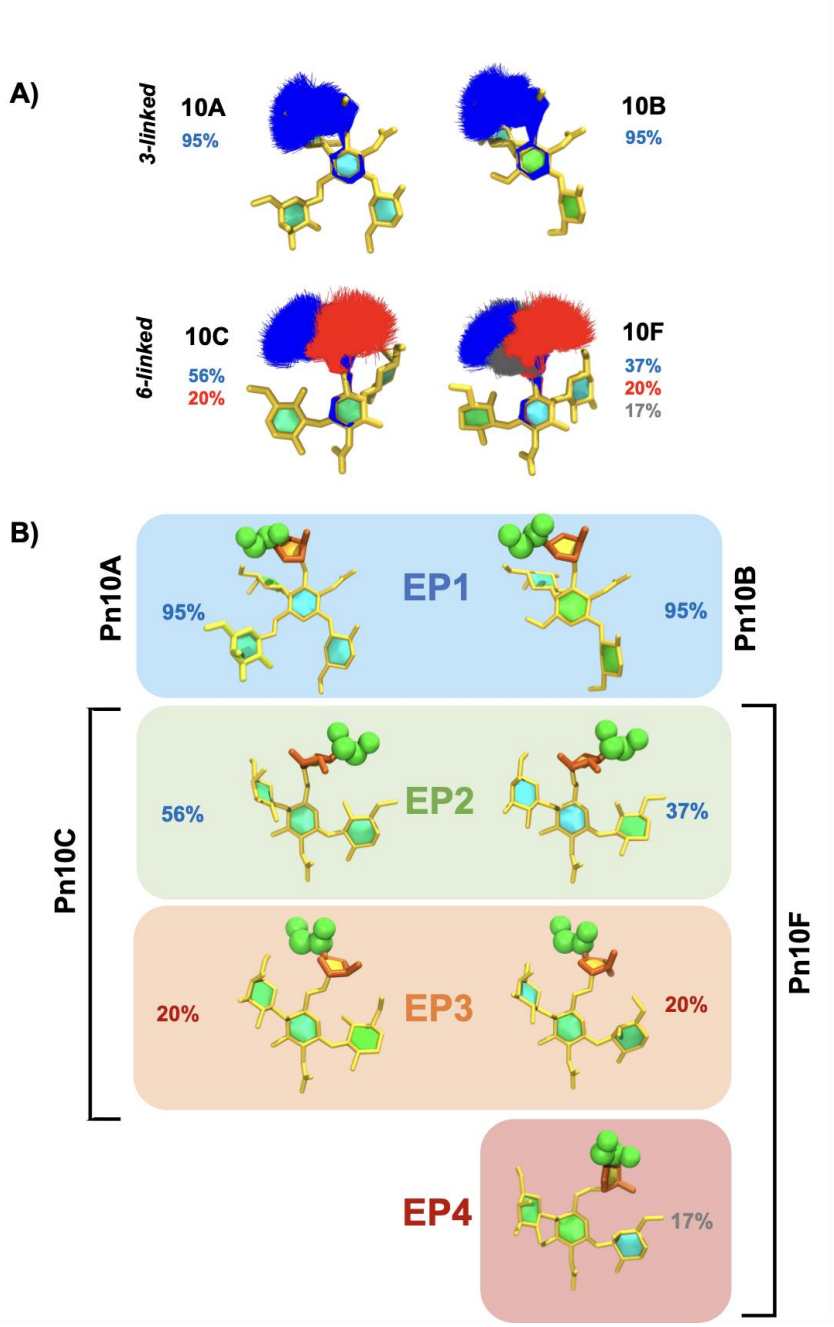


Figure 5.6: *Conformational epitopes of the βDGalNAc – βDGalf disaccharide in S. pneumoniae serogroup 10, RU 4.* ***(A)*** *Super-imposed conformational families (> 5 %) of the βDGalNAc –βDGalf disaccharide with associated percentages for the four CPS molecules.* ***(B)*** *Representative conformational epitopes (EP1-4) of theβDGalNAc – βDGalf disaccharide for the four CPS molecules, with associated percentage occupancies, grouped according to conformational similarity.*

5.5.3 hIntL-1 binding

The exocyclic 1,2-terminal diol on *β*DGal*f* (ETD) is a potential epitope for CPS binding to the hIntL-1 human innate immune protein. We used the crystal structure of hIntL-1 bound to allyl *β*DGal*f* [26,27] as a template for possible binding of the serogroup 10 CPS. Example conformations of the 3RU CPS with the central *β*DGal*f* ETD in the chain super-imposed on the ETD in the hIntL-1 binding site are shown in **Figure 5.7B-E**. These structures illustrate that it is possible for hIntL-1 to bind the EP1/EP2 epitope without involvement of the CPS backbone. However, the 3-linked *β*DGal*f* on Pn10A and Pn10B results in closer proximity of the CPS backbone to the protein than the 6-linked *β*DGal*f* on Pn10C and Pn10F, suggesting that these serotypes will bind less readily. For the 6 RU CPS chains we identified similar conformations (**Supplementary Figure S5.3**). Interestingly, the more collapsed conformations (**Supplementary Figure S5.3A** and **B**) have better potential for binding, as the *β*Gal*f* side chain projects from the backbone, with consequently less protein-CPS entanglement with the flexible backbone. Lastly, Supplementary Material **Figure 5.4** presents the potential hIntL-1 binding of the ETD of Rib-ol-5P present in Pn10B and Pn10F.

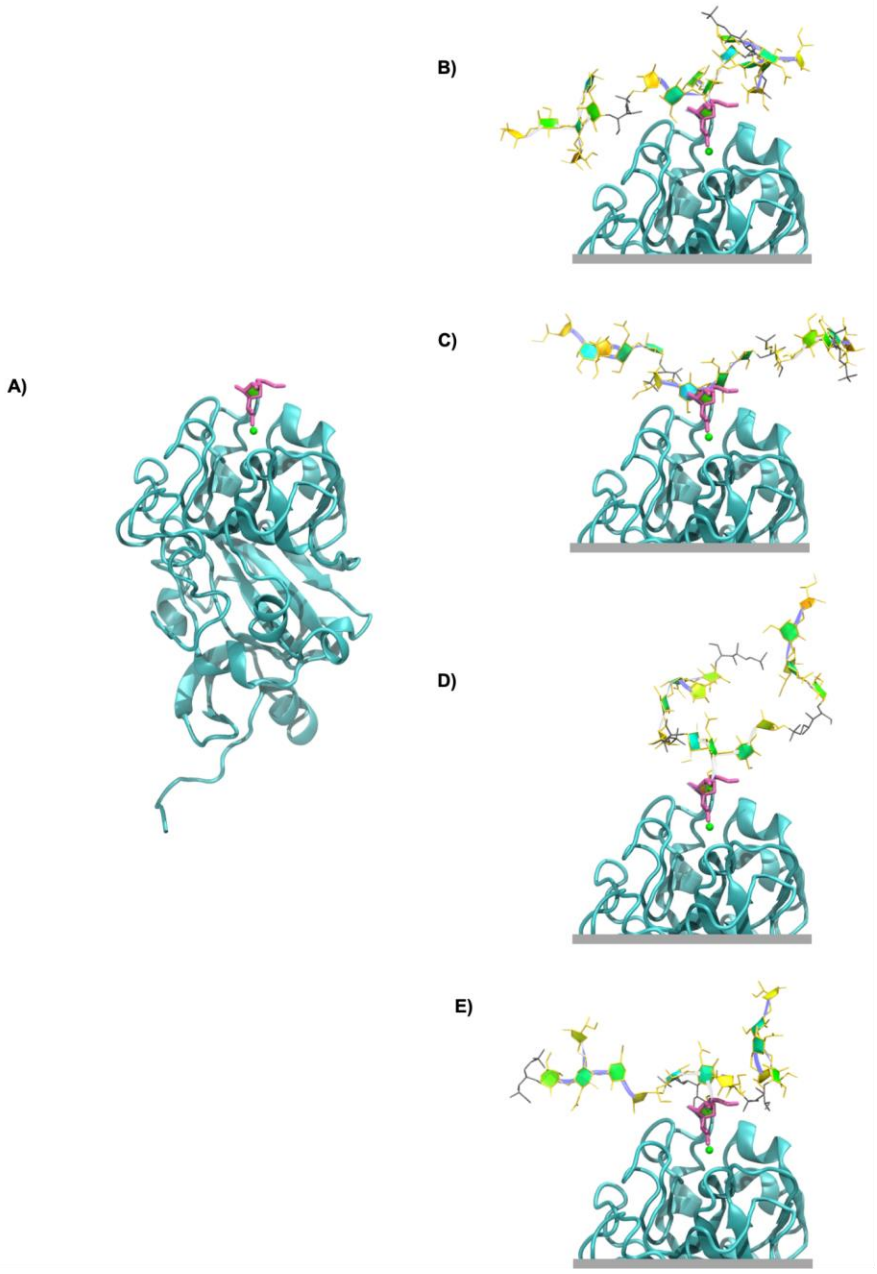


Figure 5.7: *Human intelectin-1 (hIntL-1) binding.* ***(A)*** *Monomer of the crystal structure of hIntL-1 bound with allyl-beta-galactofuranose (PDB ID: 4WMY). Example conformations of the 3 RU CPSs with the DGalf side chain positioned in the hIntL-1 binding site are shown for* ***(B)*** *Pn10A,* ***(C)*** *Pn10B,* ***(D)*** *Pn10C, and* ***(E)*** *Pn10F. The end-to-end distance (r) of these conformations are 33 Å, 41 Å, 16 Å, and 33 Å for Pn10A, Pn10B, Pn10C, and Pn10F respectively. Overlay of our CPS molecules with the hIntL-1 protein binding site were created by aligning the O4, C5, O5, C6, and O6 atoms of the central side group (RU 2) βDGalf residue from each CPS molecule with that of the allyl βDGalf in the hIntL-1 binding site. We then identified frames where the molecular conformation aligns with the hIntL-1 binding site free from protein-CPS intersection with good alignment of the galactofuranose rings and exocyclic terminal-1,2-diol (ETD) moieties.*

5.6 Discussion

Our simulations show that *S. pneumoniae* serogroup 10 have highly flexible CPSs, which is unsurprisingly due to the backbone linear Rib-ol-5P alditol. Highly flexible bacterial CPSs are common and we have previously suggested that this is a strategy for bacterial evasion of the host immune system: a flexible backbone with multiple conformations presents a "moving target" to the immune system. [60] For serogroup 10, this mobility may be especially necessary given the presence of an immunodominant PO_4 group [61] and the presence of a non-mammalian galactofuranose sugar to which the human immune system has a natural defense. Between the highly mobile phosphodiester linkages to Rib-ol-5P, the segments of the CPS chain are relatively rigid and the *β*DGal*p* and *β*DGal*f* side chains form exposed and well-defined epitopes. In particular, the *β*DGal*f* side chains are common to all four serotypes and considered to be immunodominant. [25] Our analysis of the conformational epitopes of *β*DGal*f* provides a rationalization of the observed cross-reactivity between the serotypes in serogroup 10. We suggest that the varying flexibilities found for the four DGal*f* conformational epitopes identified has relevance for their observed complex patterns of cross-reactivity, as follows. The relatively immobile 3-linked *β*DGal*f* forms the EP1 epitope in Pn10A and Pn10B; it is dominant (95 %) in both serotypes and therefore there is strong cross-reactivity between them. This suggests cross-protection of a Pn10A vaccine against Pn10B, and also accounts for the reported mistyping of a Pn10A isolate as Pn10B. [2]

The more flexible 6-linked *β*DGal*f*, present in both Pn10C and Pn10F, gives rise to the conformational epitopes EP2 and EP3, which are similar to the EP1 epitope and suggests some cross-protection for a Pn10A vaccine against both Pn10C and Pn10F. The relative proportions of EP2 and EP3 expressed in the serotypes accounts for the asymmetry of cross-reactivity (**Figure 5.1**), as the stronger cross-reactivity of Pn10A/Pn10B with Pn10C than with Pn10F corresponds to a higher prevalence of the EP2/EP3 epitopes in Pn10C (76 % versus 57 %). This is in line with our modeling of the CPS from meningococcal serogroups Y and W: [62] we found that the less flexible serogroup Y CPS has a single dominant conformation, which corresponds to only one of several conformational epitopes exhibited by the closely related, but more

flexible, serogroup W CPS. This explained the asymmetric cross-protection observed in the clinic: vaccination with serogroup Y elicited cross-protection against serogroup W in the majority of subjects, whereas minimal cross-protection was observed against serogroup Y following serogroup W vaccination. Further, the *β*DGal*f* EP1 epitope could correspond to the factor serum 10d epitope reported in the literature, as 10d reacts strongly with Pn10A and Pn10B (displaying EP1), but only weakly with Pn10C and Pn10F (displayin[7,63]2/EP3). [8,63]

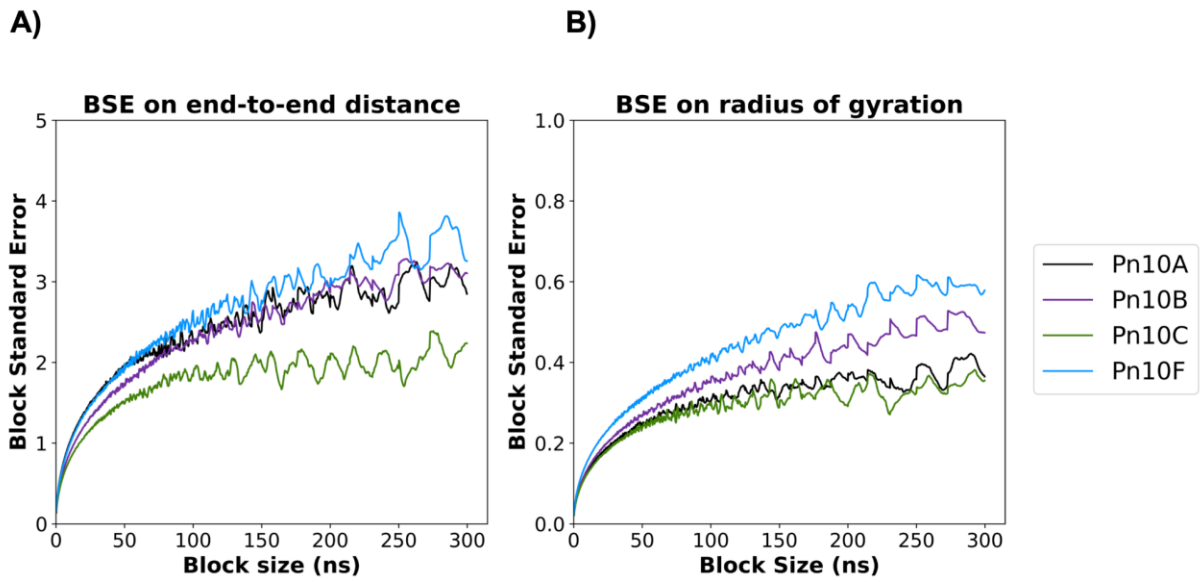
Pn10C and Pn10F also share both the *β*DGal*f* EP2 and EP3 conformational epitopes, which explains their strong mutual cross-reactivity, although Pn10F has an additional conformational epitope EP4 (17 %) not present in the other serotypes.

The recently discovered serotype Pn10D has a different backbone (and was not modeled in this study), but also exposes the 6-linked *β*DGal*p* and 3-linked *β*DGal*f* side chains found in Pn10A. [8] As expected, immunization with a polysaccharide vaccine containing Pn10A raised cross-opsonic antibodies against Pn10D, further confirming the inclusion of serotype 10A in a vaccine to best protect against serogroup 10 disease. These side chains are also present in *S. pneumoniae* serotype 39 which shows cross-reaction with factor serum 10d that reacts with Pn10A. [21] Partial O-acetylation on the terminal *β*DGal*f* in serotype 39 may be an example of masking of this key epitope to escape innate immunity. [64,65]

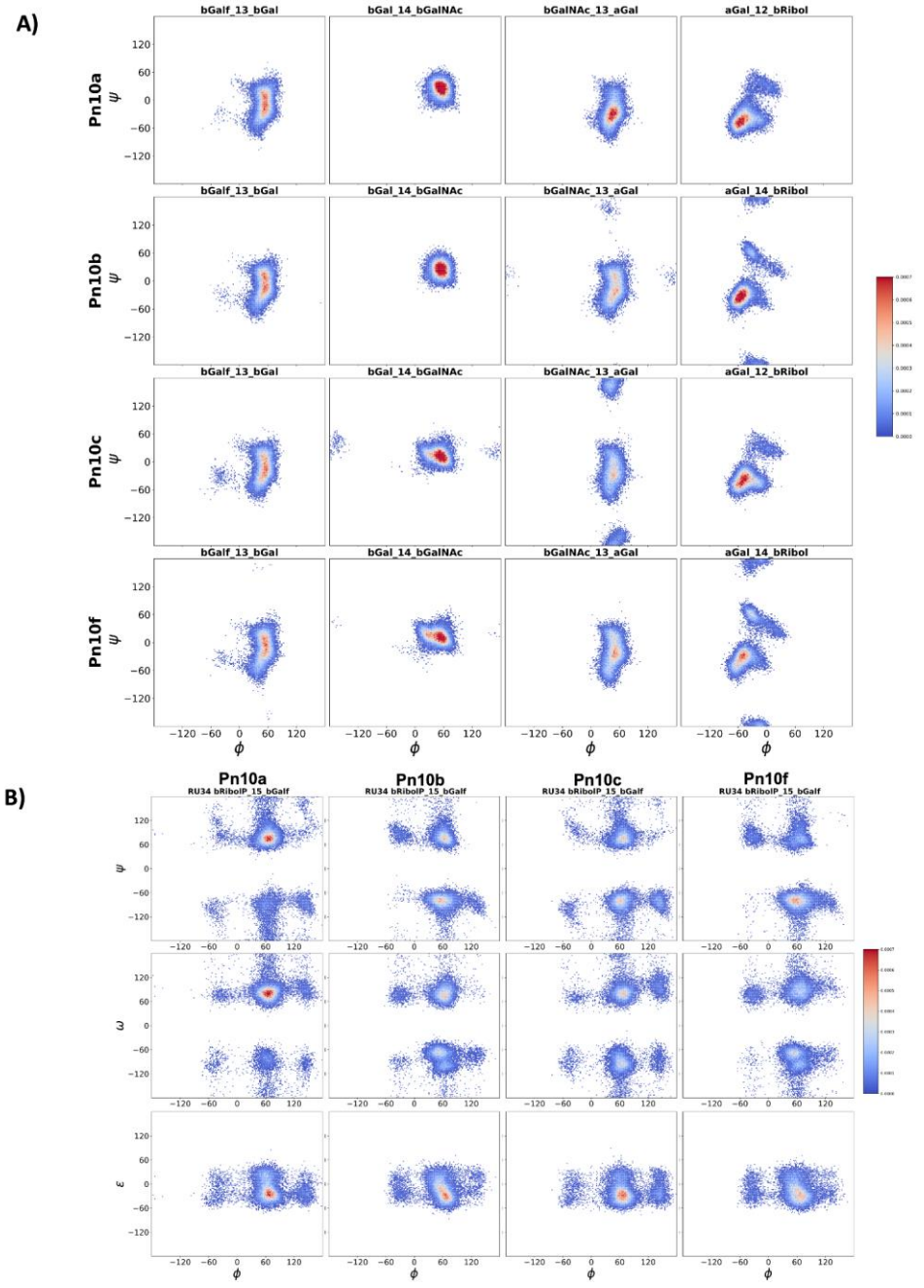
Finally, we showed that the hIntL-1 human innate immune protein has the potential to bind the EP1/EP2 epitope in serogroup 10, containing the exocyclic 1,2-terminal diol on *β*DGal*f*. Binding of 3-linked *β*DGal*f* EP1 epitope in Pn10A and Pn10B results in closer proximity of the CPS backbone to the protein than for the more accessible 6-linked *β*DGal*f* EP2 epitope in Pn10C and Pn10F, suggesting that serotypes Pn10A/Pn10B will be bound less readily. This may play a role in the strength of the innate immune response and hence serotype disease prevalence with more cases reported for Pn10A/Pn10B than Pn10C/Pn10F.

In conclusion, the common epitopes we have identified for serogroup 10 support a vaccine containing serotype 10A providing adequate cross-protection against serogroup 10 disease, as indicated by serological studies.

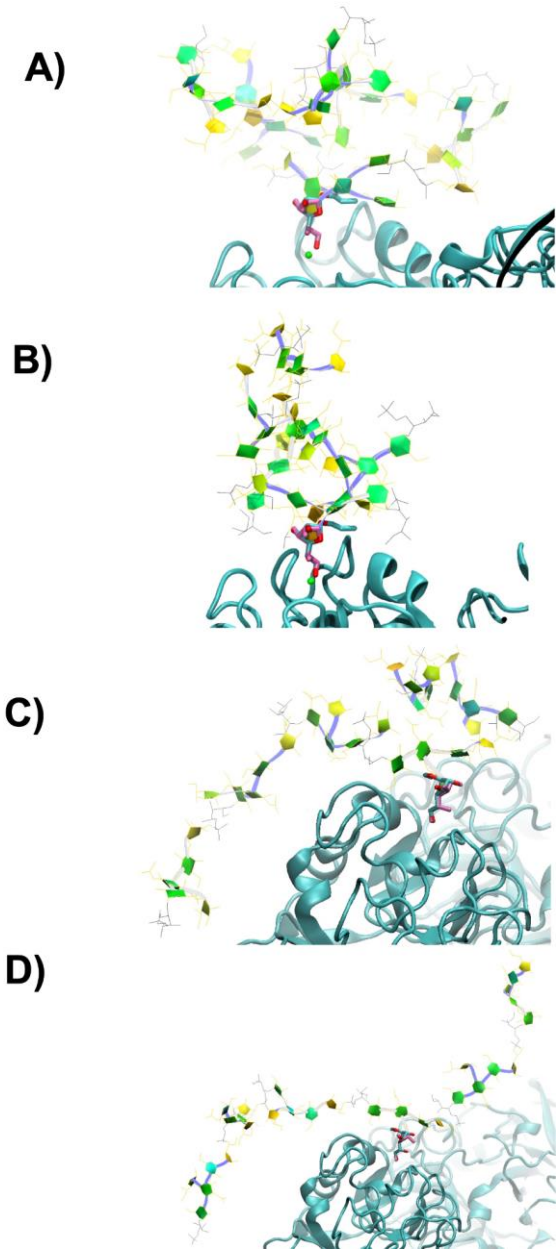
5.7 Supplementary Material



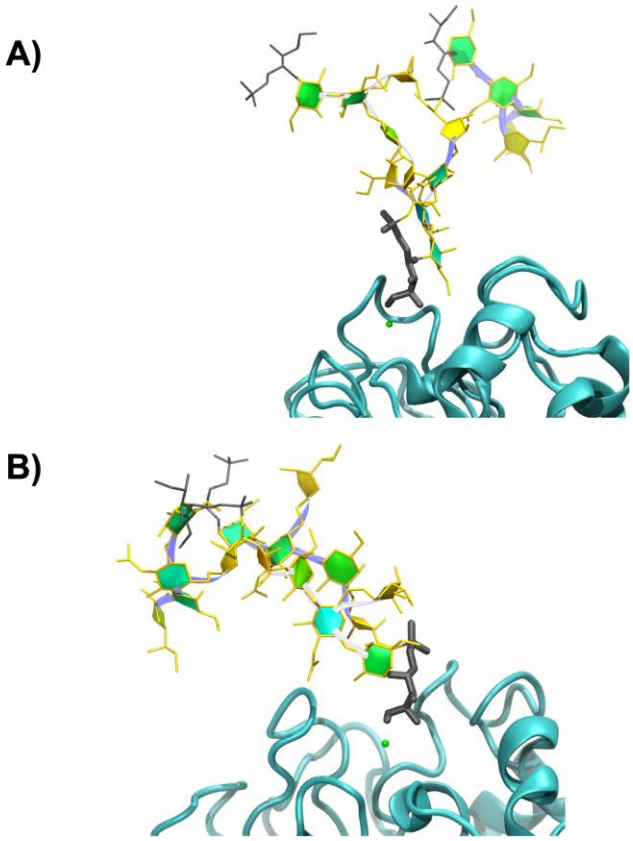
Supplementary Figure S5.1: *Block standard averaging analysis for modeled S. pneumoniae serogroup 10 CPS molecules.* ***(A)*** *block standard error (BSE) versus block size (ns) calculated on end-to-end distance.* ***(B)*** *BSE versus block size calculated on radius of gyration. For all molecules, the BSE visually reaches a plateau.*



Supplementary Figure S5.2: *Heatmap plots of dihedral angles for backbone linkages.* ***(A)*** *glycosidic linkages phi (ϕ) vs psi (ψ) for the βDGalf – βDGal; βDGal – βDGalNAc; βDGalNAc – αDGal; and αDGal – βDRib-ol-5P linkages for each serotype.* ***(B)*** *phosphodiester linkages ϕ vs ψ, omega (ω), and epsilon (ε) for the βDRib-ol-5P – βGalf linkage of each serotype. The phosphodiester linkages are more flexible than the glycosidic linkages. The linkages were defined as we have done previously for these types of linkages.* [38] *Glycosidic linkages were defined as ϕ = H1-C1-O1-Cx' and ψ = C1-O1-Cx'-Hx' and the phosphodiester linkages were defined as: ϕ = Hx-Cx-Ox-P, ψ = Cx-Ox-P-Oy, ω = Ox-P-Oy-Cy and ε = P-Oy-Cy-Hy.*



Supplementary Figure S5.3: *Alignment of human intelectin-1 bound to allyl- -galactofuranose (hIntL-1) with S. pneumoniae serogroup 10 CPS βDGalf ETD moieties. Alignment of 6 RU CPSs RU 4 side group βDGalf with hIntL-1 binding site for:* ***(A)*** *Pn10A,* ***(B)*** *Pn10B,* ***(C)*** *Pn10C, and* ***(D)*** *Pn10F. All four molecules align well and would be expected to bind. It was noted that collapsed chain conformations as in* ***(A)*** *and* ***(B)*** *were easier to align as there was less CPS-protein intersection.*



Supplementary Figure S5.4: *Alignment of human intelectin-1 bound with allyl-β-galactofuranose (hIntL-1) with S. pneumoniae serogroup 10 CPS Rib-ol-5P ETD moieties. Alignment of 3 RU CPSs RU 42Rib-ol-5P ETD with hIntL-1 binding site for:* ***(A)*** *Pn10B,* ***(B)*** *Pn10F. Adequate fitting of Rib-ol-5P in the hIntL-1 binding site was achieved by aligning C6, O6, and C3 of the allyl- -galactofuranose with O1, C1, and C4 of the Rib-Ol-5P molecule. While we were able to fit the ETD moiety in the hIntL-1 binding site, the stereochemistry of the Rib-ol-5P ETD is opposite to that of molecules expected to bind hIntL-1 and as such further studies are required to establish binding ability of the Rib-ol-5P ETD with hIntL-1.* [26,27]

5.8 References

1. N. I. Richardson, M. M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity. *Front. Mol. Biosci.*, 2022, **9**, 961532.

2. H. B. Konradsen, Validation of serotyping of Streptococcus pneumoniae in Europe. *Vaccine*, 2005, **23**, 1368-1373.

3. J. Henrichsen, Six newly recognized types of Streptococcus pneumoniae. *J. Clin. Microbiol.*, 1995, **33**, 2759-2762.

4. K. A. Geno, G. L. Gilbert, J. Y. Song, I. C. Skovsted, K. P. Klugman, et al, Pneumococcal capsules and their types: past, present, and future. *Clin. Microbiol. Rev.*, 2015, **28**, 871-899.

5. J. Yang, N. Y. Shelat, C. A. Bush and J. O. Cisar, Structure and molecular characterization of Streptococcus pneumoniae capsular polysaccharide 10F by carbohydrate engineering in Streptococcus oralis. *J. Biol. Chem.*, 2010, **285**, 24217-24227.

6. J. Yang, M. H. Nahm, C. A. Bush and J. O. Cisar, Comparative structural and molecular characterization of Streptococcus pneumoniae capsular polysaccharide serogroup 10. *J. Biol. Chem.*, 2011, **286**, 35813-35822.

7. M. Moreau and D. Schulz, Polysaccharide based vaccines for the prevention of pneumococcal infections. *J. Carbohydr. Chem.*, 2000, **19**, 419-434.

8. F. Ganaie, J. S. Saad, L. McGee, A. J. van Tonder, S. D. Bentley, et al, A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. *Am. Soc. Microbiol.*, 2020, **11**, 937.

9. A. A. Lindberg, Glycoprotein conjugate vaccines. *Vaccine*, 1999, **17**, S28-S36.

10. B. Wahl, K. L. O'Brien, A. Greenbaum, A. Majumder, L. Liu, et al, Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob. Health*, 2018, **6**, e744-e757.

11. E. Balsells, L. Guillot, H. Nair and M. H. Kyaw, Serotype distribution of Streptococcus pneumoniae causing invasive disease in children in the post-PCV era: a systematic review and meta-analysis. 2017, **12**, e0177113.

12. GBD 2016 Lower Respiratory Infections Collaborators, Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. 2018, **18**, 1191.

13. M. D. Wasserman, J. Perdrizet, L. Grant, K. Hayford, S. Singh, et al, Clinical and Economic Burden of Pneumococcal Disease Due to Serotypes Contained in Current and Investigational Pneumococcal Conjugate Vaccines in Children Under Five Years of Age. 2021, **10**, 2701-2720.

14. K. P. Klugman and G. L. Rodgers, Time for a third-generation pneumococcal conjugate vaccine. *Lancet Infect. Dis.*, 2021, **21**, 14-16.

15. E. Janssens, J. Flamaing, C. Vandermeulen, W. E. Peetermans, S. Desmet and P. De Munter, The 20-valent pneumococcal conjugate vaccine (PCV20): expected added value. *Acta Clin. Belg.*, 2022, **78**, 78-86.

16. M. Kobayashi, Use of 15-Valent Pneumococcal Conjugate Vaccine and 20-Valent Pneumococcal Conjugate Vaccine Among US Adults: Updated Recommendations of the Advisory Committee on Immunization Practices—United States, 2022. *MMWR Morb Mortal Wkly Rep*, 2022, **71.**

17. S. R. dos Santos, L. F. Passadore, E. H. Takagi, C. M. Fujii, C. R. Yoshioka, et al, Serotype distribution of Streptococcus pneumoniae isolated from patients with invasive pneumococcal disease in Brazil before and after ten-pneumococcal conjugate vaccine implementation. *Vaccine,* 2013, **31**, 6150-6154.

18. E. Jauneikaite, J. M. Jefferies, M. L. Hibberd and S. C. Clarke, Prevalence of Streptococcus pneumoniae serotypes causing invasive and non-invasive disease in South East Asia: a review. *Vaccine,* 2012, **30**, 3503-3514.

19. M. Messaoudi, M. Milenkov, W. C. Albrich, M. P. van der Linden, T. Bénet, et al, The relevance of a novel quantitative assay to detect up to 40 major Streptococcus pneumoniae serotypes directly in clinical nasopharyngeal and blood specimens. *PLoS One,* 2016, **11**, e0151428.

20. I. Wouters, S. Desmet, L. Van Heirstraeten, S. Blaizot, J. Verhaegen, et al, Follow-up of serotype distribution and antimicrobial susceptibility of Streptococcus pneumoniae in child carriage after a PCV13-to-PCV10 vaccine switch in Belgium. *Vaccine,* 2019, **37**, 1080-1086.

21. C. Jones, Full assignment of the NMR spectrum of the capsular polysaccharide from Streptococcus pneumoniae serotype 10A. *Carbohydr. Res.*, 1995, **269**, 175-181.

22. D. McGuinness, R. M. Kaufhold, P. M. McHugh, M. A. Winters, W. J. Smith, et al, Immunogenicity of PCV24, an expanded pneumococcal conjugate vaccine, in adult monkeys and protection in mice. *Vaccine,* 2021, **39**, 4231-4237.

23. A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, et al, Symbol nomenclature for graphical representations of glycans. *Glycobiology,* 2015, **25**, 1323-1324.

24. B. Tefsen, A. F. Ram, I. van Die and F. H. Routier, Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology,* 2012, **22**, 456-469.

25. C. Marino, A. Rinflerch and R. M. de Lederkremer, *Galactofuranose antigens, a target for diagnosis of fungal infections in humans*, FSO199, Future Science, 2017.

26. C. R. Isabella, *Carbohydrate and bacterial binding specificity of human intelectin-1*, Massachusetts Institute of Technology, 2021.

27. L. L. Kiessling, Chemistry-driven glycoscience. *Bioorg. Med. Chem.*, 2018, **26**, 5229-5238.

28. D. A. Wesener, K. Wangkanont, R. McBride, X. Song, M. B. Kraft, et al, Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.*, 2015, **22**, 603-610.

29. L. Chen, J. Li and G. Yang, A comparative review of intelectins. *Scand. J. Immunol.*, 2020, **92**, e12882.

30. B. L. Spencer, J. S. Saad, A. T. Shenoy, C. J. Orihuela and M. H. Nahm, Position of O-acetylation within the capsular repeat unit impacts the biological properties of pneumococcal serotypes 33A and 33F. *Infect. Immun.*, 2017, **85**, 132.

31. A. H. Aytenfisu, R. Simon and A. D. MacKerell Jr, Impact of branching on the conformational heterogeneity of the lipopolysaccharide from Klebsiella pneumoniae: Implications for vaccine design. *Carbohydr. Res.*, 2019, **475**, 39-47.

32. Y. Kang, S. Barbirz, R. Lipowsky and M. Santer, Conformational diversity of O-antigen polysaccharides of the Gram-negative bacterium Shigella flexneri serotype Y. *J. Phys. Chem. B*, 2014, **118**, 2523-2534.

33. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

34. A. Sarkar, C. Fontana, A. Imberty, S. Perez and G. Widmalm, Conformational preferences of the O-antigen polysaccharides of Escherichia coli O5ac and O5ab using NMR spectroscopy and molecular modeling. *Biomacromolecules*, 2013, **14**, 2215-2224.

35. M. Yang, R. Simon and A. D. MacKerell Jr, Conformational preference of serogroup B Salmonella O polysaccharide in presence and absence of the monoclonal antibody Se155–4. *J. Phys. Chem. B*, 2017, **121**, 3412-3423.

36. M. Zaccheus, R. Pendrill, T. A. Jackson, A. Wang, F. Auzanneau and G. Widmalm, Conformational dynamics of a central trisaccharide fragment of the LeaLex tumor associated antigen studied by NMR spectroscopy and molecular dynamics simulations. *Eur. J. Org.*, 2012, **2012**, 4705-4715.

37. M. M. Kuttel, Comparative Molecular Modelling of Capsular Polysaccharide Conformations in Streptococcus suis Serotypes 1, 2, 1/2 and 14 Identifies Common Epitopes for Antibody Binding. *Front. Mol. Biosci.*, 2022, **9.**

38. N. I. Richardson, M. M. Kuttel, F. St Michael, C. Cairns, A. D. Cox and N. Ravenscroft, Cross-reactivity of Haemophilus influenzae type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies. *Glycoconj. J.*, 2021, **38**, 735–746.

39. J. Hlozek, S. Owen, N. Ravenscroft and M. M. Kuttel, Molecular Modeling of the Shigella flexneri Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a. *Vaccines*, 2020, **8**, 643.

40. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines*, 2021, **9**, 432.

41. E. A. Kabat, The nature of an antigenic determinant. *J. Immunol.,* 1966, **97**, 1-11.

42. M. Kuttel, Y. Mao, G. Widmalm and M. Lundborg, *in Proceedings of 2011 IEEE Seventh International Conference on eScience,* ed. Anonymous, 2011.

43. M. M. Kuttel, J. Ståhle and G. Widmalm, CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.,* 2016, **37**, 2098-2105.

44. W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.,* 1996, **14**, 33-38.

45. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, et al, Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.,* 2008, **29**, 2543-2564.

46. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Comput. Chem.,* 2009, **5**, 2353-2370.

47. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.,* 1983, **79**, 926-935.

48. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al, Scalable molecular dynamics with NAMD. *J. Comput. Chem.,* 2005, **26**, 1781-1802.

49. J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco and K. Schulten, Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.,* 2007, **28**, 2618-2640.

50. T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.,* 1993, **98**, 10089-10092.

51. W. F. Van Gunsteren and H. J. Berendsen, A leap-frog algorithm for stochastic dynamics. *Mol. Simul.,* 1988, **1**, 173-185.

52. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.,* 1995, **103**, 4613-4621.

53. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A.,* 1985, **31**, 1695.

54. S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems. *Mol. Phys.,* 1983, **50**, 1055-1076.

55. A. Grossfield and D. M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.,* 2009, **5**, 23-48.

56. S. Cross, M. M. Kuttel, J. E. Stone and J. E. Gain, Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.,* 2009, **28**, 131-139.

57. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.,* 2007, **9**, 90-95.

58. Gracia Luis, *WMC PhysBio Clustering*, Weill Cornell Medical College, 2012.

59. M. H. Nahm, J. Yu, J. Vlach and M. Bar-Peled, A common food glycan, pectin, shares an antigen with Streptococcus pneumoniae capsule. *Am. Soc. Microbiol.*, 2020, **5**, 74.

60. M. M. Kuttel and N. Ravenscroft, Conformation and cross-protection in Group B Streptococcus serotype III and Streptococcus pneumoniae serotype 14: a molecular modeling study. *Pharmaceuticals*, 2019, **12**, 28.

61. S. Zhang, M. Sella, J. Sianturi, P. Priegue, D. Shen and P. H. Seeberger, Discovery of oligosaccharide antigens for semi-synthetic glycoconjugate vaccine leads against Streptococcus suis serotypes 2, 3, 9 and 14. *Angew. Chem. Int. Ed.*, 2021, **60**, 14679-14692.

62. M. M. Kuttel, Z. Timol and N. Ravenscroft, Cross-protection in Neisseria meningitidis serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr. Res.*, 2017, **446-447**, 40-47.

63. J. Yu, J. Lin, W. H. Benjamin Jr, K. B. Waites, C. Lee and M. H. Nahm, Rapid multiplex assay for serotyping pneumococci with monoclonal and polyclonal antibodies. *J. Clin. Microbiol.*, 2005, **43**, 156-162.

64. C. A. Bush, J. Yang, B. Yu and J. O. Cisar, Chemical structures of Streptococcus pneumoniae capsular polysaccharide type 39 (CPS39), CPS47F, and CPS34 characterized by nuclear magnetic resonance spectroscopy and their relation to CPS10A. *J. Bacteriol.*, 2014, **196**, 3271-3278.

65. B. O. Petersen, S. Meier, B. S. Paulsen, A. R. Redondo and I. C. Skovsted, Determination of native capsular polysaccharide structures of Streptococcus pneumoniae serotypes 39, 42, and 47F and comparison to genetically or serologically related strains. *Carbohydr. Res.*, 2014, **395**, 38-46.

6 Chapter 6

6.1 Summary and Conclusions

In this work we have applied molecular modeling methodology to compare structurally similar antigens across four different case studies using the results to propose potential cross reactivity and resultant cross protection to inform future vaccine design. We further attempted to rationalize observed immunological phenomena and provide insights into important epitopes and mechanisms for antigen-antibody binding.

Understanding the conformational aspects of these target vaccine antigens is important for informing efficient and low-cost vaccine design as many of the insights gained from a conformational analysis are not available through other experimental or theoretical analyses. [1]

The first publication included in this thesis examined the capsular polysaccharides (CPSs) of *Haemophilus influenzae* types b and a. [2] These data showed high levels of flexibility (due to a linear sugar and phosphodiester bond) and predicted a lack of cross-protection between the two serotypes supporting the need for a bivalent vaccine to protect against both serotypes. Further study of a more stable methyl protected, synthetic, type b antigen showed conformational and surface differences compared to the type b antigen. On the basis of immunological studies, however, the synthetic antigen would still be expected to provide protection and could improve vaccine formulations. [3,4]

The second publication systematically compared the CPSs of *Pasteurella multocida* types B and E as well as the structurally similar *H. influenzae* types e and d. [5] These CPSs were conformationally well defined and there were significant conformational differences between the CPSs including different helical handedness resulting from different backbone linkage positions. Taking a closer look at potential binding epitopes we further determined differences between the CPSs. Comparing the backbone (CPS without side chains) epitopes to the type B and E epitopes, we noted the lack of common epitopes as well as the use of

side groups in disrupting potential binding epitopes. The backbones are amino-sugar rich and highly immunogenic; side groups effectively shield the backbone and add steric hindrance to the backbone epitopes. Low levels of cross reactivity may be explained by the conserved monosaccharide composition and defined conformations. However, as suggested by available immunological studies, effective cross reactivity and cross protection is not expected and a bivalent vaccine would be required for effective disease protection. [6-9]

The third publication considers the lipopolysaccharide (LPS) O-antigen of *Shigella flexneri* serotype 6 in relation to the common serotype Y backbone. [10] These antigens exhibit conformational flexibility; however, structural differences result in different conformations: notably a difference in helical handedness from a difference in backbone linkage position (as for *Pasteurella multocida*). Furthermore, the presence of non-stoichiometric O-acetylation did not appear to affect conformation, suggesting this modification is not critical for immunogenicity. Verification in better models (preferably from human clinical data) is required as current animal models are not representative. The conformational differences between serotype 6 and other *S. flexneri* serotypes indicated the need to include serotype 6 as part of an effective vaccine as proposed. [11]

The fourth publication included in this thesis investigated complex asymmetric cross-reactivity seen in the *Streptococcus pneumoniae* serogroup 10 CPSs. [12] High flexibility was again observed (as previously due to a linear sugar with a phosphodiester bond) and further conformational analysis of smaller epitopes was able to rationalize the observed cross-reactivity. [13] The repeating units (RUs) included a non-mammalian galactofuranose sugar - to which the human immune system has a natural defense. [14-16] Overlaying the dominant antigen conformations with the crystal structure of the bound immune protein that recognizes galactofuranose, (human intelectin-1), explained the observed serotype epidemiology. By varying the positioning of the galactofuranose residue to make binding more difficult, some serotypes are better able to evade immune detection. Further conformational study of the galactofuranose epitope rationalized the observed complex asymmetric cross-reactivity in the serogroup. These results support the inclusion of

serotype 10A in a vaccine as it would be expected to provide adequate cross-protection against serogroup 10 disease.

As a cohesive body of work, these case studies expand the applicability of the methodology to new antigens and new pathogens. Furthermore, examining high-level patterns provides insights into different mechanisms utilized by pathogens to evade the immune system and evidences the usefulness of this methodology in informing vaccine design. While each individual problem is substantial, the overall culmination of these findings provides even greater insights; proposing overarching mechanistic rationalization for structural and conformational features and attempting to connect structure, conformation, and function.

In this work we encountered several different strategies based in structure and conformation for immune evasion. For one, high conformational flexibility creates a moving target that is harder for the immune system to recognize. Another is implementing small structural changes (such as a linkage difference) which result in significant conformational differences and provides a low metabolic "cost", efficient means of producing a completely different molecular surface (such as a change in helical handedness). Similarly, shielding of immunostimulatory components/epitopes by disguising them with non-immunogenic substituents or hindering binding through steric bulk or incompatible conformations.

We encountered high flexibility of antigens in *H. influenzae* (Chapter 2) and *S. pneumoniae* (Chapter 5) making the conformational analysis of the whole molecules relatively uninformative. *H. influenzae* type a and b have short (disaccharide), linear RUs while *S. pneumoniae* serogroup 10 have longer (hexa/heptasaccharide) RUs and include side groups. In both cases, a phosphodiester linkage and a linear sugar (ribitol) are present; this provides more flexibility than that of a normal glycosidic linkage and cyclic monosaccharide as there are more degrees of freedom in the linear sugar and greater rotational freedom in the phosphodiester linkage. The observed flexibility was more pronounced in *H. influenzae* than *S.*

pneumoniae and can be linked to structure as the steric strain of the side groups and longer RU in *S. pneumoniae* would be expected to reduce flexibility compared to the less hindered *H. influenzae*.

We suggest that the flexibility confers an advantage to these pathogens as the lack of conformational definition makes the antigen harder to recognize and more difficult for antibodies to bind. Despite this hypothesis, we note the significant success of the *H. influenzae* type b vaccine and in Chapter 2 suggest immunodominance of the charged phosphate group (which is highly exposed for binding) as the mechanism of immune recognition. In *S. pneumoniae* the use of the non-mammalian galactofuranose sugar improves pathogen virulence but is also a target for existing immune clearance mechanisms. In this case, the high flexibility appears to help hinder binding of the galactofuranose residue; only certain molecular conformations allow for galactofuranose epitope binding by innate immune proteins without significant steric hindrance from the rest of the polysaccharide chain.

When flexibility was less prevalent and conformation more defined, as in *P. multocida* (Chapter 3) different strategies were employed. *P. multocida* type B and E have short (trisaccharide), branched RUs with immunogenic amino-sugar rich backbones. Of the antigens studied in this work, *P. multocida* had the greatest conformational definition and – with an immunogenic backbone – could be considered an easy target for immune clearance. The pathogen, however, adds non-immunogenic side group residues which appear to both add steric bulk, and disguise the backbone potential binding epitopes. Interestingly these fructofuranose side groups are extremely labile, removed by very mild acid hydrolysis, and may allow environmental modulation of the antigen.

Despite shared monosaccharide composition, the *P. multocida* type B and E pathogens exhibited different conformations primarily due to a difference in helical handedness completely changing the molecular surface and potential binding epitopes. This difference was brought about by a single linkage position difference in the backbone and is an example of how a small structural difference can significantly affect conformation and in turn prevent cross reactivity by creating two conformationally distinct antigens.

Similarly, in *S. flexneri* (Chapter 4), which has a linear, tetrasaccharide RU, a shift in helical handedness due to a single linkage position difference was a key conformational difference between the serotype 6 and serotype Y backbones. The change in helical shape and handedness brought about by structural changes effectively change the epitopes and prevents cross reactivity and protection.

A key theme seen throughout these case studies was the importance of potential binding epitopes in rationalization of the observed cross-reactivity (or lack thereof). This was especially valuable in the context of high molecular flexibility and low conformational definition when other qualitative analyses were unhelpful as for *S. pneumoniae* serogroup 10 (Chapter 5). Despite a lack of conformational definition on a molecular level, observed cross-reactivity could be rationalized by looking at smaller epitopes and identifying common potential binding epitopes. In contrast, the lack of significant common potential binding epitopes in the flexible *H. influenzae* type a and b (Chapter 2) as well as *S. flexneri* serotype 6 (Chapter 4) correlates with a lack of cross-reactivity. In *P. multocida* type B and E (Chapter 3) – despite having greater conformational definition on a molecular level – a lack of shared epitopes correlates with (and further supports) a lack of observed cross-reactivity.

This work adds valuable case studies and applications of the molecular modeling methodology to the conformational analysis of vaccine antigens that have not yet been modeled. Furthermore, it expands on the established methodology and by linking in-silico modeling with immunological observations and results it contributes insights into the functional nature of the conformational and structural features observed. Additional links are drawn between the individual case studies to highlight motifs commonly employed by pathogens.

The continued application of the methodology to different vaccine antigen targets will continue to grow the knowledge in the field and potentially bring to light additional features exploited by pathogens in evading the immune system. Further validation of work though immunological studies, especially in the case of *P. multocida,* can further support the interpretations presented here and help to validate the use of molecular

modeling as a complementary technique informing the rational design of vaccines. We further hope that the constant increase in computational power will assist in reducing computational cost and time allowing the modeling of systems with greater ease and making this methodology more accessible.

6.2 References

1. M. M. Kuttel and N. Ravenscroft, in *Carbohydrate-Based Vaccines: From Concept to Clinic*, ed. A. Prasad Krishna, ACS Publications, 2018, ch. 7, pp. 139-173.

2. N. I. Richardson, M. M. Kuttel, F. St Michael, C. Cairns, A. D. Cox and N. Ravenscroft, Cross-reactivity of Haemophilus influenzae type a and b polysaccharides: molecular modeling and conjugate immunogenicity studies. *Glycoconj. J.*, 2021, **38**, 735–746.

3. M. Emmadi, M. P. Lisboa, B. Monnanda, et al., A Liquid Stable Biological Active Semi-Synthetic Glycoconjugate Vaccine against Haemophilus influenzae type b, unpublished work.

4. US Pat., US20190153015A1, 2019.

5. N. I. Richardson, N. Ravenscroft and M. M. Kuttel, Conformational comparisons of *Pasteurella multocida* type B and E and structurally related capsular polysaccharides, *Glycobiology,* 2023, https://doi.org/10.1093/glycob/cwad049.

6. F. St Michael, C. M. Cairns, P. Fleming, E. V. Vinogradov, J. D. Boyce, et al, The capsular polysaccharides of Pasteurella multocida serotypes B and E: Structural, genetic and serological comparisons. *Glycobiology,* 2021, **31**, 307-314.

7. C. W. Penn and L. K. Nagy, Isolation of a protective, non-toxic capsular antigen from Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1976, **20**, 90-96.

8. C. W. Penn and L. K. Nagy, Capsular and somatic antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1974, **16**, 251-259.

9. L. K. Nagy and C. W. Penn, Protection of cattle against experimental haemorrhagic septicaemia by the capsular antigens of Pasteurella multocida, types B and E. *Res. Vet. Sci.,* 1976, **20**, 249-253.

10. N. I. Richardson, N. Ravenscroft, V. Arato, D. Oldrini, F. Micoli and M. M. Kuttel, Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines,* 2021, **9**, 432.

11. K. L. Kotloff, M. S. Riddle, J. A. Platts-Mills, P. Pavlinac and A. K. M. Zaidi, Shigellosis. *Lancet,* 2018, **391**, 801-812.

12. N. I. Richardson, M. M. Kuttel and N. Ravenscroft, Modeling of pneumococcal serogroup 10 capsular polysaccharide molecular conformations provides insight into epitopes and observed cross-reactivity. *Front. Mol. Biosci.,* 2022, **9**, 961532.

13. J. Henrichsen, Six newly recognized types of Streptococcus pneumoniae. *J. Clin. Microbiol.,* 1995, **33**, 2759-2762.

14. L. L. Kiessling, Chemistry-driven glycoscience. *Bioorg. Med. Chem.,* 2018, **26**, 5229-5238.

15. C. R. Isabella, *Carbohydrate and bacterial binding specificity of human intelectin-1*, Massachusetts Institute of Technology, 2021.

16. D. A. Wesener, K. Wangkanont, R. McBride, X. Song, M. B. Kraft, et al, Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.*, 2015, **22**, 603-610.