Molecular Modelling of the *Streptococcus pneumoniae* serogroup 6 capsular polysaccharide antigens.

Neann Mathai

Department of Computer Science

University of Cape Town

A thesis submitted for the degree of

MSc. in Computer Science

June 2013

Plaigirism Declaration

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

Neann Mathai
June 2013

To my loving family.

Acknowledgements

I would like to thank my supervisors for all of their help and guidance. Assoc/Prof. Michelle Kuttel, thank-you for the constant support, enthusiasm, thorough proof-reading, guidance and so much more. I am very grateful for the impact you have had on my life. Assoc/Prof Neil Ravenscroft, thank-you for your guidance, all last minute editing, encouragement and never ending enthusiasm.

To my late mother, Susan Mathai, thank-you. You were my first teacher and my biggest inspiration. Everything I am is because of you. To my father, Promod Mani and sister Nicole Mathai, thank-you for supporting my move back to Cape Town and this Masters.

To the friends who have been a part of my life during this masters: thanks for all the love, friendship, prayers, concerned phone calls, singing, dancing, food packages, campus lunches and coffees, drinks, discussions, bananagrams, Skype calls and lab camaraderie. Your generosity despite my grumpy demeanor over the last few months has taught me the meaning of true friendship. I would not have made it without you! To those of you living in Cape Town, you made this place home; it is very hard to say goodbye and I hope to be back here with you all soon. To my GosComm family past and present, your support is invaluable.

Thanks to Andrew Lewis at ICTS-HPC for his generosity and help with the HPC clusters on which some simulations were run. To the University of Cape Town and the Department of Computer Science, thank-you for this post-graduate opportunity. This project would not have been possible without financial support for the last 2.5 years: Michelle Kuttel, your constant generosity (UCT-NRF grant-holder's fund, conference travel funding, and research funding) has made this possible. I would also like to acknowledge the Department of Science and Technology for the TATA Women in Science fellowship.

Finally, to my God, thank-you for your constant provision, protection and the promise of hope. "Whatever you do, work at it with all your heart, as working for the Lord .. " - I tried :) .

Thank-you.

Abstract

In this thesis, a systematic study of the structural characterization of the capsular polysaccharides of *Streptococcus pneumoniae* is conducted using Molecular Modelling methods. *S.pneumoniae* causes invasive pneumococcal disease (IPD), a leading cause of death in children under five.

The serotypes in group 6 are amongst the most common of IPD causing serotypes. We performed structural characterization of serogroup 6 to understand the structural relationships between serotypes 6A, 6B, 6C and 6D in an attempt to understand the cross protection seen within the group. The 6B saccharide has been included in the early conjugate vaccine (PCV-7), and has shown to elicit protection against the 6B as well as offer some cross-protection against 6A. 6A has since been included in the latter conjugate vaccines in the hopes of eliciting stronger protection against 6A and 6C.

Molecular Dynamics simulations were used to investigate the conformations of oligosaccharides with the aim of elucidating a conformational rationale for why small changes in the carbohydrate primary structure result in variable efficacy. We began by examining the Potential of Mean Force (PMF) plots of the disaccharide subunits which make up the Serogroup 6 oligosaccharides. The PMFs showed the free energy profiles along the torsional angles space of the disaccharides. This conformational information was then used to build the four oligosaccharides on which simulations were conducted. These simulations showed that serotype pairs 6A/6C and 6B/6D have similar structures.

Conference Presentations

Parts of this work have been presented at poster sessions at the following conferences:

Mathai, N. S.; Kuttel M.; Ravenscroft N. 2012. *Conformational Studies of Serogroup 6 Pneumococcal Antigens*, 26th International Carbohydrate Symposium, Madrid, Spain.

Mathai, N. S.; Kuttel M.; Ravenscroft N. 2012. *Optimizing multivalent conjugate vaccine composition: A systematic approach to conformational analysis of carbohydrate antigens*, 244th American Chemical Society National Meeting: Materials for Health and Medicine, Philadelphia, USA.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Invasive pneumococcal disease (IPD) is a leading cause of death in children under five, killing around a million children globally each year.[2–6] IPD is the clinical condition where *Streptococcus pneumoniae* infects normally sterile sites such as blood, cerebrospinal fluid or pleural fluid. In addition to pneumonia, *S.pneumoniae* also causes other diseases, such as: bacteremia, meningitis, sepsis, otitis media and other mucosal and invasive diseases.[1,4,7,8]

S. pneumoniae is a gram-positive bacterium surrounded by a protective polysaccharide capsule. Polysaccharides are carbohydrates comprised of long chains of linked sugar monomers, which can join together in a variety of possible combinations. There are also a number of linkages possible between two monosaccharides and structural variation is further increased by the fact that these linkages have great flexibility, which allows a carbohydrate to adopt a number of conformations. It is suspected that the flexibility of the polysaccharide capsule accounts for the powerful virulence of the bacteria, by delaying the host's immune response.[9–12]

On the basis of the structure of the capsular polysaccharides, *S.pneumoniae* is classified into over 90 different serotypes, a handful of which cause most of the IPDs in the world.[2,4,13] In addition to differentiation, the capsular polysaccharide acts as a virulence factor, as it inhibits the phagocytosis of the bacteria by the host's immune system.

Over the years, vaccine development against *S.pneumoniae* has advanced and the latest conjugate vaccines have been shown to be effective in preventing disease in high risk population groups. Conjugate vaccines comprise of bacterial

capsular polysaccharide antigens covalently linked to a carrier protein to enhance the immunogenicity of the vaccine. While conjugate vaccines are effective, they do not protect against all the serotypes causing IPDs. Nearly all the fatalities caused by IPDs, occur in developing countries in Asia and Africa.[7]

IPD in South Africa

Consideration of the IPD burden is especially important in the South African context. *S.pneumoniae* is commonly present in the South African population, with bacteria present in six in every ten healthy adults, with immune systems able to fight off the infection. However, the risk of IPD is greater in those who are immunocompromised.[14–16] South Africa has high HIV rates, which has in turn increased the IPD burden in the country. This is specifically true within the cohort of HIV infected children under five, which has contributed to the steadily increasing under-five mortality rate in South Africa.[17]

The incidence of documented IPD in South African children has doubled in the years following the onset of the HIV epidemic.[17] 75% of severe IPD in South African children occurs within 6% of the population of HIV positive children under five.[16] A trial of a 9-valent conjugate vaccine in South Africa showed a 83% reduction in HIV-negative children and a 65% reduction in HIV-positive children of vaccine serotype specific IPDs.[16,17]

South Africa has prioritized pneumococcal prevention, despite the high cost of vaccines, as prevention of this disease is more cost effective than treatment. South Africa was among the first of the industrializing countries and the first country in Africa to introduce the pneumococcal vaccine into the public immunization programme.[14,17–19]

Understanding Vaccines

One approach to improving conjugate vaccines is to increase the valency and thus provide better coverage.[20,21] This requires knowledge of which polysaccharides might elicit an antibody response. Unfortunately, the relationship between carbohydrate sequence, antigenicity and immunogenicity is still not well understood. An understanding of the three-dimensional structure of a polysaccharide will help

explain their properties and functions and thus expand the range of serotypes covered by a conjugate vaccine.[22] However, the extreme flexibility of carbohydrates makes it difficult to determine their conformation experimentally. Traditional experimental structural analysis methods, such as X-ray crystallography and nuclear magnetic resonance (NMR), can be used to determine saccharide structure however, they are not without their shortcomings. Oligo- and polysaccharides often do not have a single conformation, but occupy an ensemble of conformations in solution. It is this dynamic property of carbohydrates which makes them difficult to characterize experimentally. X-ray crystallography provides a static crystal structure and NMR data can provide structural information of carbohydrates in solution. However, this data is often an average of all the populated conformation states and often requires Molecular Modelling studies for confirmation.[23–25] Molecular dynamics (MD) simulations of molecules are conducted with well-defined empirical force fields which aim to realistically describe the system. The results of these simulations shed light on the conformation adopted by these molecules, which can then be used to explain the behaviour and functions of the molecules.

1.1 Problem Statement - Serogroup 6

The Pneumococcal Vaccine Project (PATH) supports the development of affordable pneumococcal vaccines and vaccine manufacturers in the developing world have to carefully choose the serotypes included in their vaccine to ensure a coverage of at least 60% of the invasive disease isolates in the target region[26] and yet have sufficiently low valency to be affordable. The 7-valent pneumococcal vaccine (PCV-7) has been one of the biggest success stories of conjugate vaccines and is available in South Africa. PCV-7 has dramatically reduced the incidence of IPD, however its efficacy varies globally. It is most successful in the United States and least in the developing world, where the majority of childhood deaths occur.[7,27] There has also been an increase in IPD cases due to non-vaccine serotypes (serotype replacement) which has motivated the licensure of higher valent vaccines PCV-10 and PCV-13 in order to increase vaccine coverage.

Serogroup 6 serotypes are amongst the most common cause of IPD in children

and is the focus of this work.[6] Serogroup 6 was originally comprised of type 6A and 6B. Serotypes 6C and 6D were originally mistyped 6A and 6B respectively and have recently been added to the group.[28,29] Two points of difference account for these four serotypes: a galactose or a glucose residue, and a (1→3) or a (1→4) rhamnose-ribitol5P linkage:

6A: [→2)αD**Gal***p*(1→3)αDGlc*p*(1→3)αLRha*p***(1→3)**DRibol5P(O→]
6B: [→2)αD**Gal***p*(1→3)αDGlc*p*(1→3)αLRha*p***(1→4)**DRibol5P(O→]
6C: [→2)αD**Glc***p*(1→3)αDGlc*p*(1→3)αLRha*p***(1→3)**DRibol5P(O→]
6D: [→2)αD**Glc***p*(1→3)αDGlc*p*(1→3)αLRha*p***(1→4)**DRibol5P(O→]

Serotype 6B is more stable than 6A and is included in the early conjugate vaccines (PCV-7 and PCV-10).[7,30] Serotype 6B has been shown to not only protect against IPD caused by 6B, but also offers some cross-protection against 6A.[13,31–33] Serotype 6C was originally mistyped as 6A and is now found to cause IPDs worldwide. The 6B antigen does not appear to offer any cross-protection against 6C, which is currently the most prevalent serotype within group 6 to cause IPD.[33–36] Serotype 6D, was initially synthesized in the lab in 2009, and not considered to be present naturally.[37] However IPD caused by naturally occurring 6D has since been reported.[6,28,32,38,39] The later conjugate vaccines (PCV-13 and PCV-15) also include serotype 6A, and is suspected to induce protection against 6C and 6D but, this is yet to be clinically established.[13,29]

Currently there is no conformational rationale for the observed cross-protection (the limited cross protection 6B offers 6A) and the predicted cross-protection (6A and 6B cross-protecting against 6C and 6D) for serogroup 6.

1.2 Aim

The aim of this project is to develop a computational methodology to structurally characterize the serogroup 6 serotypes, so as to investigate a structural explanation for the cross-protectivity seen within the serogroup 6.

1.3 Research Questions

The aim is to answer the following questions:

1. **What effect does the αLRha*p*-DRibol-5P (1→3)/(1→4) linkage change have on the structure of the serotypes?** Serotype 6B, included in conjugate vaccines PCV-7 and PCV-10, has shown to offer some cross-protection against 6A.[13,31,32,40] The only difference between 6A and 6B is a (1→3)/(1→4) linkage. Can the structures explain this cross-protection?

2. **What effect does a galactose/glucose residue change have on the structure of the serotypes?** With the introduction of PCV-7 into immunization schedules, there has been a global reduction in IPD incidences brought on by 6A and 6B, however serotype replacement has taken place and there has been an increase in 6C and 6D incidences. Serotypes 6C and 6D were previously identified as 6A and 6B and differ in a galactose/glucose residue. Serotype 6A has thus been included in later conjugate vaccines PCV-13 and PCV-15 in the hopes that the combination of 6A and 6B in the vaccine will protect against the entire serogroup. Preliminary studies have suggested that 6A offers 6C some cross-protection.[13,29,41,42] Can the structures shed light on this possible cross-protection?

3. **Can a systematic approach to computational modelling of serogroup 6 provide insight into cross-protection observed between serotypes 6A/B/C/D?** *S.pneumoniae* serogroup 6 are phosphodiester containing saccharides with a linear ribitol sugar. This moiety increases the complexity of the serogroup. Does the methodology used in this dissertation shed light on the structures of this serogroup?

1.4 Approach

A systematic approach is used to study the oligosaccharide structures of serogroup 6 using Molecular Mechanics and Dynamics simulations.

The first step is to investigate the glycosidic linkage conformation by analyzing the disaccharide components of serogroup 6. Ramachandran-like contour maps

will be produced for these linkages (based on their ϕ and ψ torsion angles) which detail the entire linkage conformational space and the disaccharide preferences. A metadynamics routine, using ϕ and ψ dihedral angles as collective variables, will be used to produce these potential of mean force landscapes giving insight into the low energy conformations of the disaccharides. The optimum disaccharide conformations are thus established and are used to build likely conformations of the oligosaccharides on which Molecular Dynamics simulations are conducted.

After analyzing each of the disaccharide components, 3 repeating unit long oligosaccharides will be built with their likely conformations. Molecular Dynamics simulations in vacuum of these oligosaccharides will be run to observe convergence. The dihedral angles of the middle repeating unit will be plotted on their previously calculated contour maps showing how interresidue interactions affect the preferred conformations of the glycosidic linkages.

The converged conformations of the middle repeating unit of the serotype MD simulations will be used to build 5-repeating unit long oligosaccharides which will be minimized. These structures will allow for preliminary structural comparisons of serotypes allowing for some indication of the effects of residue and linkage differences on the conformations and hence possible sources of cross-protection. However due to time constraints, no solution or ion effects can be established at this point.

1.5 Thesis Overview

This thesis is organized as follows:

Chapter 2 is an introduction to carbohydrates and conjugate vaccines. It begins with an outline of the various structural aspects of carbohydrates which allow them to perform a variety of functions. This is followed by the history of conjugate vaccines specifically against *S.pneumoniae* serogroup 6, the focus of this work.

Chapter 3 then follows, outlining the computational theory and methods used in Molecular Modelling, with attention given to the methods used in this work.

Chapter 4 details the methodology and specifications used in modelling both the disaccharide subunits of serogroup 6 and the serogroup 6 oligosaccharides.

Chapter 5 analyzes the disaccharide subunits of serogroup 6. We start by looking at the effects the highly flexible and charged ribitol-phosphate disaccharides. This is followed by an analysis of the remaining disaccharide structures, comparing points of difference within the serogroup.

Chapter 6 studies the serogroup 6 oligosaccharide structures from Molecular Dynamics simulations.

Lastly, Chapter 7 concludes this dissertation and provides possible areas of further study.

Chapter 2

Carbohydrate Structure and Conjugate Vaccines

Carbohydrates, also known as saccharides, are an important class of bio-molecules, making up most of the organic matter on Earth. The functions of saccharides extend beyond primary energy stores and fuels to: metabolic intermediates, structural framework for DNA and RNA, and structural support molecules, for both plant cell walls and animal exoskeletons. Carbohydrates are also found on the surfaces of cells and play a key role in cell-cell recognition, and thus immunological responses. The wide spectrum of functions that carbohydrates perform is due to the large variety of structures and functionalities that they can adopt. The diversity of carbohydrates can be seen from the smallest class, the monosaccharides, to the large complex polysaccharide molecules.[24,43–48]

2.1 Carbohydrate Structure

2.1.1 Monosaccharides

The simplest carbohydrates, known as monosaccharides, have the empirical formula $C_n(H_2O)_n$, where n is greater than 2. In a monosaccharide, one carbon atom bears a carbonyl group (C=O), while the other carbon atoms bear hydroxyl groups (O-H). The position of the carbonyl group classifies a monosaccharide as an aldose or ketose. If the sugar contains an aldehyde group (RC(=O)H), it is

known as an aldose, while, if it contains a keto group (RC(=O)R'), it is a ketose. This difference can be seen in Figure 2.1 below which shows two hexoses, the aldohexose glucose and the ketohexose fructose.

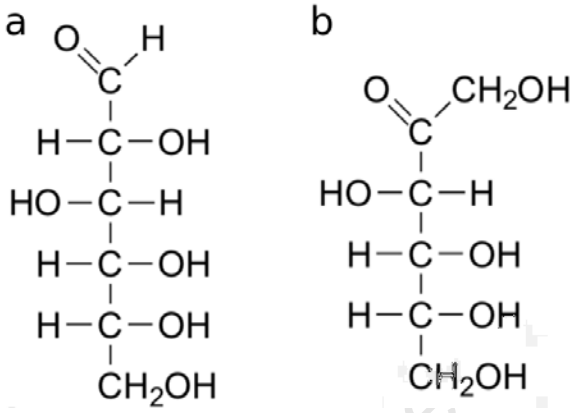


Figure 2.1: Aldose glucose (a) and ketose fructose (b)

An aldose with n carbon atoms contains n-2 chiral centres and as a result has 2^{n-2} stereoisomers. While, a ketose, with n carbon atoms, has n-3 chiral centres and therefore possess 2^{n-3} stereoisomers. These stereoisomers are further classified based on the position of the hydroxyl group at the C_{n-1}: L, when left, and D when right. Most naturally occurring sugars are found as D stereoisomers. Figure 2.2 shows the D and L stereoisomers of glucose.

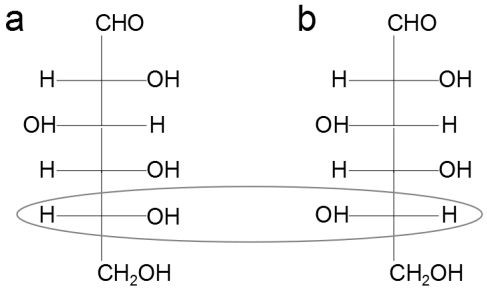


Figure 2.2: D-glucose (a) and L-glucose (b)

Hydroxyls react with an aldol/ketone group to form hemiacetals/hemiketals respectively. These reactions can occur through a neutral mechanism or an acid catalyzed mechanism as seen in Figure 2.3.

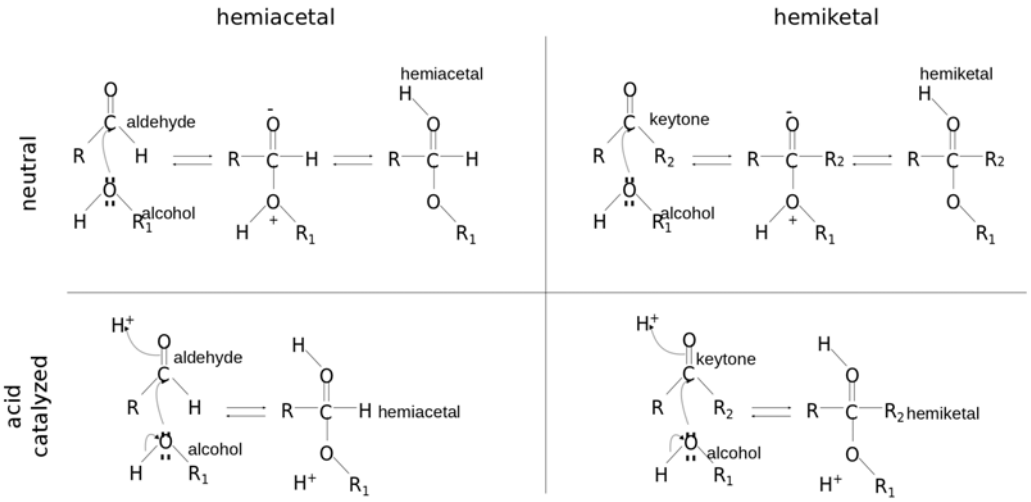


Figure 2.3: Neutral and acid catalyzed mechanisms of hemiacetal and hemiketal formation

Monosaccharides often do not exist as open chains, but they often from ring hemiacetals/hemiketals known as lactols when the aldol/ketone group react with a hydroxyl group via the neutral mechanism shown in Figure 2.3. Most often these ring structures are six-membered rings (pyranoses) or five-membered rings (furanoses).

The formation of the ring structures gives rise to another possibility for variety in saccharides. The closing of a saccharide produces two possibilities, an α and a β anomer, as seen in Figure 2.4. Typically, hexose sugars in solution exist in a dynamic equilibrium between the α and β anomers.

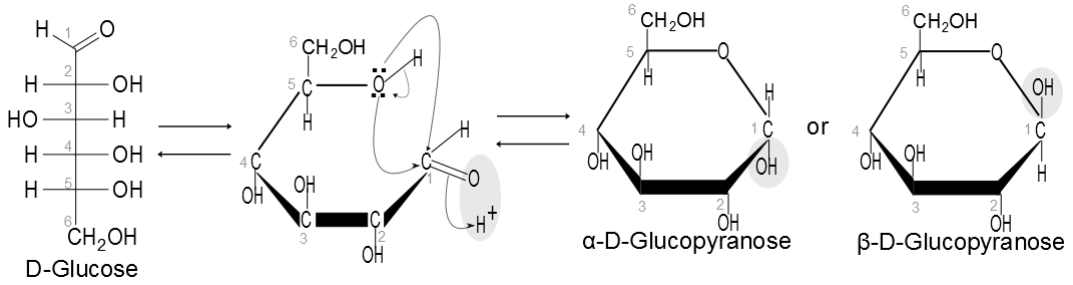


Figure 2.4: Intramolecular cyclization of glucose

Once cyclised these ring structures are not flat but rather take on a variety of

configurations known as puckers, due to the tetrahedral geometry of the carbon atoms. For pyranose rings, chair conformations are generally preferred, while the configurations such as the boat and half-chair cause steric crowding of atoms, leading to higher energies. The chair conformations are classified as either $^4C_1$, where C4 lies above and C1 lies below the plane of the ring, and $^1C_4$, with C1 and C4 lying above and below the plane respectively. Pyranose saccharide rings are said to be relatively rigid.[22,43,46,49]

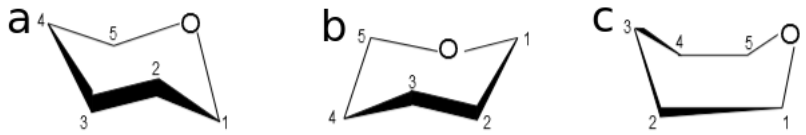


Figure 2.5: $^4C_1$ chair (a), $^1C_4$ chair (b) and boat (c) conformations of pyranose rings

2.1.2 Gycosidic Linkages and Disaccharides

Two monosaccharides may be joined together by a glycosidic linkage to form a disaccharide. A condensation reaction occurs, joining the anomeric carbon of one monosaccharide to any of the hydroxyl groups on the second monosaccharide. If the glycosidic linkage occurs between the two anomeric carbons of the monosaccharides, there is no free lactol group and the disaccharide is known as non-reducing, as it can no longer form further linkages with other saccharides. If, on the other hand, the glycosidic linkage is from the anomeric carbon to any of the non-anomeric hydroxyls, there will be a free lactol group and the disaccharide is known as a reducing disaccharide. Reducing disaccharides are free to increase in length by forming glycosidic linkages to further residues. Examples of reducing and non-reducing disaccharides can be seen in Figure 2.6.

The conformation of disaccharides are primarily determined by their glycosidic linkage. The glycosidic bond can be characterized by torsion angles ϕ and ψ and in the case of a (1→6) linkage, an additional torsional angle ω. The definitions of ϕ, ψ and ω are shown below (2.1). The ϕ and ψ torsion angle definitions are different to the standard IUPAC definitions, but are analogues to IUPAC's ϕ_H

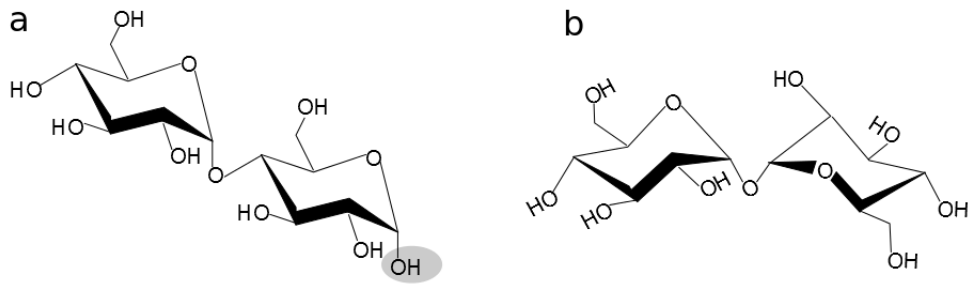


Figure 2.6: Reducing (a) and non-reducing (b) disaccharides

and ψ_H, and can therefore be translated when necessary.

$$\begin{aligned} \phi &= H_1 - C_1 - O_1 - C_x \\ \psi &= C_1 - O_1 - C_x - H_x \\ \omega &= O_1 - C_6 - C_5 - O_5 \end{aligned} \tag{2.1}$$

The conformational preference of the glycosidic torsional angles determines how the monomers are oriented with respect to each other. Since monomers rings are often comparatively rigid, the glycosidic linkages, determine the overall conformation of the disaccharide.

Disaccharides are the smallest, and simplest, level of carbohydrate which possesses the qualities of a long polysaccharide chain. As a result studying disaccharide conformation is the natural first step in studying polysaccharide conformations.[46,47]

2.1.3 Polysaccharides

Most carbohydrates exist as long chains of monosaccharides, which adopt complex three-dimensional tertiary structures. These structures may be rigid, but most often they are highly flexible and can adopt several conformations, allowing for the variety of roles that they play. Determining the three-dimensional structure of these flexible saccharides is a major challenge in glycobiology. It is essential for a better understanding of their chemistry and in order to develop carbohydrate based drugs.[43,45,46,49,50] Typically, a saccharide chain of up to 20

monosaccharides is termed an oligosaccharide and one greater than 20 residues is called a polysaccharide.

Nearly all polysaccharides are comprised of pyranoside residues which tend to adopt the rigid chair conformations. As a result, the main contributor to carbohydrate tertiary structures are the glycosidic bonds (ϕ, ψ and ω torsion angles) that link the monomers. Therefore, a key step to determine carbohydrate structure is characterizing the glycosidic linkages. These linkages are not easily characterized as they are flexible, and the degree of flexibility varies from linkage to linkage. Within a particular saccharide the linkages need not act independently but often influence each other.[22,43,46,49,51]

2.2 Experimental Techniques to Study Carbohydrate Structure

Carbohydrates structure is primarily studied using ^{1}H and ^{13}C NMR spectroscopy. Homonuclear coupling constants, $^{3}J_{H-H}$, are important in determining ring conformations and estimating the ω torsion angle. Linewidth temperature coefficients and low temperature hydroxyl $^{3}J_{HO,C,H}$ couplings are used to study hydrogen bonds.[52] Nuclear Overhauser effects are also key in determining carbohydrate structure, ^{1}H-^{1}H NOE constants, shed light on intraresidue protons and, in rare cases, on interresidue protons between consecutive monosaccharides. ^{1}H-^{13}C NOEs as well as T_1, T_2 and T_{1p} relaxation times detail distances between carbons and hydrogens, and are used to determine glycosidic linkage conformations.[53,54]

Oligosaccharides show flexibility on short timescales, through angle and bond vibrations, and longer times scales, through rotations of the dihedral angles. This flexibility and internal motions complicate results obtained from NMR studies, which do not sample at these timescales. Therefore, results obtained from NMR studies represent an "average" structure and one has to be aware of the various conformational permutations when interpreting spectra.[45,49,50] As a result, Molecular Mechanics (MM) and Molecular Dynamics (MD) simulations are often used in conjunction with NMR techniques when studying carbohydrates, as they help shed light on the dynamic properties of these molecules.[24,25,55]

Other techniques such as Mass Spectrometry (MS) and X-ray crystallography can also be used to shed light on structural information. Carbohydrates are often hard to crystallize for X-rays.[24,49,50,56] One way of obtaining carbohydrate crystals is by trapping them with proteins, either through covalent bonds or as carbohydrate ligands for protein complexes. Results from X-ray crystallography do not show dynamic characteristics of the molecules, but detail structural information such as torsion angles and hydrogen bonding with surrounding water molecules. Mass Spectrometry is also used, as it is a highly sensitive requiring requiring small sample amounts. MS, while useful, is unable to distinguish among different stereoisomers, as they have the same mass.[53,54]

2.3 Carbohydrates and the Immune System

Bacterial capsular polysaccharides have several characteristics that aid in the pathogen's survival. The capsule, which is hydrophilic in nature, protects the bacteria from dessication, which aids in host-to-host transmission. Once in a host, the capsule also prevents opsonins, molecules that act as immune system targets, from coating the bacteria, thus preventing phagocytosis and the activation of the complement pathway. In some instances, the outer polysaccharide is the chemical structure of molecules produced by human cells and as a result not recognized as a foreign body by the immune system[10,11,20,57]

When recognized as a foreign body, the capsular polysaccharides of most pathogenic bacteria act as T cell independent antigens. In this type immunity, the T cell independent antigens directly interact with polysaccharide specific B-cells which produce antibodies against the pathogen. For this mechanism, immunity is often not developed in children under 2 years old. In addition to this, the mechanism does not induce sustained memory.[11]

Glycoconjugates aim to change polysaccharides to T cell dependent antigens, like proteins, and thus better the immune response. A glycoconjugate is formed by covalently linking the carbohydrate to a protein carrier, most often a highly immunogenic modified bacterial proteins such as: tetanus toxoid (TT), mutated diphtheria toxin (CRM197) or outer membrane protein (OMP).

With glycoconjugates, the T cell dependent immune response is now possi-

ble and involves two types of cells: B cells, which recognize the polysaccharide, and helper T cells, which recognize the carrier. The T-cell dependent immune response will now be induced through a series of steps. The response first starts when the polysaccharide-protein hapten binds to the B cell receptors of polysaccharide specific pre-B cells, and is enveloped into the endosome. In the cell, the protein portion of the hapten is digested into peptide epitopes, regions of the antigen which are recognized by the immune system. These peptides then bind to protein molecules known as Major Histocompatibility Complex-Class II (MCHII) molecules. The MCHII molecules carry the small regions of the antigen to the cell surface and present them to the $\alpha\beta$ receptor of the $CD4^+$ T cells, activating them. Once active, the T cells produce cytokines which are then released to stimulate the maturation of the pre-B cells to B-cells, as well as induce immunoglobin class switching from IgM to polysaccharide IgG. The B cells now release saccharide specific IgG antibodies.[11]

2.4 Conjugate Vaccines

2.4.1 Use of Conjugate Vaccines

The *Haemophilus influenzae* type b, or Hib, conjugate vaccine was first licensed in 1987 in the United States. Since its introduction into the vaccine schedules, Hib vaccine has significantly reduced the incidence rate of disease in both developed and developing countries. The efficacy of the Hib conjugate vaccine has been reported to be between 95-100%, which is a marked improvement over the earlier polysaccharide vaccine, which was discontinued in 1988 in the United States.[58] Unlike its polysaccharide predecessor, Hib conjugate vaccines are immunogenic in high risk patients (infants, HIV-positive patients, etc.). Over the years, the Hib polysaccharide, polyribosylribitol phosphate (PRP), has been conjugated to various protein carries: tetanus toxoid (TT), diphtheria toxoid (D), mutant diphtheria toxin (CRM197) and outer membrane protein (OMP), to form the conjugate polysaccharide vaccines.[59]

The success of the Hib conjugate vaccine promoted the development of conjugate vaccines against *Neisseria meningitidis*. The first *N.meningitidis* vaccine

was a polysaccharide against serogroups A and C.[60] Following this, a polysaccharide vaccines against serogroups A,C and W135 as well as serogroups A, C, W135, and Y were developed. These vaccines were not immunogenic in infants which led to the development of conjugate vaccines. Meningococcal C conjugate (MCC) vaccines were first developed and licensed in 1999. The serogroup C saccharides are conjugated to diphtheria or tetanus toxoid proteins.[60,61] The use of MCC has shown remarkable decrease incidence not just in those vaccinated, but in non-vaccinated people as well due to herd immunity.[61] A tetravalent conjugate vaccine (A, C, Y, and W135) is also available in the United States and has shown to be highly effective. *N.meningitidis* serogroup B causes 50% of meningococcal disease in the world and is unfortunately is not currently protected against by vaccines, as its outer polysaccharide (an $\alpha(2\rightarrow8)$linked N-acety-neuraminic acid polymer) is present in human tissue and as a result is not recognized by the immune system.[60]

2.4.2 Pneumococcal Vaccines

History of Pneumocccal Vaccines

The history of pneumococcal vaccines can be seen in Figure 2.7. The original pneumococcal vaccines were whole cell vaccines, first introduced in 1918, administered to the patient in two doses.[10,62] These vaccines had very strong side effects, and as a result, the 1920s saw the development of polysaccharide-based vaccines.

In 1920, the capsular polysaccharides and nucleoproteins of the bacteria were chemically isolated from bacterial cultures. When animals were immunized with the capsular polysaccharides, it was found that serotype specific antibodies were produced, and with the immunization of the nucleoproteins, general antibodies to pneumococcus were produced. Unfortunately, the general antibodies were poorly immunogenic while the type specific antibodies did elicit protection. This lead to the development of vaccines, with the aim of combining the different

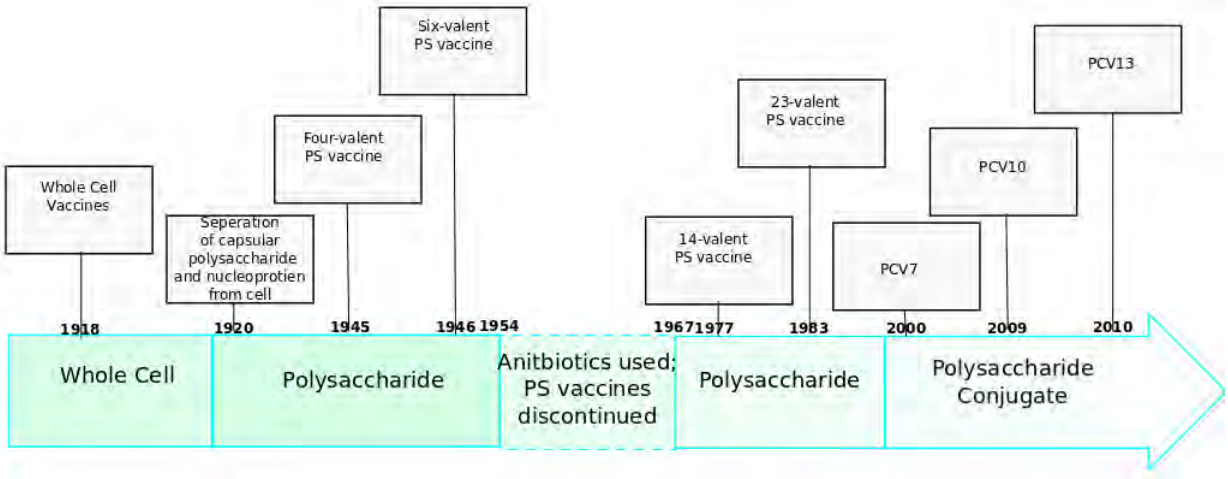


Figure 2.7: History of Pneumoccoal Vaccines

polysaccharides to from a polyvalent vaccine which would protect against a range of IPD causing serotypes.[62]

The first of these polyvalent vaccines, a four-valent vaccine, showed good efficacy, protecting against serotypes 1, 2, 5, and 7.[21,62] Despite its efficacy, the then newly-developed hexavalent polysaccharide vaccine was withdrawn due to the wide use of antibiotics such as penicillin.[10] The eventual rise of antibiotic resistant pneumococci renewed the interest in polysaccharide vaccines.[10,62] From the late 1960's this interest led to the successful development of 14-valent and then 23-valent polysaccharide vaccines.[62] However, while polysaccharide vaccines work well in healthy adults, they have failed to produce the same protection for high-risk population groups, such as young children, the elderly, and people infected with HIV.[10,21]

It was found that coupling pneumococcal polysaccharides to protein carriers enhances the efficacy in the high risk population groups.[20] These highly effective conjugate vaccines do pose some challenges: the chemical conjugation of carbohydrate to protein is technically difficult, the production of conjugate vaccines is expensive, and conjugate vaccines only offer coverage for a limited number of serotypes.[20]

Current Status of Conjugate Vaccines against *Streptococcus pneumoniae* serogroup 6

Over the years, the development of vaccines against *S.pneumoniae* has made progress and can be seen in Table 2.1. The first conjugate vaccine licensed against *S.pneumoniae* is Prevnar/Prevenar (PCV-7), contained seven serotypes, including serotype 6B. PCV-7 has been one of the biggest milestones in the prevention of pneumococcal disease, and with its introduction into vaccine schedules new carriage and IPD due to PCV-7 serotypes has dramatically reduced, disease caused by serotype 6B was nearly eliminated. In addition to this, it was also observed that IPD due to 6A reduced, suggesting that 6B in PCV-7 offered some cross-protection against 6A.[63] While the introduction of PCV-7 has reduced IPD rates, serotype replacement has been observed, and there was a noted rise in IPD from non-vaccine serotypes (NVT). With respect to serogroup 6, there was a noted increase in IPD caused by 6C, a serotype previously undistinguished from 6A.[21,64] In Spain, it was shown that the prevalence of 6C increased with the introduction of PCV-7. In children the prevalence of 6C, in the collected samples, was 0.1% pre PCV-7, which increased to 1% post PCV-7.[65,66] In adults, the study showed that prevalence increased from 0.3% to 1.7% in the same period.[65] A study in Cleveland Ohio, noted that since the distribution of PCV-7, there has been an increase in incidences caused by vaccine-related serotypes (207.4%), and non-vaccine serotypes (18.4%).[34] These increases were accounted for by mostly 19A, 6C, and 22F and serogroup 15.[34] Serotype 6C is currently the most prominent disease causing serotype of serogroup 6.

Vaccine	Carrier	Serotypes
PCV-7	CRM197	4, 6B, 9V, 14, 18C, 19F, 23F
PCV-10	DD/TT	PCV-7 + 1, 5, 7F
PCV-13	CRM197	PCV-10 + 3, 6A, 19A
PCV-15	CRM197	PCV-13 + 22F, 33F

Table 2.1: Pneumococcal Conjugate Vaccines[1]

This prompted the development of higher valency conjugate vaccines, PCV-10 and PCV-13, the latter also included serotype 6A. The inclusion of 6A in

PCV-13 has been shown to induce a ten-fold greater response against 6A than what was induced by PCV-7. In addition to this, the 6A conjugate also seems to induce protection against 6C in recent studies. Cohen et. al. showed that when vaccinated with PCV-13, the incidence of 6C was significantly lower than in the population group that received the PCV-7 vaccine.[4,31,67] Serotype 6D, was initially synthesized in the lab in 2009, and not considered to be present naturally. It has since been reported 6D has been identified to cause IPD in South America, Fiji and Korea.[6,28,32,38] It is yet to be determined if PCV-13 serotypes offer protection against 6D disease, as has been suggested, or if 6D is an emerging serotype due to serotype replacement.

It is clear that there is a constantly changing epidemiology of predominant serotypes, which calls for an increased coverage by the conjugate vaccines, as a result PCV-15 is currently in development.[42]

Chapter 3

Molecular Simulation

Molecular Modelling attempts to determine the "best" structure of a molecule, and to use this information to explain the physical properties exhibited. Due to their flexibility, carbohydrates are particularly difficult to simulate. Quantum Mechanical principles are used to predict observable chemical properties from first principles, by describing microscopic systems by "wave functions" that fully describe all the physical properties of that system. However, it is often computationally impractical to simulate larger systems, such as oligosaccharide systems, using Quantum Mechanics and thus various other methods, such as force field methods are employed. Typically, Molecular Mechanics (MM) and Molecular Dynamics (MD) force field methods are used to for carbohydrate simulations.

The basic principle of force field methods is to treat a system as a collection of balls (atoms) and springs (bonds). The molecular system's energy can then be described mechanically, as a function of its resistance to bond stretching, angle bending atom crowding etc.

3.1 Force Fields

The potential energy function of a molecule is defined as the sum of the energy contributions from all bond stretching, angle bending, dihedral rotations, and all

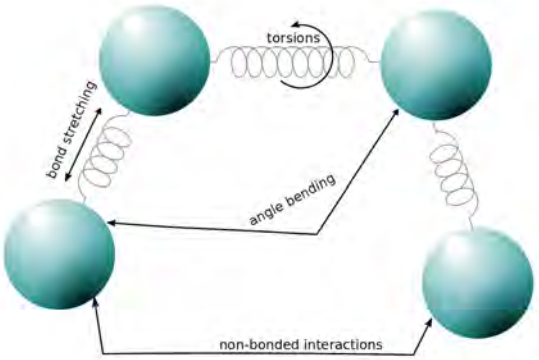


Figure 3.1: Ball and spring model

significant non-bonded atom interactions (Eq. 3.1).

$$E = \sum_{bonds} E_{stretch} + \sum_{angles} E_{bend} + \sum_{dihedrals} E_{torsion} + \sum_{pairs} E_{non-bond} \tag{3.1}$$

A force field consists of mathematical expressions and numerical parameters of the terms above, the first three of which represent bonded potential terms and the last non-bonded potential.

The *bond stretch* term in this equation models the energy of the bond (represented as a spring) when it is stretched or shortened past its equilibrium/natural bond length. In the definition of $E_{stretch}$ (Eq. 3.2), $k_{stretch}$ is the bond force constant, b is the bond length, and b_{eq} is the equilibrium bond length.

$$E_{stretch} = k_{stretch}(b - b_{eq})^2 \tag{3.2}$$

The second term of the potential energy function, the *angle bending* term, models the energies of angles corresponding to triatomic units, say atoms *i-j-k* (Eq. 3.3). Just as with the bond stretching term, k_{bend} is the bond force constant, a is the angle currently formed by *i,j,k*, and a_{eq} is the *i,j,k* angle equilibrium. The second half of E_{bend} is known as the Urey-Bradley term, and models a non-covalent spring between the outer atoms i and k. In this term, r_{ik} is the distance between atoms i and k, r_{ub} is the equilibrium distance and k_{ub}, is the force constant. The

Urey-Bradley term is only active when $k_{ub} \neq 0$.

$$E_{bend} = k_{bend}(a - a_{eq})^2 + k_{ub}(r_{ik} - r_{ub})^2 \tag{3.3}$$

The *torsional term* describes four sequentially bonded atoms, say *i-j-k-l* (Eq. 3.4). The dihedral angle, α is the angle between the plane created by atoms *i-j-k* and the plane *j-k-l*. The integer constant n, where $n \geq 0$, is a multiplicity term which indicates periodicity. When $n>0$, k is the multiplicative constant and β, is the phase shift angle which when $n=0$, is the equilibrium angle.

$$E_{torsion} \begin{cases} k(1 + cos(n\alpha + \beta)), & \text{if } n > 0 \\ k(\alpha - \beta)^2, & \text{if } n = 0 \end{cases} \tag{3.4}$$

The final term, the *non-bonded term*, calculates the van der Waals interactions (the Lennard-Jones potential) and electrostatic interactions (the Coulomb potential) of pairs of non-bonded atoms (Eq. 3.5). Calculating the non-bonded interactions is the most computationally expensive step of a MD simulation. In the Lennard-Jones potential, r_{ij} is the distance between a pair of atoms, E_{min} is the well depth, and R_{min} is the minimum interaction radius for the pair of atoms. As the distance r_{ij} increases, the potential approaches 0 quickly, and so a cut-off radius, after which all potentials are assigned as 0, is often specified to reduce computational cost. The Coulomb potential measures the attractive and repulsive forces between atoms based on atomic charges, here $q_i q_j$ are the charges on atoms i and j, C is the Coulomb's constant, ϵ_o is the dielectric constant, and ϵ_{14} is a scaling constant.

$$\begin{aligned} &E_{non-bonded} = E_{LJ} + E_{Coulomb} \\ &where: \\ &E_{LJ} = (E_{min}) \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min}}{r_{ij}} \right)^{6} \right] and, \\ &E_{Coulomb} = \epsilon_{14} \left(\frac{C q_i q_j}{\epsilon_o r_{ij}} \right) \end{aligned} \tag{3.5}$$

3.2 Force Field Parameterization

The potential energy function is a mathematical function which has to be transformed into an empirical force field. The empirical forcefield then relates chemical structure and conformations to their energies. The process of fitting actual numbers to the equation parameters for each atom type (force constants, equilibrium values angles etc.), pair (bond and non-bonded parameters), triplet (angle parameters), and quartets (dihedral parameters), is known as parameterization.

Parameterization of a force field is a difficult and tedious task as different bonds and atom types have to be considered. A refined force field accounts for multiple atom types and bonds. That is for example, there will be different parameters to account for a sp^2/sp^2 C-C bond and a sp^3/sp^3 C-C bond.

Parameter values are obtained from a combination of experimental and *ab initio* calculations. The bond stretch and the angles bending parameters can usually be obtained from experimental data, however obtaining experimental data for the torsion and non-bonded parameters is considerably harder. As a result these parameters are often determined using theoretical quantum mechanics calculations.

3.3 Carbohydrate Force Fields

Over the years there have been significant improvements in carbohydrate force fields, brought on, in large part, by the advances in computing and NMR spectroscopy. The increase in computing power has increased the number and complexity of quantum mechanical simulations, which has allowed for parameterization of atomic charge and torsional parameters from first principles, which were not previously assignable based on existing experimental data. Improvements in experimental techniques, particularly NMR, have also positively informed carbohydrate parameters. In return, MD simulations are now routinely performed to generate models which have been used to explain NMR results.[68]

Force fields are characterized into various classes, based on the properties of their functional forms. The most widely used force fields belong to the Class I followed by the Class II groups. Class I force fields are also known as quadratic force fields, as their bond and angle terms are defined by quadratic terms. The pop-

ular force fields: CHARMM, AMBER, and GROMOS all fall into this category. These force fields model atoms as soft balls which are given static and partial charges, held together with bonds modelled as harmonic springs. The functional from of such force fields is generally of the form shown in Equations 3.1 - 3.5. These biomolecular force fields were all initially developed for protein research, but have since added parameter sets to model a wider range of biomolecules, including carbohydrates. Class II force fields, such as MM3, MMFF, UFF and CFF, have the same philosophy and ball-spring molecular model however their functional forms are significantly more complex as they attempt to better handle electronic effects. The sole force field used for the purpose of this work has been the CHARMM force field.[68]

3.3.1 CHARMM Force Fields

CHARMM force fields were initially developed to model proteins, and parameters for DNA, RNA, and lipids were subsequently added to CHARMM. Ha et al. were the first to develop carbohydrate parameters for monosaccharaides, namely α-D-glucopyranose.[69]

Following this, Reiling et al. developed carbohydrate parameters to extend the CHARMM22 force field. The focus of this work was to develop accurate torsional parameters using high-level *ab initio* calculations on molecules representing carbohydrate fragments, which showcased all the possible dihedrals in non-substituted carbohydrates.[70]

Kuttel et al. derived a CHARMM Carbohydrate Solution Force Field (CSFF) which took into account solvation and rotation of the primary and secondary alcohols of a pyran ring.[71]

Hatcher et al. created new CHARMM carbohydrate parameters to include not just hexopyranoses, but all possible glycosidic linkages between them, accounting for chiralities, thus allowing for all un-substituted polysaccharides to be modelled. The parameter set was further extended to include not just hexopyranoses, but linear polyalcohols, inositol and linear sugars, using both experimental data and QM calculations. These parameters were added to the CHARMM additive all-atom force field which allows for the simulation of glycoproteins and glycolipids in

addition to substituted polysaccharides.[72–75] The CHARMM forcefield has been further extended to include glyco-phosphates and glyco-sulphates.[76]

3.4 Energy Minimization

Energy minimization is often used in molecular conformational analysis. Here, the conformational space is explored with the intention of finding atom positions for which the force field function (Eq. 3.1) takes on its minimum value. At the minimum, the first derivative of this function is zero and the second derivatives are positive.

Minimization typically involves iteratively moving atomic positions to locate the energy minimum. Several algorithms can be used for minimization, the most popular being the steepest descent and conjugate gradient methods. The first derivative minimization method, known as the conjugate gradient method, is used to calculate the minima in this work, where points are moved based on the initial direction of the point and its negative gradient. When minimized, atoms are moved along gradients so as to reduce the forces on them, thus reducing their potential energy.

Minimization does not guarantee that the global minimum in the conformational space is reached, but that the molecule is in its local minimum that it is in a well on the potential energy surface. Molecules often have several local minima, and as a result a minimization would find a local minimum that is closest to the starting configuration of the search. In order to obtain a global minimum, one will often have to carry out minimizations from various starting conformations.

3.5 Molecular Dynamics

Molecular Dynamics (MD) uses Newton's equation of motion (Eq. 3.6a) in a stepwise fashion, to calculate the movement of particles in a system caused by forces brought on by the interaction of the system's particles, where F_i, m_i *and* a_i is the force, mass and acceleration of particle i. The equation can also be represented as a gradient of the potential energy, E (Eq. 3.6b). The combination

of Equations 3.6a and 3.6b produces Equation 3.6c, where E is the potential energy of the entire system (Eq. 3.1) and r is the position of atom i.

$$F_i = m_i a_i \tag{3.6a}$$

$$F_i = -\nabla_i E \tag{3.6b}$$

$$m_i a_i = -\frac{d}{dr_i} E \tag{3.6c}$$

The primary advantage of running a MD simulation is that one is able to see details of the route taken, that is, a detailed time-dependent history of properties and interaction is observable. In addition to this, a MD simulation will produce an ensemble of conformations, unlike MM.

3.5.1 Numerical Integration

There are various possible methods that can be used to integrate the equations of motion in a simulation such as: Gear predictor-corrector algorithms, Euler algorithm, Verlet algorithm etc. The Verlet algorithm is the most popular and for this work, the "velocity-Verlet" integration was used. This method starts by calculating the velocity of the atoms between time steps ($v_{n+\frac{1}{2}}$) using the current velocity (v_n) and acceleration ($M^{-1}F_n$) of the atoms. This is then used to obtain the positions and the forces of atoms for the next time step ($r_{n+1},\ v_{n+1}$), which in turn is used to calculate the velocity in the next time step (v_{n+1})

$$\begin{aligned} v_{n+\frac{1}{2}} &= v_n + M^{-1}F_n \cdot \Delta t/2, \\ r_{n+1} &= r_n + v_{n+\frac{1}{2}}\Delta t, \\ F_{n+1} &= F(r_{n+1}), \\ v_{n+1} &= v_{n+\frac{1}{2}} + M^{-1}F_{n+1} \cdot \Delta t/2 \end{aligned} \tag{3.7}$$

In order to reduce computational costs, NAMD uses a multiple time step integration scheme. Here short-range forces consisting of bonded terms and van

der Waals and electrostatic terms within a local distance are calculated at every time step. Long-range forces, van der Waals and electrostatic interactions at greater than the local distances, are calculated less frequently as they change more slowly. The choice of time step size is very important, the smaller the time step the more accurate the simulation, which comes at a computational cost. Typically, a time step of 1 to 2fs is chosen for the short range calculations while the long range forces may be calculated up to every 5fs.

3.6 Simulation Ensembles

A MD simulation can be conducted under various ensembles. Often a microcanonical ensemble is used and here the number of atoms n, the volume of the unit cell V and the total energy E or the total entropy S of the system is conserved. Under the nVE ensemble as energy is conserved, entropy is maximised and conversely, as entropy is conserved energy is minimized under the nVS ensemble.

An alternative to the microcanonical ensembles, are the canonical ensembles. Here the system is immersed in an infinite heat bath, which maintains an almost constant temperature, without any particle exchange. The two alternatives are nVT, a constant number of atoms, volume, and temperature, and nPT, a constant number of atoms, pressure and temperature, ensembles.

Temperature can be maintained by a variety of methods such as velocity rescaling, weak coupling with the heat bath using a Berendsen thermostat, stochastic collision such as an Anderson thermostat, or Langevin dynamics.

3.7 Simplifications

MD simulations are very tedious and computationally expensive, and there are several techniques used to reduce the calculations in the simulation.

3.7.1 Periodic Boundary Conditions

Bulk systems are modelled using periodic boundary conditions. Using this method, a central cell, often cubically shaped, is replicated and surrounded by its images.

All the particles, in the cell and its images, move identically. When a particle leaves the central cell from one side, an image of this particle re-enters from the opposite end of the cell (Figure 3.2). In doing so, the surface effects are eliminated, that is a particle doesn't disappear from the cell as it exists. This allows a large bulk system to be simulated, while keeping track of just a few particles at a low computational cost.

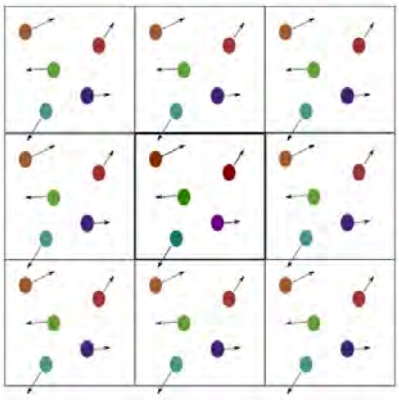


Figure 3.2: 2D representation of periodic boundary conditions

3.7.2 Non-bonded Interactions

Calculating the non-bonded interactions for the potential energy function is the most expensive part of a simulation (Eq. 3.5). This is a pairwise calculation, calculating the interaction of all the atom pairs, and as a result computational cost is N^2, proportional to the square of the number of atoms. This expense is reduced by using non-bonded cut-off distances.

van der Waals

The Lennard-Jones potential term is used to calculate the van der Waals interactions between non-bonded atoms. When a cut-off distance is specified, van der Waals interactions are not calculated between atoms which are further than the cut-off distance. All atom pairs, beyond this distance, are assumed to have a potential of zero. Smooth switching and a switching distance may also be set;

after the switching distance the potential is smoothly lowered so as to arrive at zero by the cut-off distance. This can be seen in the figure below (Figure 3.3).

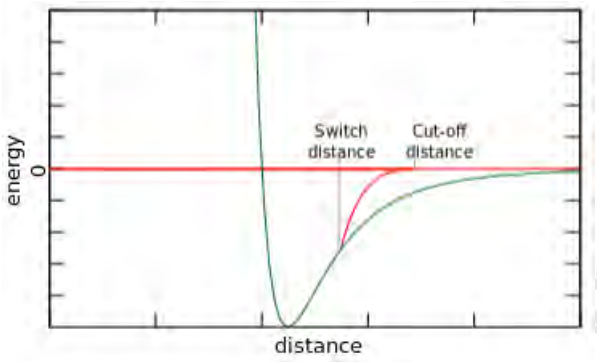


Figure 3.3: Lennard-Jones potential when cut-off is applied

Electrostatic Interactions

Electrostatic interactions are calculated in a similar fashion as the van der Waals interactions (Figure 3.4). When a cut-off is specified, atom pairs that are at a greater distance than the specified cut-off distance are assumed to have an electrostatic interaction energy of zero. When switching is activated (Figure 3.4 a), the potential is shifted so as to reach zero by the cut-off. Full electrostatic interactions (Figure 3.4 b), can also be calculated when multiple time step integration is employed; electrostatic interactions below the cut-off distance are calculated every time step and those beyond this distance are only calculated periodically.

Full electrostatic interactions may also be calculated using the Particle Mesh Ewald method (PME). The PME method calculates the short-range electrostatic interactions as a direct sum, and long range interactions are a sum in Fourier space. The terms in the Fourier space converge quickly and as a result can be truncated effectively, reducing computational cost.

3.7.3 Langevin Dynamics

The previous simplifications are used to reduce computational resources when explicit solvent molecules are part to the simulated system system. However often

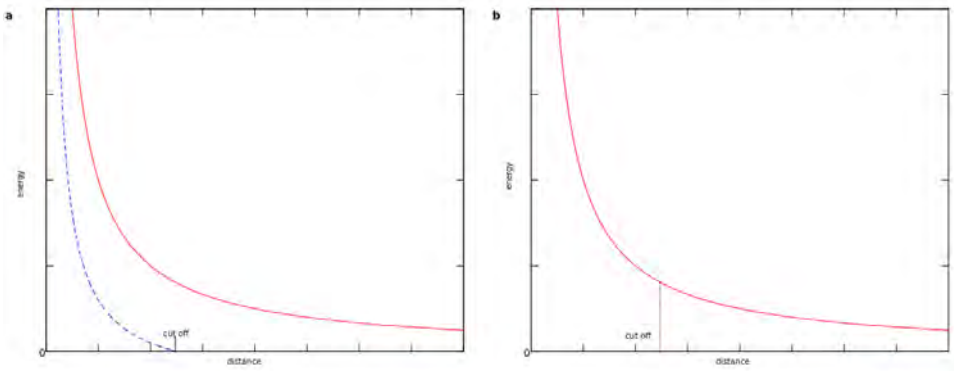


Figure 3.4: Electrostatic interactions with switching (a) or with full electrostatics (b)

even with these simplifications, the simulations are too computationally expensive and solvents are thus implicitly included. Implicit solvents can be simulated using Langevin Dynamics, which simulate the effects of solvent molecules on the solute, without adding them to the system.

Langevin Dynamics (Eq. 3.8) modifies the equation of motion of atom i with the addition of two terms which simulate solvent interactions.

$$m_i a_i = F_i - \gamma_i \mathbf{v} + R_i(t) \tag{3.8}$$

$\gamma_i \mathbf{v}$ describes the frictional force of the implicit solvent on the solute particles. This term opposes motion and is proportional to the particle's velocity. Different solvents can be simulated by altering the frictional coefficient γ_i.

The $R_i(t)$ term describes the stochastic forces caused by thermal fluctuations of the solvent on the solute molecule. When the two terms, frictional forces and stochastic forces, are set to zero, the Langevin equation (Eq. 3.8) reduces to Newton's equation of motion (Eq. 3.6a).

3.7.4 Constraint Dynamics

Constraint algorithms are used to make certain high frequency bonds rigid, that is holding them at a fixed bond length and angle. In doing so, it allows the simulation to take larger time steps. This is particularly useful when large systems are being modelled.

The most common method used to do this is the SHAKE algorithm which works in the Cartesian coordinates. During the dynamics simulation, atoms initially move first without any constraints, the magnitude of the constraints are then calculated and the algorithm is readjusted. The SHAKE algorithm is used most often to constrain water bonds or bonds involving hydrogen atoms.

3.8 Enhanced Sampling Methods

Molecular simulations have a huge influence on a variety of problems, but these techniques are often limited by computational time. It is useful to calculate the a Potential of Mean Force (PMF) which details the change in free energy along reaction coordinates, such as a change in conformation. A PMF can be computed when the simulation has sampled the potential energy surface throughly.

However, the progression of a simulation is hugely impacted by the starting structure(s) of the system; that is, a simulation started near one minimum well might find it very difficult to move into another well on the potential energy surface unless highly favourable conditions are met. The probability of these favourable conditions occurring, and thus a transition from one well to another, might be low. As a result the progression of a simulation is heavily time dependent, and enhanced sampling methods can be used to get over energy barriers. There are various methods to accelerate the occurrence of such favourable conditions such as: Adaptive Biasing Force, Umbrella Sampling and Free Energy Perturbation, to name a few.[77]

3.8.1 Metadynamics

Metadynamics is used in this work and is closely related to Umbrella Sampling. It computes free energies and aims to accelerate rare events in the simulations. Metadynamics works by moving along a trajectory defined by a chosen set of Collective Variables (CVs) which adds a bias to influence the dynamics of the system. A normal MD simulation is biased in the direction of lower free energy and Metadynamics aims to correct this bias by adding repulsive Gaussians potentials centred on the current position of the system on the potential energy

surface. As the system moves along the trajectory followed by the CVs, Gaussian potentials are added to the free energy space. Adding these potentials discourages the system to revisit previous configurations and enabling the system to jump over saddle points into new minimum wells. The history dependent potential function of the energy surface (V_G(s,t)) can then be calculated as a sum of these Gaussians, producing a PMF surface.

Critical to Metadynamics is the choice of the CVs (s = s_x, x = 1,...,n), which define the trajectory. The CVs need to be chosen so as to describe the configurational properties of the investigated molecules as completely as possible, thus reducing the investigation to just the dimensions of the CVs.

The height (h), width (w), and frequency of deposition (t) define the Gaussians (Eq. 3.9), and is critical to how accurately and quickly the PMF can be constructed, larger Gaussians lead to a faster the exploration of the conformational space but produce a coarser PMF. The width (w) of the Gaussian determine the resolution of the PMF, and the rate at which the Gaussians are added is the Gaussian height (h) divided by the frequency of deposition (t).

$$h \cdot exp(-\frac{|s - s(t)|^2}{2w^2}) \tag{3.9}$$

The history dependent potential V_G(s,t) is then constructed as a sum of Gaussians centered at the explored point (s), value of the CV, up till time t (Eq. 3.10).

$$V_G(s,t) = h \sum_{t_i} exp\left(-\frac{|s - s(t_i)|^2}{2w^2}\right) \tag{3.10}$$

Over time, the sums of the Gaussians will reproduce the free energy F(s) (Eq. 3.11). That is, if dynamics is run for long enough the inverse of V_G(s,t) will produce PMF.

$$\lim_{t\to\infty} V_G(s,t) \approx -F(s) \tag{3.11}$$

The use of an enhanced sampling method such as Metadynamics allows for the production of a PMF surface which will shed light on the the conformational preferences of the molecules being investigated.

Chapter 4

Modelling Methods

This work follows a systematic approach to study the structures of the serogroup 6 saccharides. The methodology used for this investigation can be seen in Figure 4.1.

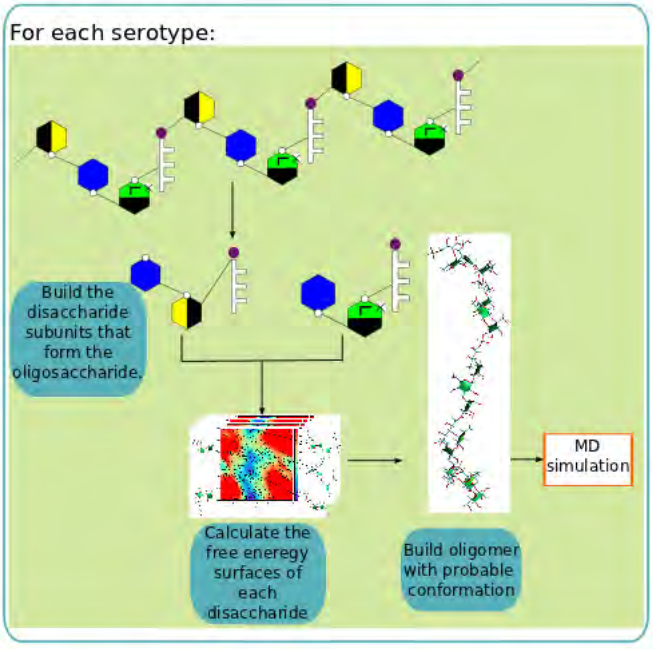


Figure 4.1: Overall methodology used for this study

The progression of a simulation depends on the starting structure(s) of the

system; that is, a simulation started near one minimum well might find it very difficult to move into another well on the potential energy surface unless highly favourable and unlikely conditions are met. As a result, it is important to get a good starting structure for the oligosaccharide.

This investigation begins by modelling disaccharide subunits that make up the serotypes. This is done to investigate the properties of the disaccharides themselves, and to use their optimum configurations to build a good starting structure for the serotype oligosaccharides.

4.1 Disaccharides

Studying disaccharide substructures of an oligosaccharide so as to understand the conformational preferences of the glycosidic linkages is often the first step to studying oligosaccharide structure, and has been used in previous studies.[78–82]

Torsion angles of a glycosidic linkage are the major contributor to saccharide flexibility and conformation. Producing Ramachandaran-like contour plots to represent the linkage's conformational preferences is becoming increasingly common.[22,51,83] The energy landscape, or potential of mean force (PMF), plots of a linkage are thus a function of the ϕ and ψ torsion angles of the disaccharide. These PMF plots indicate regions of low energy and therefore predict the disaccharide structure.

Two classes of disaccharide components were investigated. In the first class, the disaccharides were a result of treating the ribitol-phosphate naïvely as a residue and thus the following disaccharides were studied:

αDGal$p$(1→3)αDGlc$p$ and	αLRha$p$(1→3)DRibol5P and
αDGlc$p$(1→3)αDGlc$p$	αLRha$p$(1→4)DRibol5P
DRibol5P(O→2)αDGlc$p$ and	αDGlc$p$(1→3)αLRha$p$
DRibol5P(O→2)αDGlc$p$	

Ribitol-phosphate, being a linear sugar residue, is highly flexible and as a result a second disaccharide class was considered where the DRibol5P was treated a substituent to the αDGal$p$/αDGlc$p$(1→3)αDGlc$p$ and αDGlc$p$(1→3)αLRha$p$

linkages. In this class, the glycosidic linkage involving the DRibol5P was left to relax while the linkage between the other two residues was intensionally explored. As a result the following disaccharides were studied:

DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ and
DRibol5P(O→2)αDGlc$p$(1→3)αDGlc$p$

αDGlc$p$(1→3)αLRha$p$(1→3)DRibol5P and
αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P

The torsion angles for the DRibol5P(O→2)αDGal$p$/Glc$p$ linkages are defined as ϕ = PO_3-P-O_2-C_2 and ψ = P-O_2-C_2-H_2. For all the other disaccharides, the torsion angles for a (1→n) linkage are defined as ϕ = H_1-C_1-O_1-C_n and ψ = C_1-O_1-C_n-H_n. These dihedral angles can be seen in Figure 4.2 on disaccharides αDGal$p$(1→3)αDGal$p$ and DRibol5P(O→2)αDGal$p$.

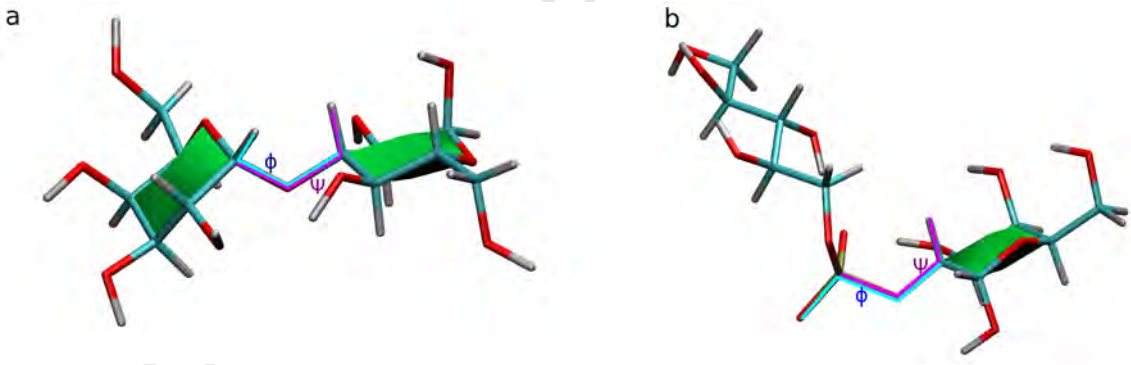


Figure 4.2: Dihedral angles on disaccharides αDGal$p$(1→3)αDGal$p$ (a) and DRibol5P(O→2)αDGal$p$ (b). ϕ is seen in blue and ψ is seen in purple.

Torsion angles are classified as syn when they fall in between 0° and ±90°, or anti when they are in between ±90° and ±180°. Disaccharides torsion angles ϕ/ψ can thus be classified as syn/syn, syn/anti, anti/syn, or anti/anti depending on their dihedral angles, which can be seen in Figure 4.3.

In this work, the atoms of the non-reducing residue are labeled using numbers, atoms on the reducing residue are labeled with a primed number and when ribitol-phosphate is treated as a substituent, its atoms are labeled with double primed numbers.

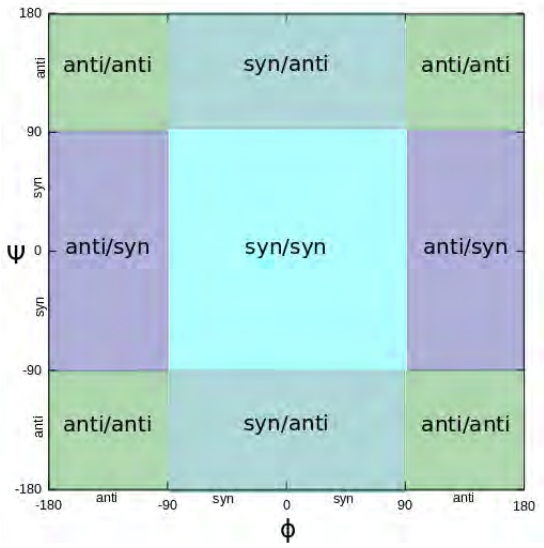


Figure 4.3: Disaccharides' syn and anti classification

4.2 Oligosaccharides

Three repeating unit long (12mers) oligosaccharides were built using the global minima calculated from the disaccharide energy landscapes. The serogroup 6 oligosaccharides are:

6A: $[\rightarrow 2)\alpha$DGal$p(1\rightarrow 3)\alpha$DGlc$p(1\rightarrow 3)\alpha$LRha$p(1\rightarrow 3)$DRibol5P(O$\rightarrow]$

6B: $[\rightarrow 2)\alpha$DGal$p(1\rightarrow 3)\alpha$DGlc$p(1\rightarrow 3)\alpha$LRha$p(1\rightarrow 4)$DRibol5P(O$\rightarrow]$

6C: $[\rightarrow 2)\alpha$DGlc$p(1\rightarrow 3)\alpha$DGlc$p(1\rightarrow 3)\alpha$LRha$p(1\rightarrow 3)$DRibol5P(O$\rightarrow]$

6D: $[\rightarrow 2)\alpha$DGlc$p(1\rightarrow 3)\alpha$DGlc$p(1\rightarrow 3)\alpha$LRha$p(1\rightarrow 4)$DRibol5P(O$\rightarrow]$

Two sets of oligosaccharides were built first using the global minima from the class 1 disaccharides and secondly using the global minima from the class 2 disaccharides.

Molecular Dynamics simulations were conducted on these oligosaccharides. Simulations were conducted in an nVT ensemble with implicit solvation. Implicit solvation was accomplished by conducting the simulations in vacuum with a high dielectric constant so as to reduce computational costs.

The conformation of the glycosidic linkages of the middle repeating unit through the course of the simulation were analyzed as they will most accurately relay polysaccharide information. These dihedral angles of the middle repeating

unit are plotted on their previously calculated contour maps showing how inter-residue interactions affect the preferred conformations of the glycosidic linkages.

20mer (five repeating unit) oligosaccharides were then built using the conformational preferences of the oligosaccharides seen from the MD simulations. The most populated conformation of the glycosidic linkages of the middle repeating unit from the final 100 ns of the MD simulations were used to build the 20mers, with subsequent minimization, to show the two most prevalent structures for each of the serotypes. This gives some insight into the possible structures of the serogroup 6 saccharides.

4.3 Simulation Conditions

4.3.1 NAMD, VMD, and force fields

NAMD was used to conduct all the simulations in this work. Version 2.8 for Linux-x86_64-multicore was used for the metadynamics simulations, and version 2.9 for Linux-x86_64-MPI was used for oligosaccharide simulations.[84] VMD version 1.8.7 for LINUXAMD64 was used for all the visualization.[85]

The CHARMM carbohydrate force field (C36) was used for all the simulations.[76] Serogroup 6 are phosphodiester containing saccharides and the phosphate parameters in this force field were adapted for the Ribol5P(O→)αDGal$p$/Glc$p$ linkages.

4.3.2 Disaccharides - Metadynamics

A metadynamics protocol in NAMD was set up to investigate the disaccharide conformational preferences about the conformational space (-180° to 180° for ϕ and ψ torsion angles.) Collective variables are defined and used to alter the dynamics of the system around a defined set of coordinates. Therefore, in this case the collective variables were defined as ϕ and ψ. Each of the collective variables were then rotated between -180° to 180° in 2.5° increments. The Gaussians were defined with a weight and height of 3 and 0.1 respectively.

The class 1 disaccharides (standard disaccharides) simulations were run for

400ns, while the disaccharides in class 2 (disaccharides with a ribitol-phosphate substituent) were were run for 1200ns to allow the ribitol-phosphate to settle into low energy conformations. The Gaussian parameters and simulation times have to be in the chosen carefully in order to accurately recreate the energy landscape.[77] Well chosen parameters will ensure that the entire collective variable space is explored through the course of the simulation. Figure 4.4 shows that this is the case for disaccharides αDGal$p$(1→3)αDGlc$p$ and DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ and that therefore appropriate parameters were chosen.

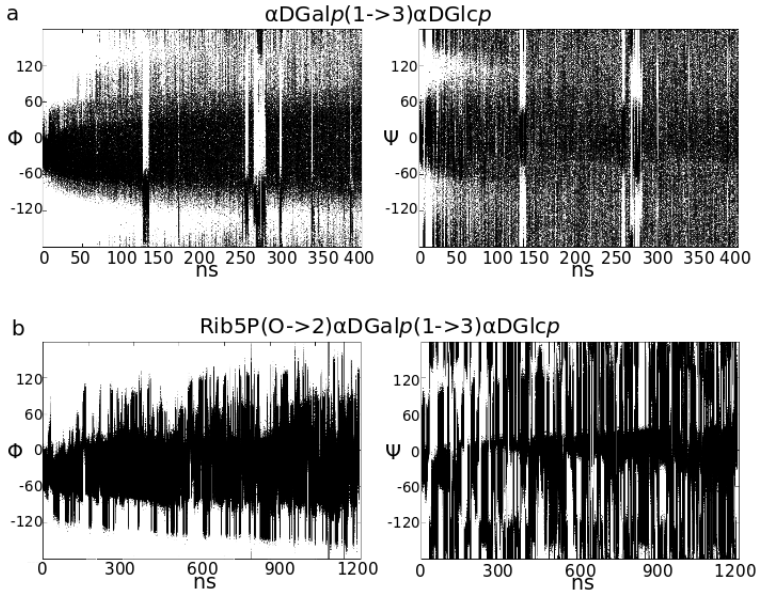


Figure 4.4: Distribution of the ϕ and ψ angles of disaccharides αDGal$p$(1→3)αDGlc$p$ (a) and DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ (b)

Validation: β-maltose

In order to validate the metadynamics protocol, a simulation of maltose (αDGlc$p$(1→4)βDGlc$p$) was conducted. The PMF of maltose, calculated with the metadynamics method, compares to the previously calculated umbrella sampling vacuum PMF (Figure 4.5).[86] Both PMF plots were produced using the same CSFF[71] forcefield, therefore allowing for all similarities and differences to be attributed to the sampling methods themselves.

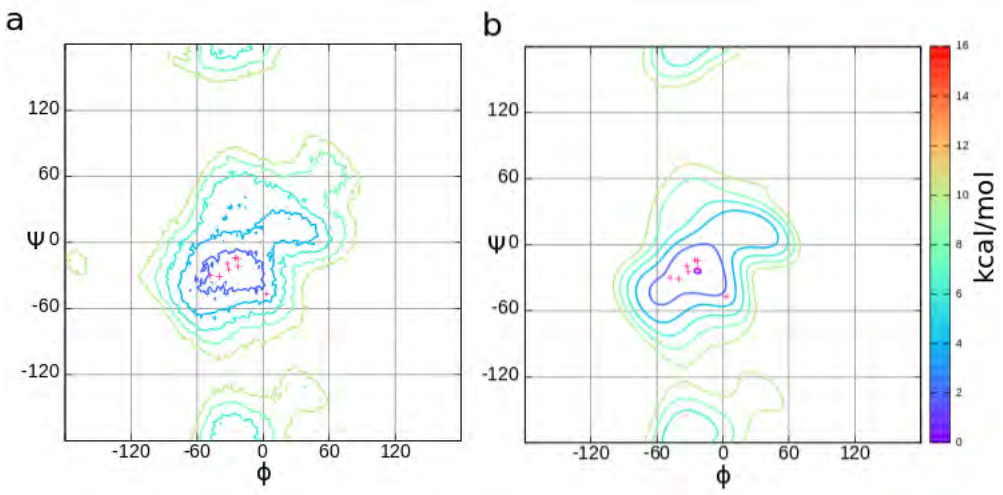


Figure 4.5: PMFs of maltose (αDglc$p$(1→4)βDglc$p$) calculated using method detailed in Section 4.3.2 (a) and by Kuttel et. al. (2005)[86] (b). Both PMFs show the energy contours at 2 kcal/mol increments from 0 kcal/mol to 10 kcal/mol. The pink crosses show the distribution of ϕ/ψ vaules of maltose from Table 4.1

A visual inspection of the two PMFs shows that the two PMFs are very similar. The low energy regions in both plots are conserved with the lowest energy well located at the syn/syn position. Metadynamics does produce PMFs that have more jagged contours than umbrella sampling, however the general shape of the syn/syn, and syn/anti wells are conserved. The global minimum from both these calculations are also in agreement, with the umbrella sampling producing a global minimum of ϕ,ψ= -22.5°, -25°, while the global minimum of maltose calculated in this work is at ϕ,ψ= -21.25°,-26.25°. Metadynamics does show a small anti/syn well, not produced by the umbrella sampling, however this well has a much higher energy than the syn/syn and syn/anti wells and is unlikely to be occupied. Numerous other maltose studies have been conducted, both experimental and theoretical (Table 4.1). The global minimum, ϕ,ψ= -21.25°,-26.25°, of this work is in close agreement with most of these published literature values, all of which do fall into the first contour of the low energy syn/syn wells seen in 4.5. This, along with the close agreement of these two energy landscapes in Figure 4.5 confirms that the umbrella sampling and metadynamics produce similar outputs.

Technique Used			
NMR ϕ,ψ	X-Ray ϕ,ψ	MM/MD ϕ,ψ	QM ϕ,ψ
-31.5, -24.5[52]	-23.2, -14.8[87]	-32.8, -19.5[46] -48, -30[47] -22.7, -21.7[89] -25.3, -13.9[52] -40, -31[90]	-47,-32.8[88]

Table 4.1: Literature ϕ,ψ (°) values of β-maltose

4.3.3 Oligosaccharide Simulations

Molecular dynamics simulations in an nVT ensemble were conducted in vacuum on each of the eight oligosaccharides (four with unsubstituted dihedral values and four with substituted dihedral values). A 300 K temperature was maintained using a Langevin thermostat with a damping coefficient of 1/ps. No electrostatic cutoffs were set thus all electrostatic interactions were calculated. A dielectric constant of 78.5 was used to approximate water and all covalent bonds involving hydrogen atoms were constrained to their equilibrium length using the SHAKE algorithm. The simulations were minimized for 13 ps, followed by a production run of 500 ns and the equations of motion were integrated using the Leap-Frog Verlet integrator with a 1 fs integration timestep.

For the analysis of the simulation trajectories, the dihedral angles of the middle repeating unit of the three repeating unit long oligosaccharides were plotted on the corresponding PMF surfaces to show the dihedral occupancy of the linkages throughout the course of the simulation. The most populated dihedral angles of the glycosidic linkages from the final 100 ns of the simulations were used to build 20mers of the serotypes, with subsequent minimization, to show the two most prevalent structures for each of the serotypes.

4.3.4 Building Carbohydrate PDB files

All the investigated saccharides were built using CarbBuilder, inhouse software which uses existing pdb structure files of sugar residues to build the saccharides.[91] The builder retrieves sugar pdb structures from text input (for example "aDGlc(1→4)aDGlc" for maltose), creating the requested bonds between the residues, and rotating the residues to either preset or user defined dihedral values,

finally writing out a pdb file of the newly created structure. CarbBuilder allows a user to build a variety of structures from disaccharides to regular polysaccharides including branched structures.

Chapter 5

The Glycosidic Linkages of Serogroup 6

In this chapter, the energetically prefered conformations of each of the disaccharide subunits of the *S.pneumoniae* serogroup 6 serotypes are identified. As stated in Chapter 2, disaccharides are the simplest carbohydrates exhibiting properties of longer polysaccharide chains[46,47] and are often the first focus in a study of carbohydrate conformation.[25]

The first step is to analyze the PMF plots of the glycosidic linkages of the disaccharide subunits that comprise serogroup 6, by performing detailed free energy analysis of the ϕ and ψ dihedral angles of the glycosidic linkage. The disaccharides are analyzed to so as to gain an understanding of their optimal structures, which will be used to build likely starting structures for the oligosaccharide MD simulations.

The structures of the serogroup 6 serotypes are:

6A: $[\rightarrow 2)\alpha\text{D}\text{Gal}p(1\rightarrow 3)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{L}\text{Rha}p(1\rightarrow 3)\text{D}\text{Ribol5P}(\text{O}\rightarrow]$

6B: $[\rightarrow 2)\alpha\text{D}\text{Gal}p(1\rightarrow 3)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{L}\text{Rha}p(1\rightarrow 4)\text{D}\text{Ribol5P}(\text{O}\rightarrow]$

6C: $[\rightarrow 2)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{L}\text{Rha}p(1\rightarrow 3)\text{D}\text{Ribol5P}(\text{O}\rightarrow]$

6D: $[\rightarrow 2)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{D}\text{Glc}p(1\rightarrow 3)\alpha\text{L}\text{Rha}p(1\rightarrow 4)\text{D}\text{Ribol5P}(\text{O}\rightarrow]$

These oligosaccharides were divided into disaccharide subunits for individual analysis as shown in Figure 5.1. Modelling the serogroup 6 serotypes is particularly challenging, as these saccharides contain the flexible ribitol-phosphate moiety.

Ribitol is unique among the residues of the serogroup in that it is a linear alditol and possesses a high degree of internal conformational freedom. This makes analysis of the linkages containing ribitol challenging, which is further complicated by the presence of the phosphodiester linkage. Ribitol thus has a potentially large effect on the overall secondary structure of the serotypes.

The analysis of disaccharide subunits begins with ribitol-phosphate being treated as a standard residue. This treatment results in the division of the serogroup into seven disaccharides, henceforth labeled **1**, **2**, **3**, **4** and primes, which indicate serotype variations relative to 6A (class 1 in Figure 5.1).

1: DRibol5P(O→2)αDGal$p$ (6A and 6B)
1′: DRibol5P(O→2)αDGlc$p$ (6C and 6D)

2: αLRha$p$(1→3)DRibol5P (6A and 6C)
2′: αLRha$p$(1→4)DRibol5P (6B and 6D)

3: αDGal$p$(1→3)αDGlc$p$ (6A and 6B)
3′: αDGlc$p$(1→3)αDGlc$p$ (6C and 6D)

4: αDGlc$p$(1→3)αLRha$p$ (6A, 6B, 6C and 6D)

The flexibility of the ribitol-phosphate prompted a second division of the serotypes, this time into trisaccharides, effectively treating the ribitol-phosphate as a substituent on a residue. This results in the division of the serotypes into four trisaccharides, namely **r3**, **4r** and their primes (class 2 in Figure 5.1).

r3: DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ (6A and 6B)
r3′: DRibol5P(O→2)αDGlc$p$(1→3)αDGlc$p$ (6C and 6D)

4r: αDGlc$p$(1→3)αLRha$p$(1→3)DRibol5P (6A and 6C)
4r′: αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P (6B and 6D)

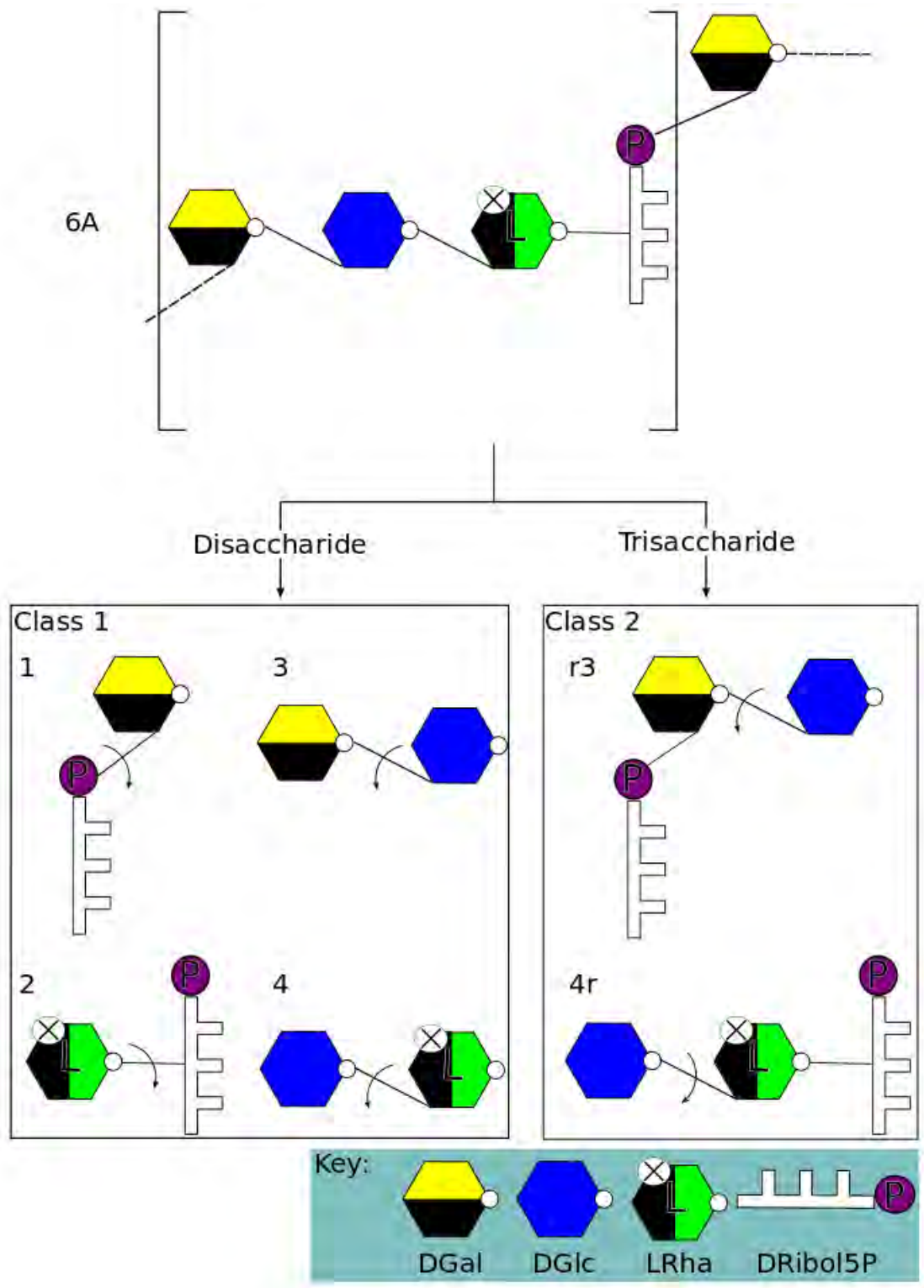


Figure 5.1: Illustration of the division of *S.pneumoniae* serotype 6A into subunits. Fragments are classified into two classes. Class 1 contains disaccharides and class 2 contains trisaccharides where the ribitol is effectively treated as a substituent. Arrows indicate the glycosidic linkage of interest. Other serotypes have equivalent divisions.

5.1 Disaccharides containing ribitol-phosphate

The disaccharides containing with ribitol-phosphate are examined so as to interrogate the flexibility the linear alditol containing disaccharides: DRibol5P(O→2) αDGal$p$ (**1**), DRibol5P(O→2)αDGlc$p$ (**1′**), αLRha$p$(1→3)DRibol5P (**2**) and αLRha$p$(1→4)DRibol5P (**2′**).

5.1.1 DRibol5P(O→2)αDGal$p$ and DRibol5P(O→2)αDGlc$p$ (Group 1)

Ribitol-phosphate is (PO→2) linked in serotypes 6A and 6B to a αDGal$p$ (**1**) and in serotypes 6C and 6D to a αDGlc$p$ (**1′**)(Figure 5.2).

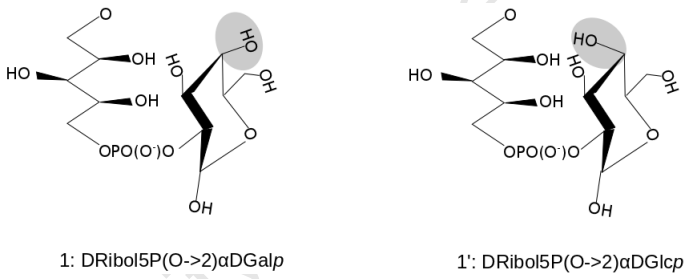


Figure 5.2: Structure of **1**: DRibol5P(O→2)αDGal$p$ and **1′**: DRibol5P(O→2)αDGlc$p$

The PMF surface of **1** and **1′** (Figure 5.3 (a and b) and Table 5.1) confirms that these are very flexible linkages with a large number of small energy wells. The global energy minimum is located in the syn/anti (well A) position, but the anti/anti conformation (well B) is almost equally favourable. Three wells (A, B, and C) are within 2 kcal/mol of the global minimum and four wells within 4 kcal/mol (C, D, E and F). This can be accounted for by the bond between the residues (PO→2) which results in a flexible ribitol tail, that may bend in numerous ways to form a variety of stabilizing hydrogen bonds.

The highly similar PMF surfaces calculated for the dihedral linkages **1** and **1′** shows that the substitution of a glucose for a galactose does not alter the preferred conformation of the linkage significantly. This is to be expected from the structural similarity of the molecules: galactose is a C4 epimer of glucose (Figure 5.2), a point of difference far removed from the glycosidic linkage. Figure

5.3 (c) shows a difference plot of molecules **1** and **1′**, and shows the minimal difference between these linkages. The difference plot highlights that the biggest difference between the two PMFs is in the ψ-syn region where the galactose disaccharide is higher in energy. As 99.7% of the area of the two PMF plots is within a difference of 2 kcal/mol, the rest of this discussion will be limited to disaccharide **1**.

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
DRibol5P(O→2) αDGal$p$ 1 (6A,6B)	-63.75 -78.75,(2.42),E 51.25 -71.25,(3.57),F	63.75 161.25,(0.00),A -46.25 153.75,(1.20),C	171.25 23.75,(1.98),D 156.25 -53.75,(2.96),G	166.25 163.75,(0.28),B
DRibol5P(O→2) αDGlc$p$ 1′ (6C,6D)	-61.25 -76.25,(1.69),E 46.25 -66.25,(2.70),F	56.25 158.75,(0.00),A -48.75 148.75,(1.34),C	173.75 21.25,(1.05),D 158.25 -66.25,(1.28),G	168.75 153.75,(0.86),B

Table 5.1: Minima in energy wells of DRibol5P(O→2)αDGal$p$ and DRibol5P(O→2)αDGlc$p$. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.3

Wells A, B and C and wells E, F, and G are separated by approximately 120° along the ϕ dihedral angle. Figure 5.3 (d) shows snapshots of the molecules from the labeled regions on the PMF, with the stabilizing hydrogen bonds indicated. Snapshots of the A, B and C wells (the 2kcal/mol wells), reveal that the 120° rotation about ϕ allows for two interresidue hydrogen bonds to form between the hydroxyl groups adjacent to the linkage (OH1′ and OH3′) and the phosphate oxygens (O5 and OPs). The snapshots from the 4kcal/mol wells (D, E, F, and G) shows one hydrogen bond between the OH1′ or OH3′ hydroxyls and the phosphate oxygens (O5 and OPs), accounting for the lower energy seen in A, B and C. Wells E, F and G give the corresponding "flipped" conformations with respect to A, B and C of the molecule with the glactose ring rotated by 180° about the ψ dihedral.

The snapshots also highlight the flexibility of the ribitol residue: in snapshots from regions: A, C, E and F ribitol-phosphate forms bended ring-like structures, yet in regions B, D and F the ribitol-phosphate has adopted a more extended structure from the galactose residue.

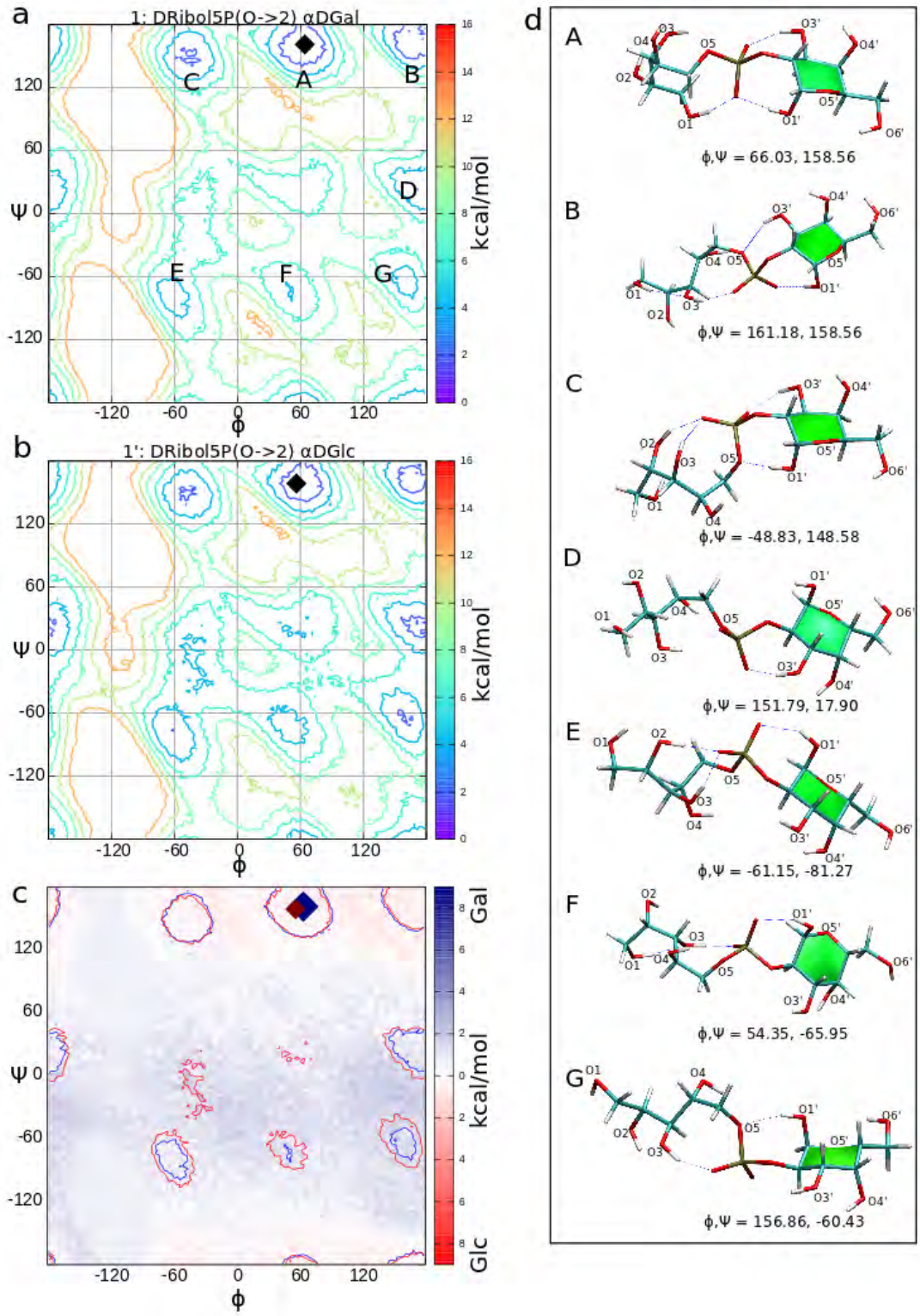


Figure 5.3: The potential of mean force (PMFs) plots of (a) DRibol5P(O→2)αDGal*p*, (b) DRibol5P(O→2)αDGlc*p* (the global minimum is indicated by the black diamonds (ϕ,ψ = 63.75°, 161.25° on (a) and ϕ,ψ = 56.25°, 158.75° on (b)) and contours are at 2kcal/mol increments). (c) Difference plot of (a) and (b), blue regions indicate higher energies seen in molecule **1** and red indicates higher energies seen in **1′**. (d) snapshots of the conformations of the molecule corresponding to the labeled regions of the PMF (a).

5.1.2 αLRha*p*(1→3)DRibol5P and αLRha*p*(1→4)DRibol5P (Group 2)

Serogroup 6 has two different rhamnose-ribitol-phosphate linkages (Figure 5.4): αLRha*p*(1→3)DRibol5P in 6A and 6C (fragment (**2**) in Figure 5.1) and αLRha*p* (1→4)DRibol5P in 6B and 6D (**2′**).

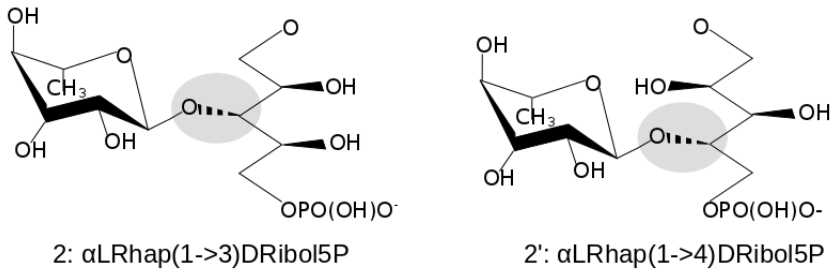


Figure 5.4: Structure of **2**:αLRha*p*(1→3)DRibol5P and **2′**: αLRha*p*(1→4)DRibol5P

The ϕ, ψ free energy profiles of the glycosidic linkages in **2** and **2′** (Figure 5.5) reveal a considerable difference between the (1→3) and (1→4) linkages. While the global minimum for both the disaccharides is in the syn/syn position (Table 5.2), there is a considerable shift between the two minima, with about a 100° flip on both ϕ and ψ: -46.25°, -36.25° for **2** and 56.25°, 63.75° for **2′**.

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
αLRha*p*(1→3)DRibol5P **2** (6A,6C)	-46.25 -36.25,(0.00),A	-21.25 -171.25,(0.88),C	-168.75 -6.25,(1.39),B	-91.25 176.25,(7.48)
αLRha*p*(1→4)DRibol5P **2′** (6B,6D)	56.25 63.75,(0.00),A	-13.75 -178.25,(5.01),C	-171.25 -21.25,(1.68),B -91.25 -46.25,(2.55),D	-91.25 146.25,(9.96)

Table 5.2: Minima in energy wells of αLRha*p*(1→3)DRibol5P and αLRha*p*(1→4)DRibol5P. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.5

Disaccharide **2** has three broad minima: A, B and C, while **2′** shows a different landscape: a shift in well A, an additional well D found in between well A and B and the loss of well C (the region has a 2 kcal/mol contour at **2** which is around 6 kcal/mol for **2′**).

This shows that the (1→3) linkage has a larger number of conformational options around ψ than the (1→4) linkage when ϕ is restricted to a syn conforma-

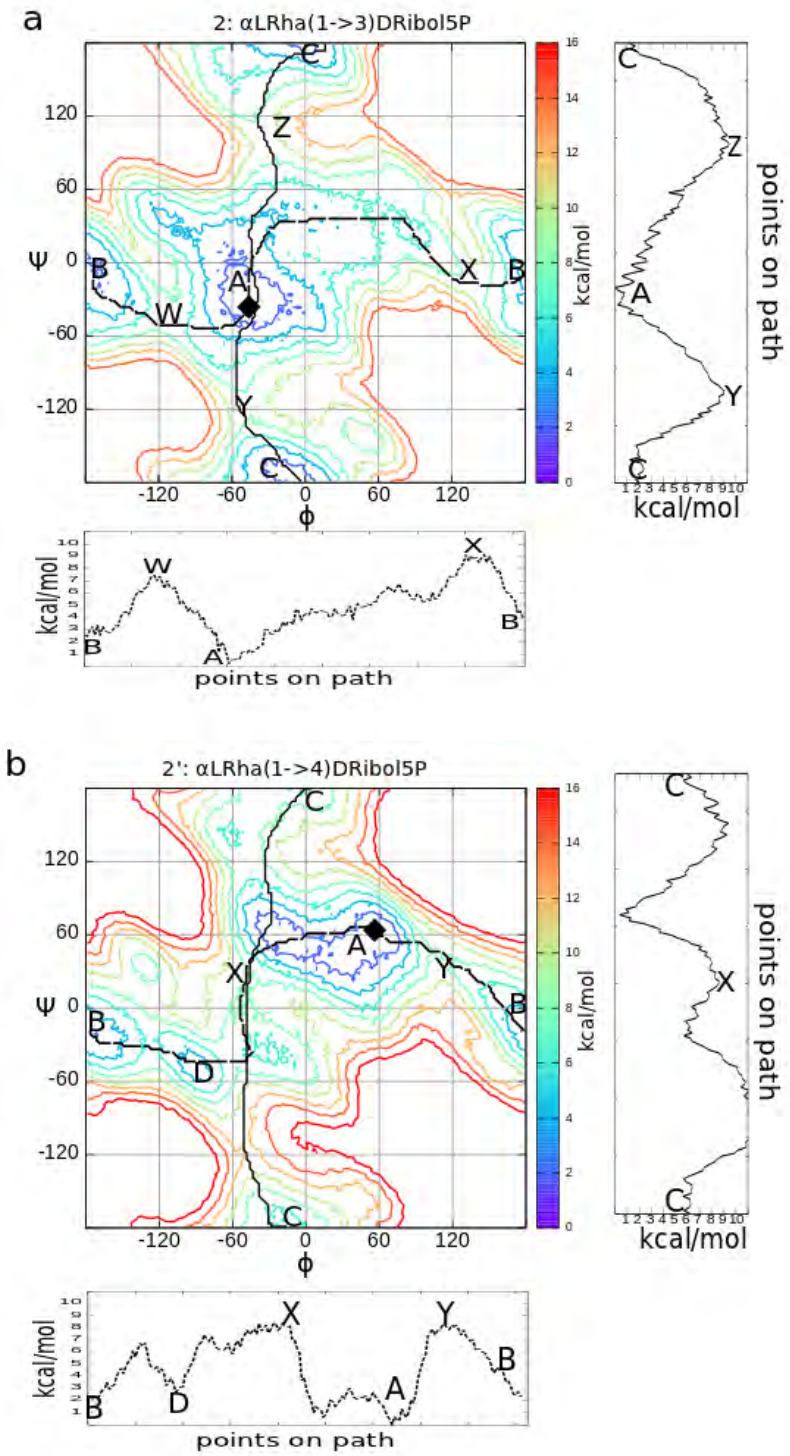


Figure 5.5: Potential of mean force (PMF) plots for (a) αLRha$p$(1→3)DRibol5P (global minimum at ϕ,ψ = -46.25°, -36.25°) and (b) αLRha$p$(1→4)DRibol5P (global minimum at ϕ,ψ = 56.25°, 63.75°). The global minima are indicated by the black diamonds, contours are at 2kcal/mol increments and the energy along the drawn paths on the PMFs are shown along side the PMF plots.

tion. However, the additional well D for the (1→4) linkage gives **2′** conformational options about the ψ angle not possible for the (1→3) linkage.

Figure 5.6 shows representative conformations corresponding to the regions labeled on the PMF plots. The highly flexible ribitol-phosphate bends to form hydrogen bonds both with other hydroxyl groups on the ribitol and to other residues. Figure 5.6, shows that most often (regions A, C and D), the ribitol bends forming a ring-like structure through an intraresidue hydrogen bond across the length of the ribitol: between PO′-O1′ and/or PO′-O2′. In addition to the ring-forming intraresidue hydrogen bonds, the B conformation snapshot for both molecules shows a stabilizing interresidue hydrogen bond between O2-O4′ in **2** and between O2-OP′ in molecule **2′**. There is a difference in the global minimum (A) of the two molecules, which is nearly a reflection about the line ψ = -ϕ. This causes the ribitol orientation, with respect to the rhamnose residue, to be flipped. That is, for molecule **2**, the phosphate end of the ribitol is far away from the glycosidic linkage, rather than adjacent to the glycosidic linkage as in molecule **2′**.

Possible paths of transitions between the low energy wells on the PMF plots were explored. The energy along these paths was extracted using the in house Heightmap Analyzer tool.[92] The energy along these paths, shown alongside the PMF plots in Figure 5.5 (a and b), gives a good indication of the energy barriers between the wells. For molecule **2** (the (1→3) linkage) transitions between A and B must occur over barriers W (7 kcal/mol) and X (9 kcal/mol). Barriers Y and Z, between the A and C regions, are approximately the same height (9.5 and 9 kcal/mol respectively), however Y is a wider and less steep. Based on these data, it is more likely that transitions between A and B would occur over the W barrier and between wells A and C would occur over barrier Y. For **2′** the two barriers X and Y, between the A and B/C conformations, are equally likely to be crossed as they are of the similar heights (8 kcal/mol).

The flexibility of the ribitol-phosphate thus demonstrated, prompted the investigation of the effects of ribitol-phosphate on ribitol adjacent disaccharide linkages.

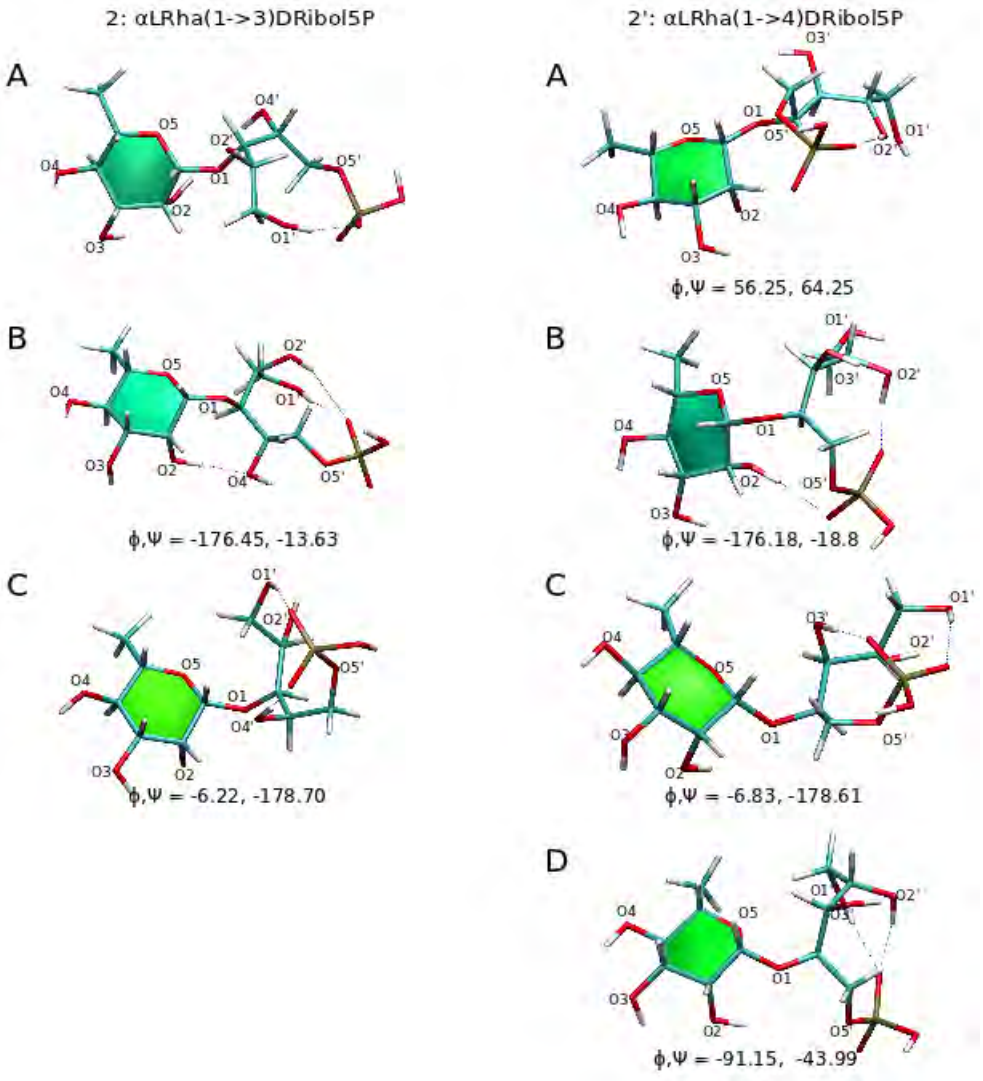


Figure 5.6: Examples of conformations of αLRha$p(1\rightarrow3)$DRibol5P and αLRha$p(1\rightarrow4)$DRibol5P corresponding to the labeled regions of the PMFs (Fig 5.5 (a and b))

5.2 Pyranose disaccharides

There are three linkages in serogroup 6 involving only pyranose residues: αDGal$p$ (1→3)αDGlcp (**3**), αDGlcp(1→3)αDGlcp (**3′**), and αDGlcp(1→3)αLRha$p$ (**4**). The PMF profiles for these linkages were calculated twice, firstly as the unsubstituted disaccharides (**3**, **3′**, and **4**) and secondly with a ribitol-phosphate substituent attached to the appropriate residue of the disaccharides (**r3**, **r3′**, **4r** and **4r′**).

Ribitol-phosphate is added to these disaccharides in order to investigate its effects on the adjacent disaccharide linkage, which will give a better indication of an appropriate disaccharide orientation in serogoup 6. The Ribitol-5P is attached to the appropriate residue of the respective disaccharides and the conformational space of the torsional angles between the galactose-glucose, glucose-glucose and glucose-rhamnose is intentionally explored while the ribitol is left unconstrained.

5.2.1 αDGal$p$(1→3)αDGlcp and αDGlcp(1→3)αDGlcp (Group 3)

The serotypes 6A/6B and 6C/6D differ in that 6A and 6B contain components **3** (αDGal$p$(1→3)αDGlcp) or **r3** (DRibol5P(O→2)αDGal$p$(1→3)αDGlcp), while 6C and 6D contain components **3′** and **r3′** (where the galactose residue is replaced by a glucose residue). The structures of the unsubstituted and substituted linkages can be seen in Figure 5.7. The PMFs around the αDGal$p$(1→3)αDGlcp and the αDGlcp (1→3)αDGlcp linkages are calculated in order to investigate the effects of the αDGal$p$/αDGlcp change on the (1→3) linkage.

Unsubstituted disaccharides

The αDGlcp(1→3)DGlcp disaccharide is commonly referred to as nigerose. A variety of experimental and theoretical studies have been conducted to determine the conformation of β-nigerose (αDGlcp(1→3)βDGlcp), an anomer of **3′** at C1 of the reducing residue. It is expected that β-nigerose would exhibit similar results to its α anomer and thus the results from this study can be compared.

Table 5.3 summarizes the previous results which are also plotted on Figure 5.8

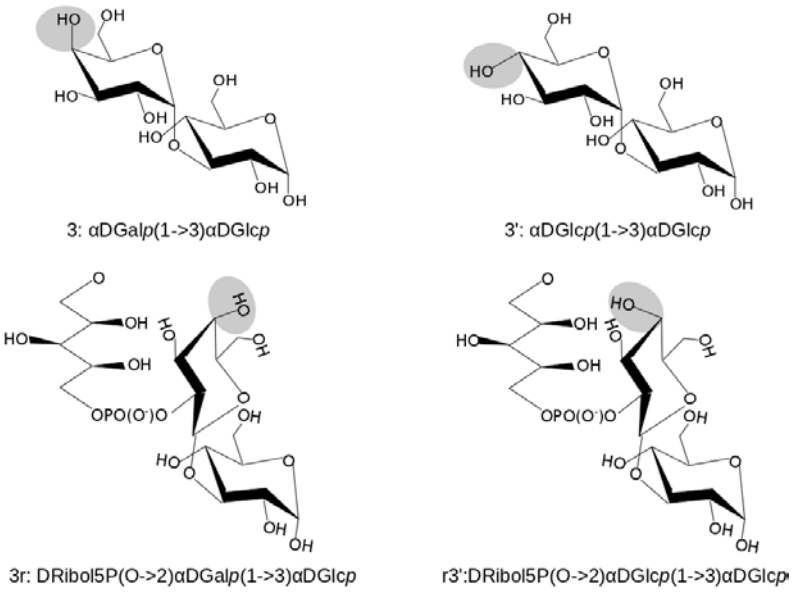


Figure 5.7: Structure of the αD-Gal$p(1\rightarrow3)\alpha$D-Glc$p$ and αD-Glc$p(1\rightarrow3)\alpha$D-Glc$p$ disaccharides

(b). The global minimum of α-nigerose, -25.25°,-23.75°, and the literature results in Table 5.3, including the most far removed quantum mechanical result, fall within the low energy syn/syn well (A) of **3′**, with nearly all the results falling within the first 2kcal/mol contour [Figure 5.8 (b)]. Taking into consideration the anomeric difference of the disaccharides, experimental shortcomings, as well variation in modelling techniques and forcefields, this agreement helps to validate the nigerose results calculated in this work.

Technique Used			
NMR ϕ,ψ	X-Ray ϕ,ψ	MM/MD ϕ,ψ	QM ϕ,ψ
-38.8, -22.6[52]	-20.1, -15.8[87]	-24, 0[47] -24.3, 0.6[46] -34.8, 47.4[89] -22.7, 1[93] -35.6, -9.6[52]	-46.77, -33[88]

Table 5.3: Literature ϕ,ψ values of β-nigerose. These values are plotted on the α-nigerose plot (Fig 5.8 (b))

The PMF plots of disaccharides **3** and **3′** [Figure 5.8 (a) and (b)], have a principal energy well (A) between -60° and 0° for ϕ and -60° and 60° for ψ. For both the unsubstituted disaccharides, the global minimum falls in this syn/syn (A) conformational region (-21.25°,-23.75°). In addition to the principal well, the disaccharides have a secondary syn/anti (B) well. The low energy wells are

approximately at 150° intervals about ψ, with ϕ constricted between -60° and 0° (Table 5.4).

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
αD-Gal$p$(1→3)αD-Glc$p$ 3 (6A,6B)	-21.25 -23.75 (0.00),A	-26.25 -173.75 (1.61),B	-91.25 -33.75 (6.49)	-91.25 166.25 (10.98)
αD-Glc$p$(1→3)αD-Glc$p$ 3′ (6C,6D)	-21.25 -23.75 (0.00),A	-28.75 -171.25 (2.11),B	-91.25 -23.75 (7.44)	-91.25 151.25 (10.40)

Table 5.4: Minima in conformational regions of αD-Gal$p$(1→3)αD-Glc$p$ and αD-Glc$p$(1→3)αD-Glc$p$. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.8

The difference plot of the **3** and **3′** PMFs [Figure 5.8 (c)] highlights the similarities of the two disaccharides, the energy profiles of these molecules is within 2.3 kcal/mol. Both **3** and **3′** show nearly identical energy profiles [Figure 5.8 (a and b)] and minima, with 99.9% of the energy difference between the two plots equal to or lower than 2 kcal/mol, as seen the the difference plot.

In the low energy conformations, two potential hydrogen bonds may form between the oxygens adjacent to the glycosidic linkage between: O2 or O5 and O2′ or O4′. Figure 5.9 shows snapshots of the two molecules in the A and B labeled minima showing stabilizing hydrogen bonds. The snapshots of the global energy minimum (A) of **3** shows stabilizing hydrogen bonds on either side of the glycosidic linkage, while the global minima of **3′** shows one of these hydrogen bonds (O2-O4′). The secondary well snapshots (B), (a 150° rotation about ψ) also show these interresidue hydrogen bonds between O5-O4′ for **3** and O2-O2′ for **3′**. Both molecules have two identical energy barriers, X and Y, between their wells A and B. However, barrier X is approximately 6.5 kcal/mol high, which is 2 kcal/mol lower than the barrier at Y. Thus, transitions between A and B wells are likely to go over barrier X.

The similarity of the profiles of these two molecules is to be expected, as the only point of difference is the C4 hydroxyl: axial in galactose and equatorial in glucose (Figure 5.7). This steric difference is far removed from the glycosidic bond and, as a result, has minimal effects on the energy profiles.

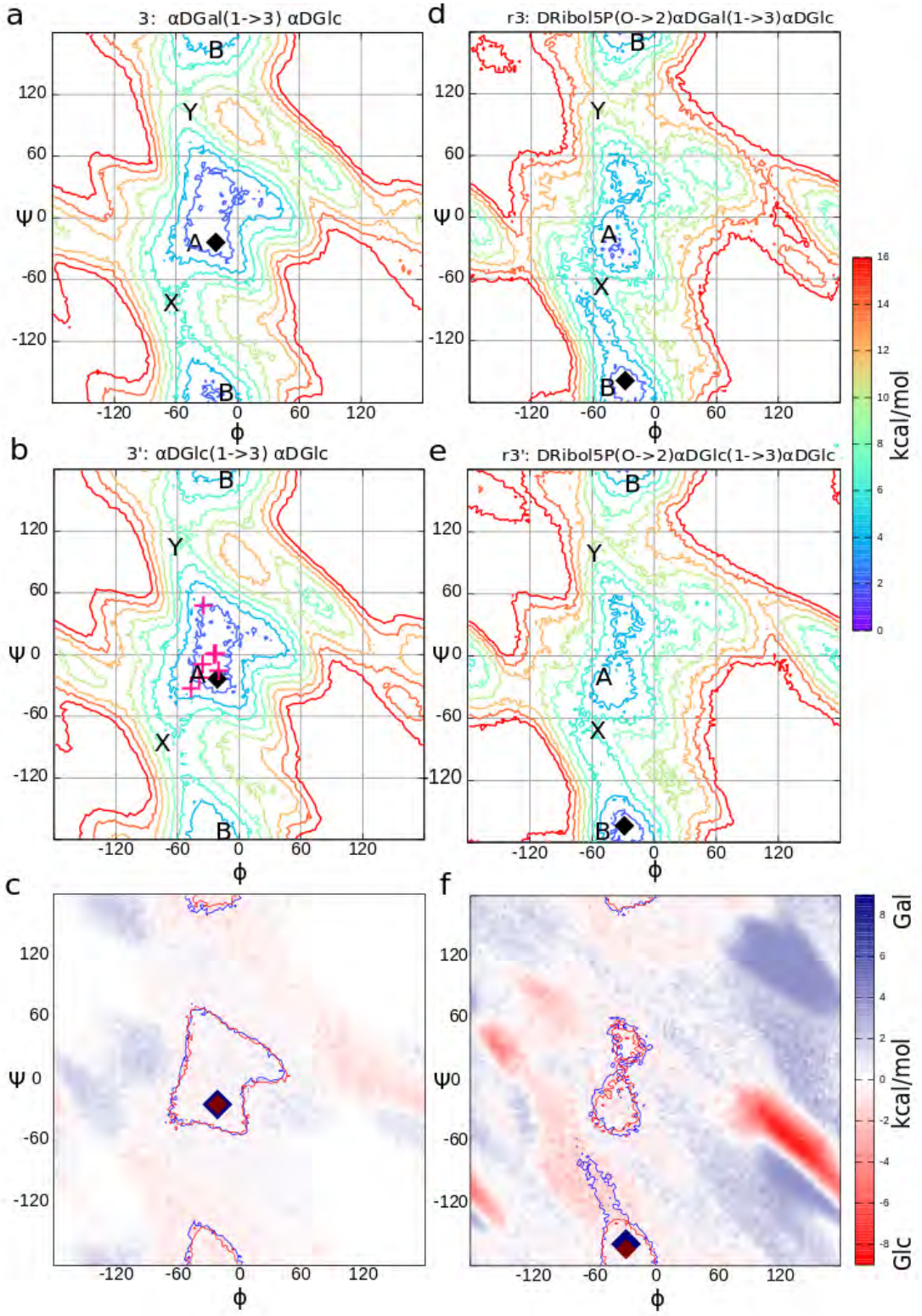


Figure 5.8: The potential of mean force (PMFs) plots of (a) αDGal*p*(1→3)αDGlc*p*, (b) αDGlc*p*(1→3)αDGlc*p*, (d) DRibol5P(O→2)αDGal*p*(1→3)αDGlc*p* and (e) DRibol5P(O→2)αDGlc*p*(1→3)αDGlc*p*. The contours are at 2kcal/mol increments and global minimum is indicated by the black diamonds (ϕ,ψ = -21.25°,-23.75° for (a) and (b), ϕ,ψ = -21.25°,-23.75° for (d) and ϕ,ψ = -21.25°,-23.75° for (e)) The pink crosses on the 3′ plot (b) indicate the locations of nigerose calculated in Table 5.3. Difference plots (c) of **3** and **3′** and (f) of **r3** and **r3′**, blue regions indicate higher energies seen in molecule **3**/**r3** and red indicates higher energies seen in **3′**/**r3′**.

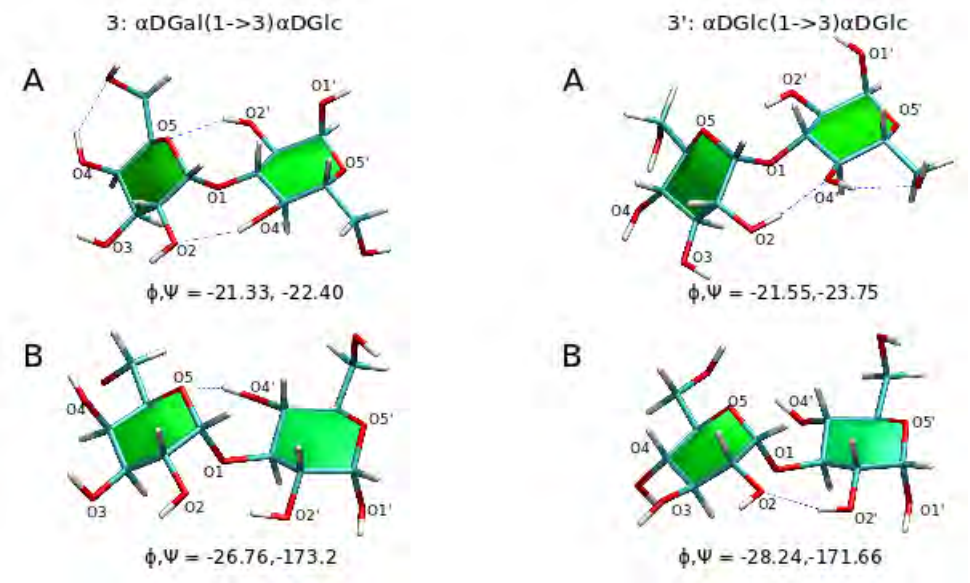


Figure 5.9: Examples of conformations of αDGal$p(1\rightarrow3)\alpha$DGlcp and αDGl$cp(1\rightarrow3)\alpha$DGlcp corresponding to the labeled regions of the PMFs (Fig 5.8 (a and b))

Substituted disaccharides

The PMF plots of **r3** and **r3′** are shown in Figure 5.8 (d) and (e). These plots show two energy wells, the primary anti/syn well (B) and a secondary syn/syn well (A).

In comparison to **3** and **3′**, the central A well has narrowed and split into two adjoining smaller wells. The biggest effect of the ribitol-phosphate on these disaccharides is to shift the global minima from the syn/syn position [Figure 5.8 (a) and (b)], to a syn/anti position: **r3** at -28.75°,-158.75° and **r3′** at -28.75°,-163.75°. With the addition of the ribitol, the syn/syn conformation is now about 3-4kcal/mol higher than the new global minimum for both molecules. This secondary minima of these molecules sits at -38.75°,-21.25° for **r3** and -53.75°,-28.75° for **r3′** (Table 5.5).

Looking at the difference plot of the two PMFs [Figure 5.8 (e)], 84.5% of the plot falls within 2 kcal/mol. While this difference (between **r3** and **r3′**) is greater than the difference between unsubstituted disaccharides (99.9%), it is located in the ϕ-anti region of the molecules, which is unlikely to be occupied as it falls above the 16 kcal/mol contour.

The wells A and B are approximately 150° apart along ψ, while ϕ is constricted mainly between -60° and 0°. This is similar to the unsubstituted disaccharides (**3** and **3′**). The addition of the ribitol-phosphate does alter the degree of conformational freedom of the molecules, and this can be seen in the energy barriers between the wells. With the addition of the ribitol-phosphate, the Y barrier is heightened, from 8kcal/mol for the unsubstituted linkages to 10 kcal/mol for the substituted linkages, and barrier X is been made wider, suggesting that the addition of the ribitol reduces the flexibility of the molecules. The extension of the B well of **r3** results in a more gradual climb to barrier X than **r3′** which would suggest that for the substituted disaccharides, **r3** has greater conformational freedom than **r3′**. Given the height difference of 4 kcal/mol between barriers X and Y, one would expect majority of the transitions between the A and B wells of these molecules to occur over barrier X, particularly for **r3**, due to the extended B well at barrier X.

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ r3 (6A,6B)	-38.75 -21.25 (1.20),B	-28.75 -158.75 (0.00),A	-91.25 -53.75 (5.36)	-91.25 -91.25 (10.80)
DRibol5P(O→2)αDGlc$p$(1→3)αDGlc$p$ r3′ (6C,6D)	-53.75 -28.75 (2.06),B	-28.75 -163.75 (0.00),A	176.25 -8.75 (5.44)	-91.25 -91.25 (11.73)

Table 5.5: Minima in energy wells of DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ and DRibol5P(O→2)αDGlc$p$(1→3)αDGlc$p$. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.8

Figure 5.10 shows snapshots of conformations of the molecules in their low energy wells. For the unsubstituted disaccharides, hydrogen bonds formed, in various combinations, between O2/O5 and O2′/O4′. However, the addition of the ribitol-phosphate to C2 has altered this, as hydrogen bonds are now formed with the phosphate oxygen atoms instead of O2. The snapshots of both molecules in their global minimum (B) shows hydrogen bonds between O5-O4′ and PO-O2′, while snapshots of the molecules in their secondary minima (A) shows a PO-O4′ hydrogen bond.

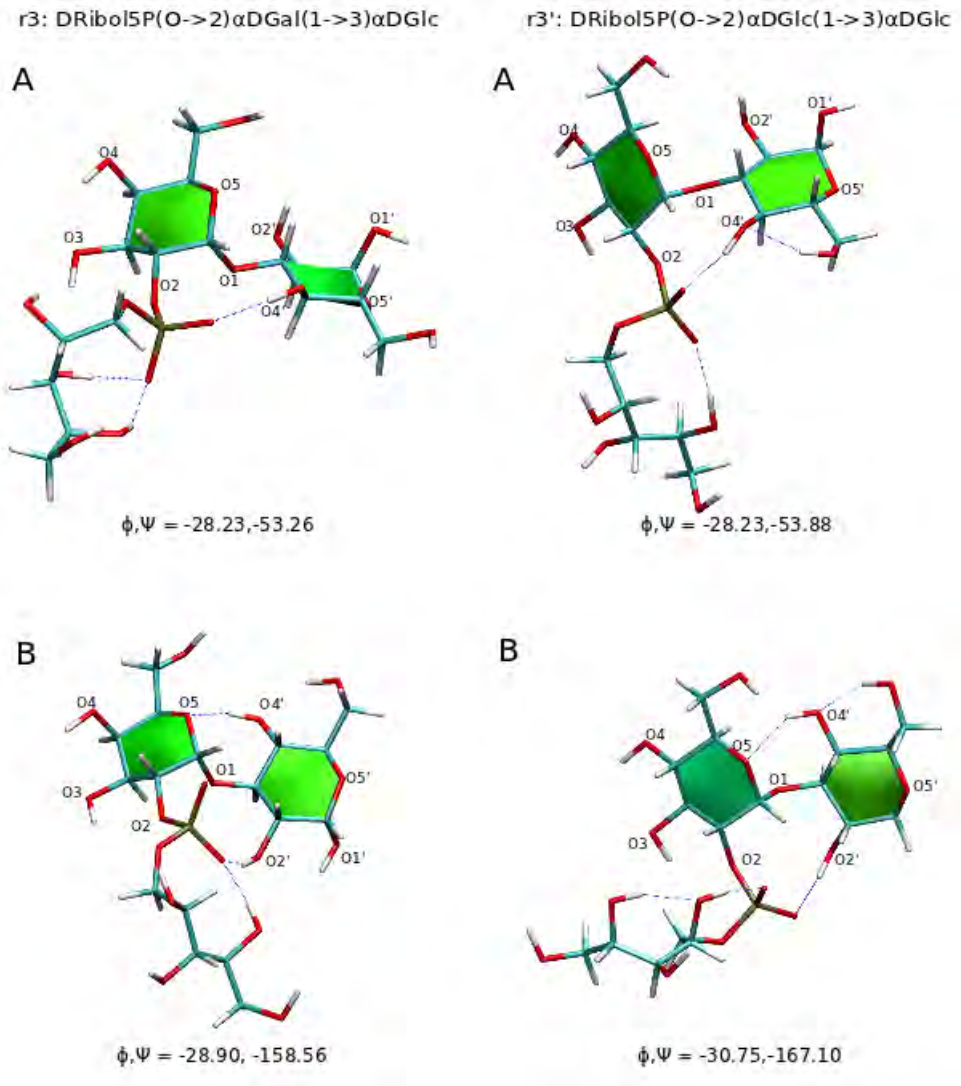


Figure 5.10: Examples of conformations of DRibol5P(O→2)αDGal*p*(1→3)αDGlc*p* and DRibol5P(O→2)αDGlc*p*(1→3)αDGlc*p* corresponding to the low energy regions of the PMFs (Fig 5.8 (d and e))

5.2.2 αDGlc*p*(1→3)αLRha*p* (Group 4)

The αDGlc*p*(1→3)αLRha*p* (**4** Figure 5.11) disaccharide is common to the four serotypes of serogoup 6. Though this linkage does not vary across the serotypes, it is important to know its conformational preferences to understand the overall structure of the serotypes. The αLRha*p* residue, the only L-sugar in the serogroup, is also linked to the ribitol-phosphate in two ways: through a (1→3) linkage (6A and 6C, **4r**) and a (1→4) (6B and 6D, **4r′**). These adjacent ribitol linkages might have an effect on the conformation that linkage **4** adopts in the serotypes, and as such it has been studied here.

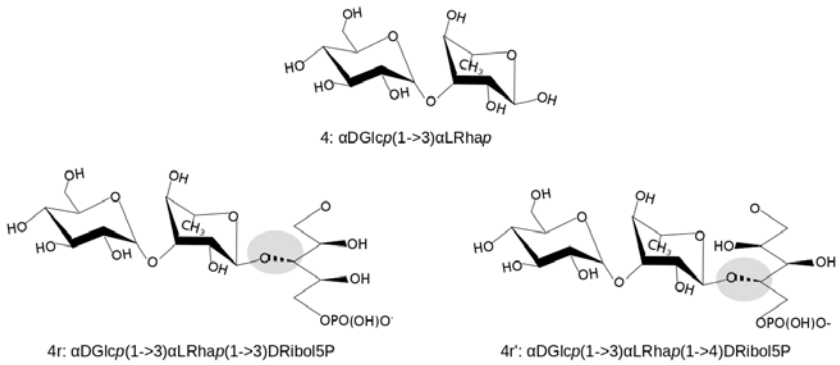


Figure 5.11: Structure of the αDGlc*p*(1→3)αLRha*p* disaccharides

Unsubstituted disaccharide

Figure 5.12 (a) shows that αDGlc*p*(1→3)αLRha*p* (**4**) has two energy wells: a large principal syn/syn energy well and a smaller anti/syn energy well. The global minimum for the molecule sites at ϕ,ψ= -43.75°, -28.75° (Table 5.6).

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
αDGlc*p*(1→3)αLRha*p* 4 (6A,6B,6C,6D)	-43.75 -28.75 (0.00), B -33.75 43.75 (1.49), C	-18.75 173.75 (2.32), A	-91.25 -51.25 (5.17)	-91.25 -91.25 (8.30)

Table 5.6: Minima in conformational regions of αDGlc*p*(1→3)αLRha*p*. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.12

The PMF plot of **4** shows a similar structure to the calculated landscapes of the other (1→3) linked disaccharides discussed in Section 5.2.1. However, the

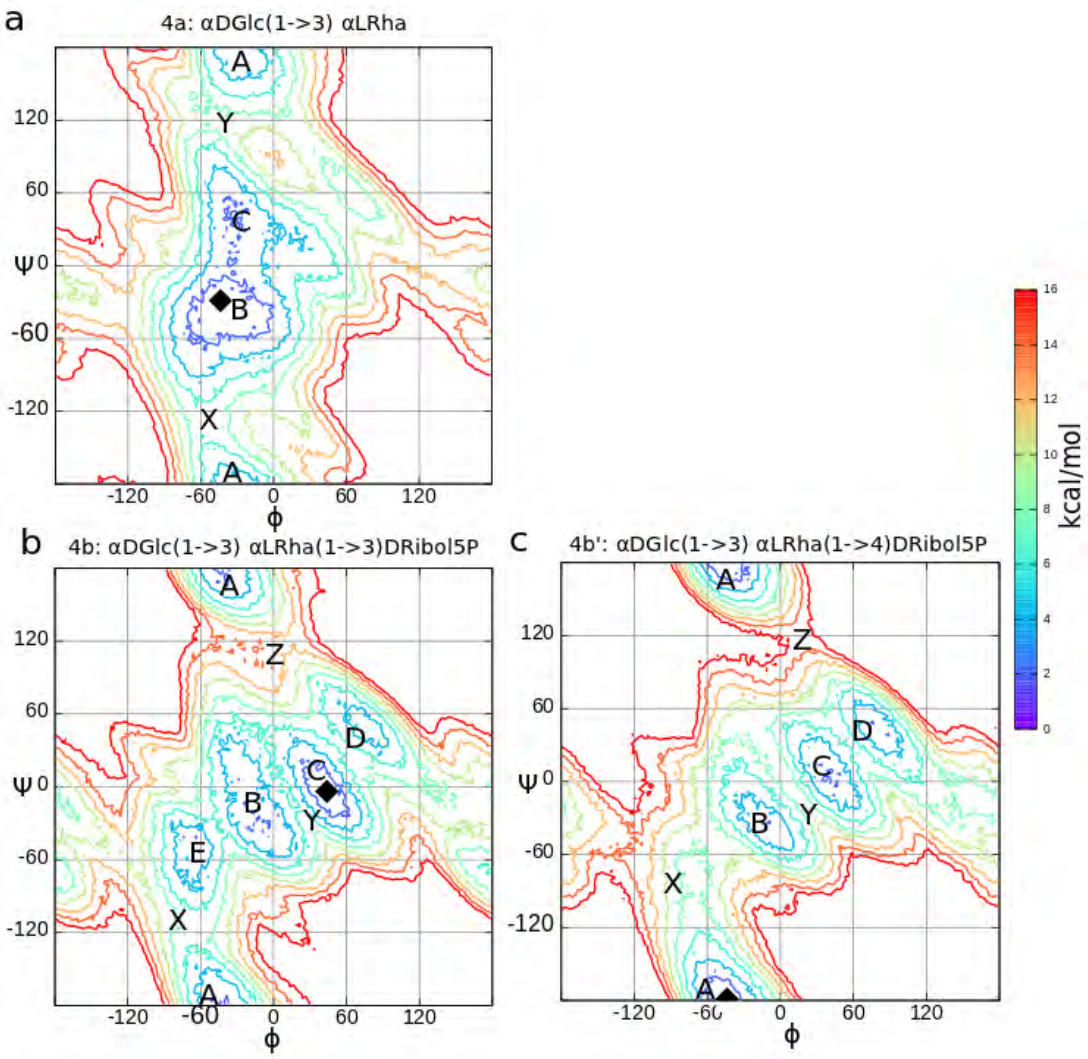


Figure 5.12: The potential of mean force (PMFs) plots of (a) αDglc$p(1\rightarrow3)\alpha$Lrha$p$, (b) αDglc$p(1\rightarrow3)\alpha$Lrha$p(1\rightarrow3)$Dribol5P and (c) αDglc$p(1\rightarrow3)\alpha$Lrha$p(1\rightarrow4)$Dribol5P. The contours are at 2kcal/mol increments and global minimum is indicated by the black diamonds (ϕ,ψ = -43.75°,-28.75° for (a), ϕ,ψ = 43.75°,-3.75° for (b) and ϕ,ψ = -43.75°,-178.75° for (c))

large syn/syn well is more extended along ψ when compared to the landscapes of **3** and **3′**, while ϕ is still limited to -60° to 0°. The two wells are approximately a 140° rotation about ψ apart. The two energy barriers X (7 kcal/mol) and Y (9 kcal/mol) between the wells are comparable to the barrier heights calculated for molecules **3** and **3′**.

Figure 5.13 (b) shows **4** in its low energy conformations with B and C falling in the syn/syn well and A from the syn/anti well. Hydrogen bonds between hydroxyls adjacent to the glycosidic linkages form in these conformations stabilizing the molecules. Conformations B and C, from the syn/syn well show stabilizing hydrogen bonds between O5-O4′ and O6-O4′. The secondary minimum snapshot, A, also shows these bonds between: O5-O2′ and O2-O4′. This is consistent to with the hydrogen bonds observed in the other pyranosyl disaccharides **3** and **3′**.

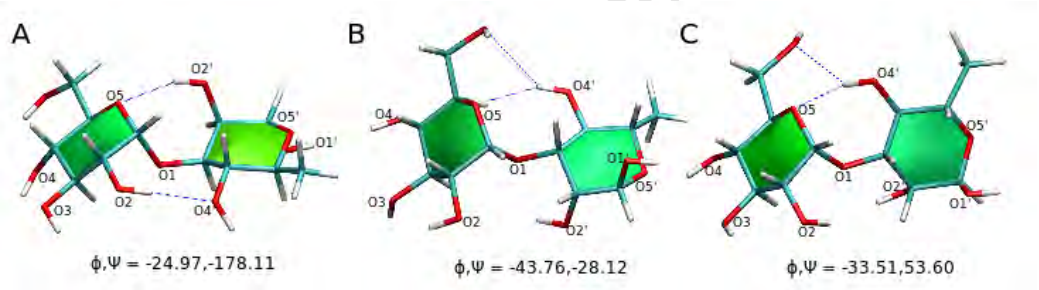


Figure 5.13: Examples of conformations of αDGlc$p$(1→3)αLRha$p$ corresponding to the labeled barriers of the PMF (Fig 5.12 (a))

Substituted disaccharide

The PMF plots [Figure 5.12 (b and c)] of the substituted molecules [αDGlc$p$(1→3) αLRha$p$ (1→3)DRibol5P (**4r**) and αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P (**4r′**)] shows a dramatic change in the energy landscape with the substitution of the ribitol-phosphate. While the anti/syn well has remained unchanged, what used to be a single syn/syn well in the unsubstituted disaccharide, has been split to four (B,C,D and E in 4r) and three (B,C, and D in 4r′) syn/syn wells. Thus the addition of the ribitol-phosphate has increased the conformational options along ϕ for the linkage. The PMFs demonstrate that has **4r** more freedom than **4r′**, with one extra syn/syn energy well (E) along the conformational space. In

addition to the extra energy well, **4r** also has lower energy barriers between wells, which suggest that the (1→3) substituted disaccharide is more flexible than the corresponding (1→4) disaccharide. The global minimum of the two molecules also differ with **4r** at 43.75°,-3.75°, and **4r′** at -43.75°,-178.75° (Table 5.7).

	Conformational Region			
	syn-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	syn-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/syn-ψ $\phi\psi$,(ΔG),PMF	anti-ϕ/anti-ψ $\phi\psi$,(ΔG),PMF
αD-Glc*p*(1→3)αL-Rha*p* (1→3)D-Ribol5P 4r (6A,6C)	43.75 -3.75 (0.00), C -23.75 -13.75 (1.53),B -73.75 -41.25 (1.76),E 68.75 56.28 (2.71),D	-38.75 173.75 (1.30), A	91.25 36.25 (3.66)	-91.25 -101.25 (8.74)
αD-Glc*p*(1→3)αL-Rha*p* (1→4)D-Ribol5P 4r′ (6B,6D)	41.25 8.75 (1.19), C -8.75 -33.75 (1.83),B 88.75 38.75 (1.53),D	-43.75 -178.75 (0.00), A	91.25 38.75 (2.92)	-91.25 -96.25 (8.57)

Table 5.7: Minima in energy wells of αD-Glc*p*(1→3)αL-Rha*p*(1→3)D-Ribol5P and αD-Glc*p*(1→3)αL-Rha*p*(1→4)D-Ribol5P. ϕ/ψ (°), ΔG (kcal/mol), corresponding region on PMF in Figure 5.12

The ribitol-phosphate forms stabilizing hydrogen bonds in the disaccharides in all the conformational regions (Figure 5.14). The conformations of these molecules are held in place not just by hydrogen bonds between the glucose and rhamnose residues, but with the ribitol-phosphate as well. The trend of hydroxyls adjacent to the glycosidic linkage forming stabilizing hydrogen bonds (between O2/O5 and O2′/O4′) is somewhat conserved, however the snapshots in Figure 5.14, shows the flexible ribitol-phosphate orientates its phosphate end causing hydrogen bonds to form with the phosphate oxygens instead.

It is worth noting, that the preference for the formation of hydrogen bonds between the ribitol-phosphate and hydroxyls adjacent to the glycosidic linkage, seen in the snapshots in Figure 5.14, is likely to be greatly reduced as the phosphate motif responsible for these hydrogen bonds is linked via a (O→2) linkage to a galactose or glucose residue. The linkage to a glucose or galactose would increase the constraints on this motif, reducing its ability to form the hydrogen bonds. In addition to this, the doubly linked ribitol (through a rhamnose (1→3) or (1→4) linkage and a (PO→2) linkage to a galactose or glucose) is likely to severely reduce the internal flexibility of the ribitol (along the C3/C4 to PO_4 backbone).

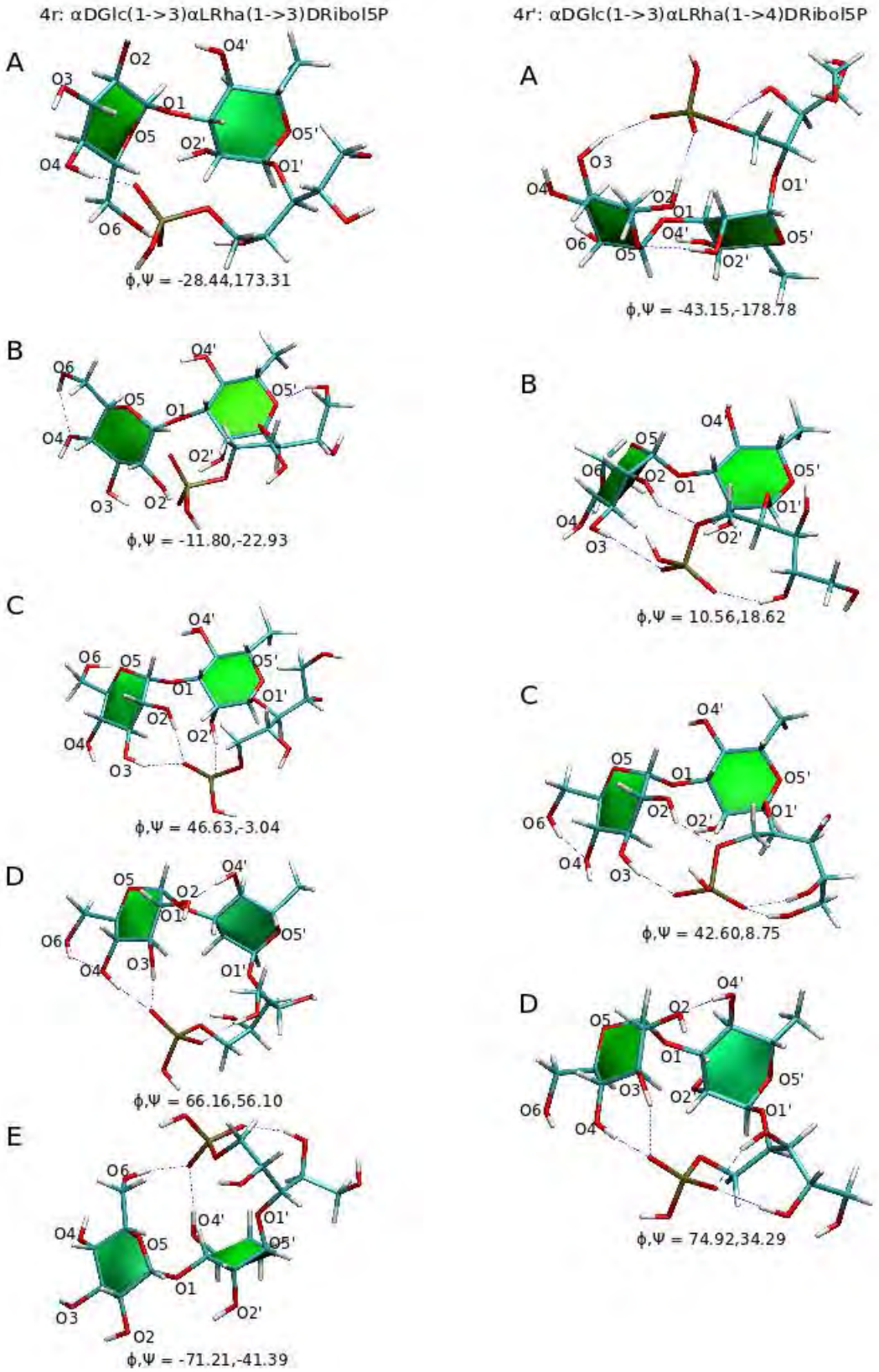


Figure 5.14: Examples of conformations of αD-Glc$p$(1→3)αL-Rha$p$(1→3)D-Ribol5P and αD-Glc$p$(1→3)αL-Rha$p$(1→4)D-Ribol5P corresponding to the labeled wells of the PMFs (Fig 5.12 (b and c))

5.3 Discussion

Serogroup 6 oligosaccharides were divided into disaccharide components in two ways: firstly treating ribitol-phosphate as a standard residue and secondly treating it as a substituent onto another residue (Figure 5.1):

1: DRibol5P(O→2)αDGal$p$ (6A and 6B)
1′: DRibol5P(O→2)αDGlc$p$ (6C and 6D)

2: αLRha$p$(1→3)DRibol5P (6A and 6C)
2′: αLRha$p$(1→4)DRibol5P (6B and 6D)

3: αDGal$p$(1→3)αDGlc$p$ (6A and 6B)
3′: αDGlc$p$(1→3)αDGlc$p$ (6C and 6D)

4: αDGlc$p$(1→3)αLRha$p$ (6A, 6B, 6C and 6D)

and

r3: DRibol5P(O→2)αDGal$p$(1→3)αDGlc$p$ (6A and 6B)
r3′: DRibol5P(O→2)αDGlc$p$(1→3)αDGlc$p$ (6C and 6D)

4r: αDGlc$p$(1→3)αLRha$p$(1→3)DRibol5P (6A and 6C)
4r′: αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P (6B and 6D)

The PMF plots of the disaccharides are in agreement with the exo-anomeric effect which states that the ϕ angle for the non-reducing sugar α-linked disaccharides (**2**, **2′**, **3**, **3′** and **4**) is approximately -60° or 180°.[47,94] The serotype points of difference **1** and **1′**, **3** and **3′** as well as **r3** and **r3′** produce nearly identical PMF plots. This is to be expected with disaccarides **3** and **3′** as these disaccharides differ in the hydroxyl orientation at C4 of their glucose and galactose residues, a difference which is far removed from their glycosidic linkage and as a result has minimal effects. Even when the C4 hydroxyl is closer to the linkage (in disaccharides **1** and **1′** and **r3** and **r3′**) a minimal difference is seen in the PMFs. This

would suggest that the change between a glucose and a galactose residue would minimally alter the structure of the higher order oligosaccharides.

The results of disaccharide pairs **2** and **2′** and **4r** and **4r′** exhibit the greatest point of difference in these serotypes. Comparing disaccharides **2** and **2′**, the αLRha$p$(1→3)DRibol5P linkage has more conformational options, with a low energy anti/syn region, than its (1→4) counterpart. When ribitol-phosphate was added as a substituent to the αDGlc$p$(1→3)αLRha$p$ disaccharide (**4r** and **4r′**), the disaccharide with the (1→3) linked substituent exhibited greater conformational freedom with an extra syn/syn well than the molecule with the (1→4) linked substituent. The greater conformational freedom observed for the (1→3) ribitol substituted disaccharide, can be explained by the longer ribitol backbone (C3 to PO_4 as opposed to C4 to PO_4 in a (1→4) linkage). The longer (1→3) linked ribitol backbone in rhamnose-ribitol, affords more rotational freedom to position the phosphate motif to form stabilizing hydrogen bonds with hydroxyls adjacent to the glucose-rhamnose glycosidic linkage than the (1→4) substituent linkage would.

The results show that the biggest point of difference between the serotypes is between: αLRha$p$(1→3)DRibol5P (6A and 6C) and αLRha$p$(1→4)DRibol5P (6B and 6D). There is minimal difference seen between the αDGal$p$ (6A and 6B) and αDGlc$p$ (6C and 6D) residues. Ribitol-phosphate is a flexible residue which has the ability to form a variety of stabilizing interresidue hydrogen bonds which change the energy landscapes of the disaccharides. However in an oligosaccharide, with ribitol bonded to both a rhamnose and a glucose/galactose residue, it is expected that this flexibility will be greatly reduced. From this study of the glycosidic linkages, it is expected that the greatest structural differences within the serogroup will be observed between serotype pairs: 6A and 6D, and 6B and 6C.

Chapter 6

Oligosaccharide Structures of Serogroup 6

The previous chapter gave insight into the conformation of the component disaccharides of *S.pneumoniae* serogroup 6. This chapter extends these results to oligosaccharides, looking at the effect of interresidue interaction on the preferred conformation of the serogroup 6 oligosaccharides. The differences between the four serotypes are highlighted in the schematic in Figure 6.1: a galactose/glucose residue and a (1→3)/(1→4) linkage. The PMFs of all the linkage constituents of serogroup 6 allowed identification of the global minimum conformations which were used to build likely starting structures for the 6A-D oligosaccharides.

6.1 Building of the Oligosaccharides

Two oligosaccharide structures were built for each serotype: the first using dihedral values from the calculations for unsubstituted disaccharides for all the glycosidic linkages and the second using values for the DRibol5P substituted disaccharides (Figures 6.2 - 6.5). Each of the oligosaccharides structures consisted of three repeating units (12 residues). Oligosaccharides built with the unsubstituted dihedral values are henceforth labeled 6A - 6D, while those built with DRibol5P substituted dihedral values are primed (6A′ - 6D′).

In most cases, the disaccharides were set to their global minimum ϕ/ψ values,

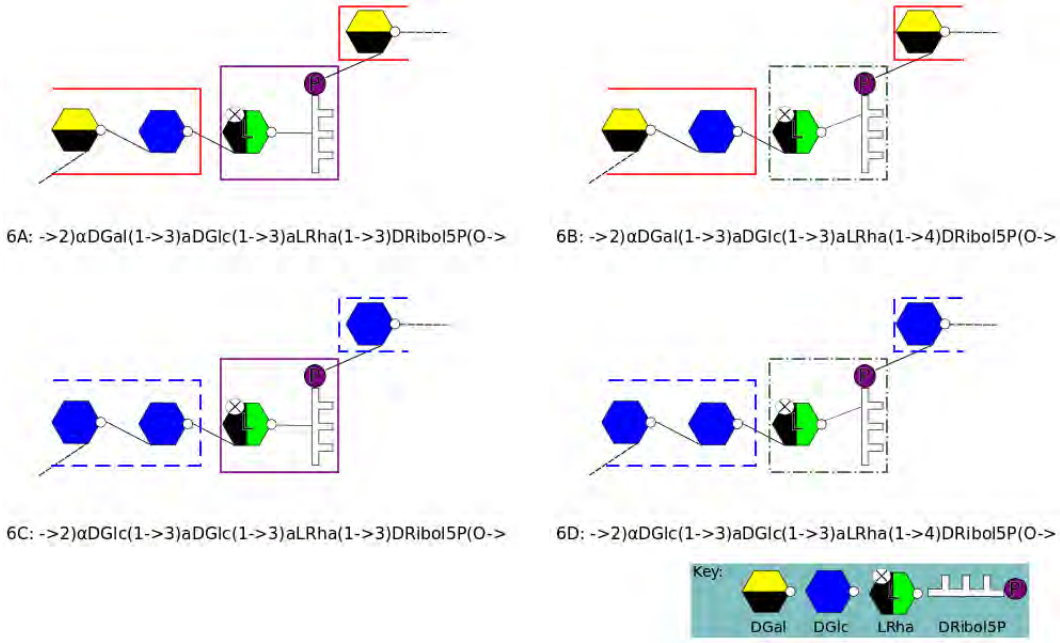


Figure 6.1: Schematic of serogroup 6, highlighting the differences between the four serotypes. The red boxes indicate a galactose residue, blue indicates a glucose residue, purple indicate a (1→3) linkage and green indicate a (1→4) linkage

as calculated from the PMFs in Chapter 5. However, for 6A′ and 6C′ this choice of dihedrals leads to self-intersecting molecular structures. Therefore, in these cases, the flexible ribitol-phosphate-galactose (6A′) or ribitol-phosphate-glucose (6C′) dihedral angles were rotated to the secondary low energy well in the linkage, as shown in Figures 6.2 and 6.4. This ribitol-phosphate-galactose/glucose dihedral shows the greatest distribution of equivalent low energy wells across the ϕ/ψ conformational space. The DRibol5P(O→2)αDGal$p$/αDGlc$p$ dihedrals in 6A′ and 6C′ were set to the lowest energy well that resolved the self-intersection: well D for 6A′ and well C for 6C′.

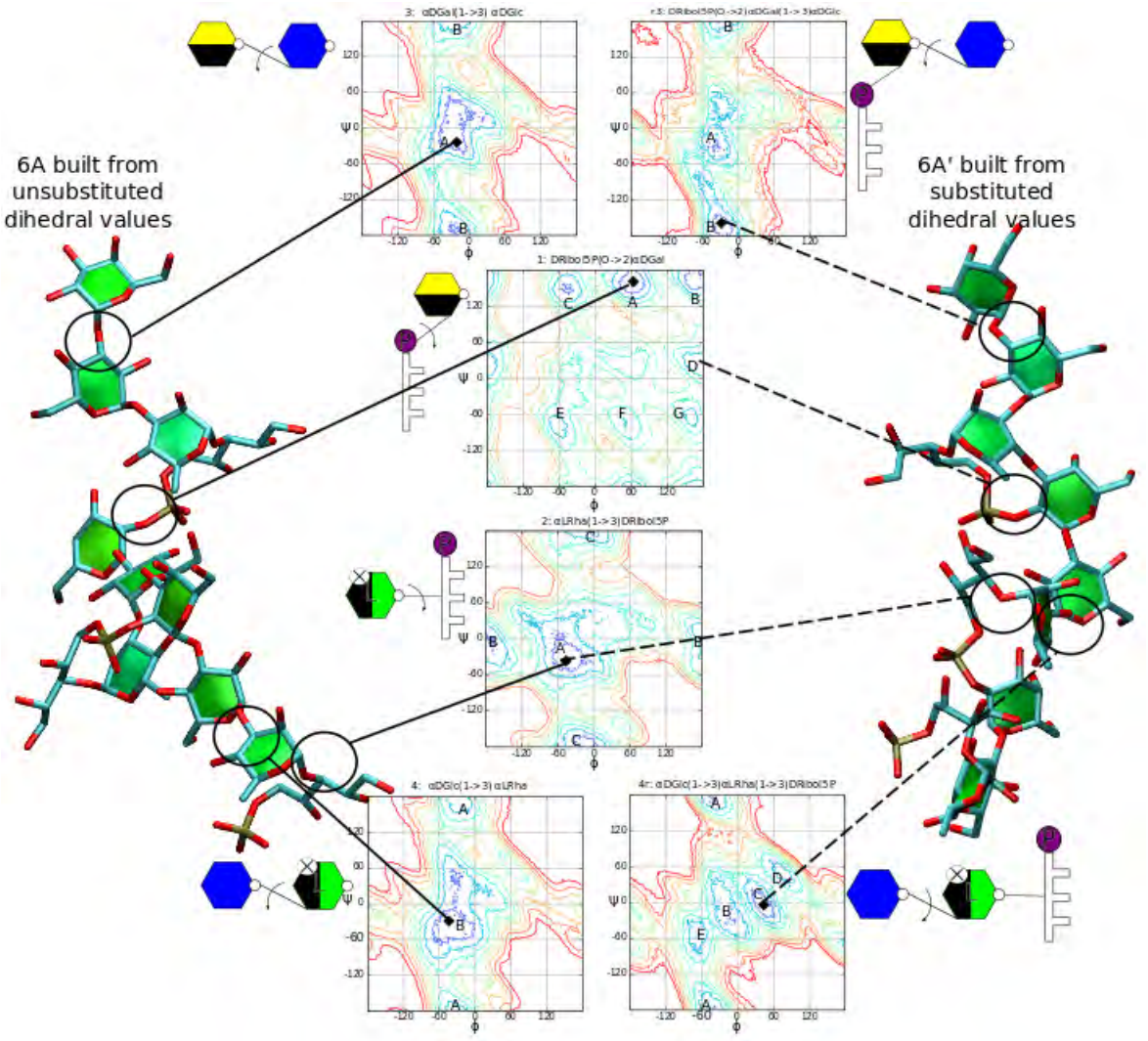


Figure 6.2: Three repeating unit 6A ([→2)αDGal*p*(1→3)αDGlc*p*(1→3)αLRha*p*(1→3)DRibol5P(O→]) oligosaccharides built from disaccharide conformations. The global minimum is indicated by the black diamonds. Note that for 6A′, the DRibol5P(O→2)αDGal*p* linkage was set to the secondary minimum D.

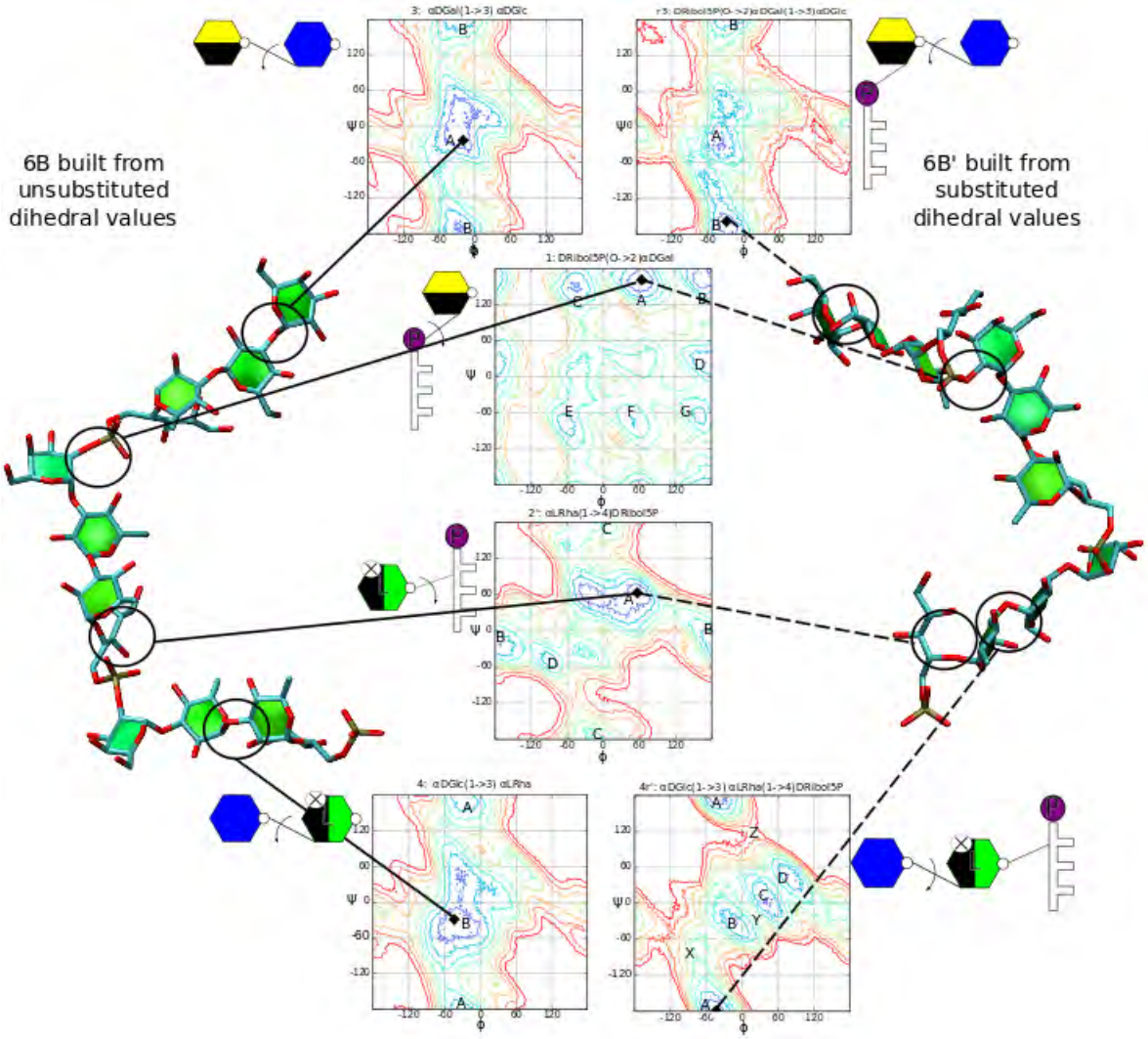


Figure 6.3: Three repeating unit 6B ([→2)α-D-Gal*p*(1→3)α-D-Glc*p*(1→3)α-L-Rha*p*(1→4)-D-Ribol5P(O→]) oligosaccharides built from disaccharide conformations. The global minimum is indicated by the black diamonds.

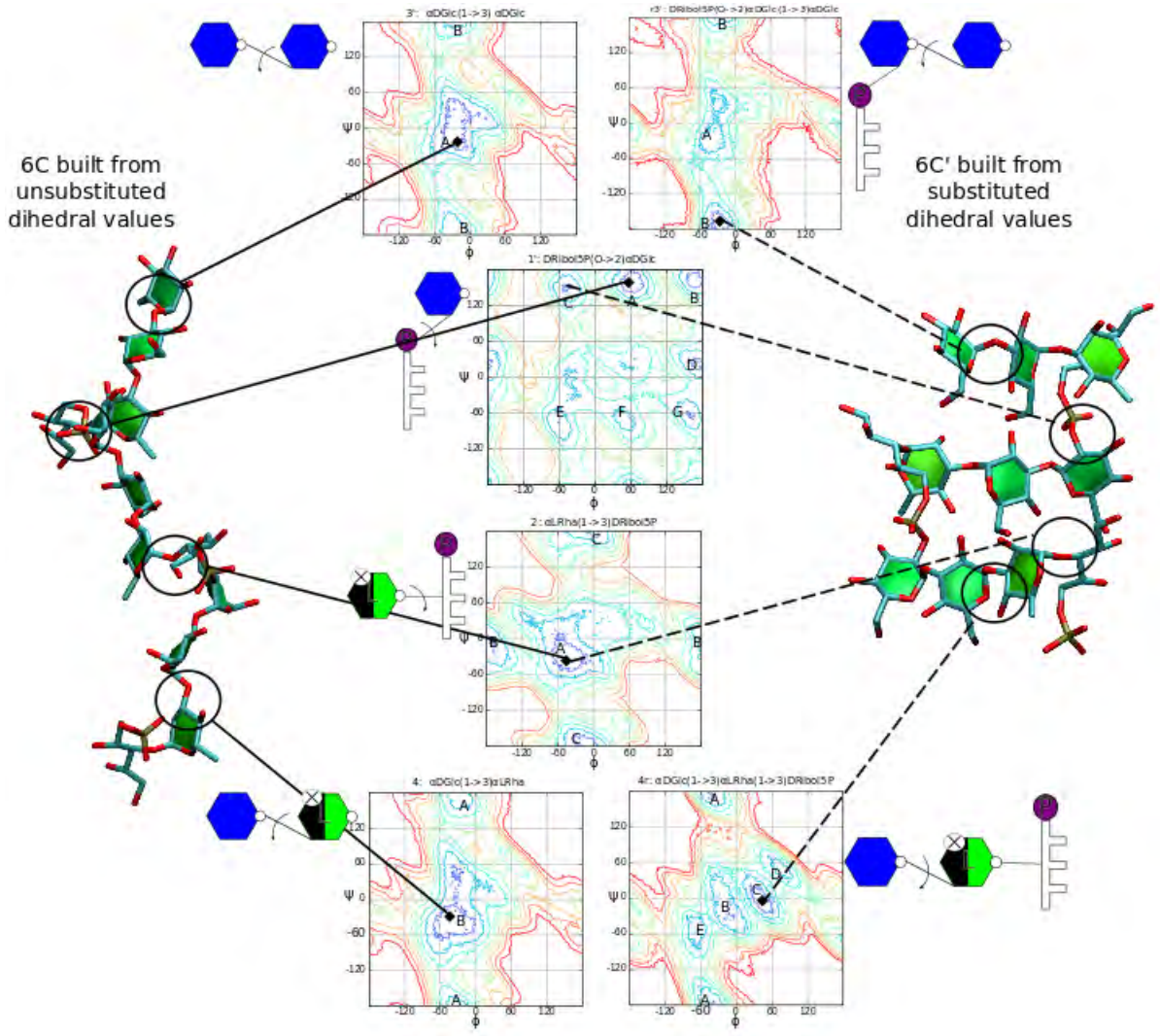


Figure 6.4: Three repeating unit 6C ([→2)αDGlc$p$(1→3)αDGlc$p$(1→3)αLRha$p$(1→3)DRibol5P(O→]) oligosaccharides built from disaccharide conformations. The global minimum is indicated by the black diamonds. Note that for 6C′, the DRibol5P(O→2)αDGlc$p$ linkage was set to the secondary minimum C.

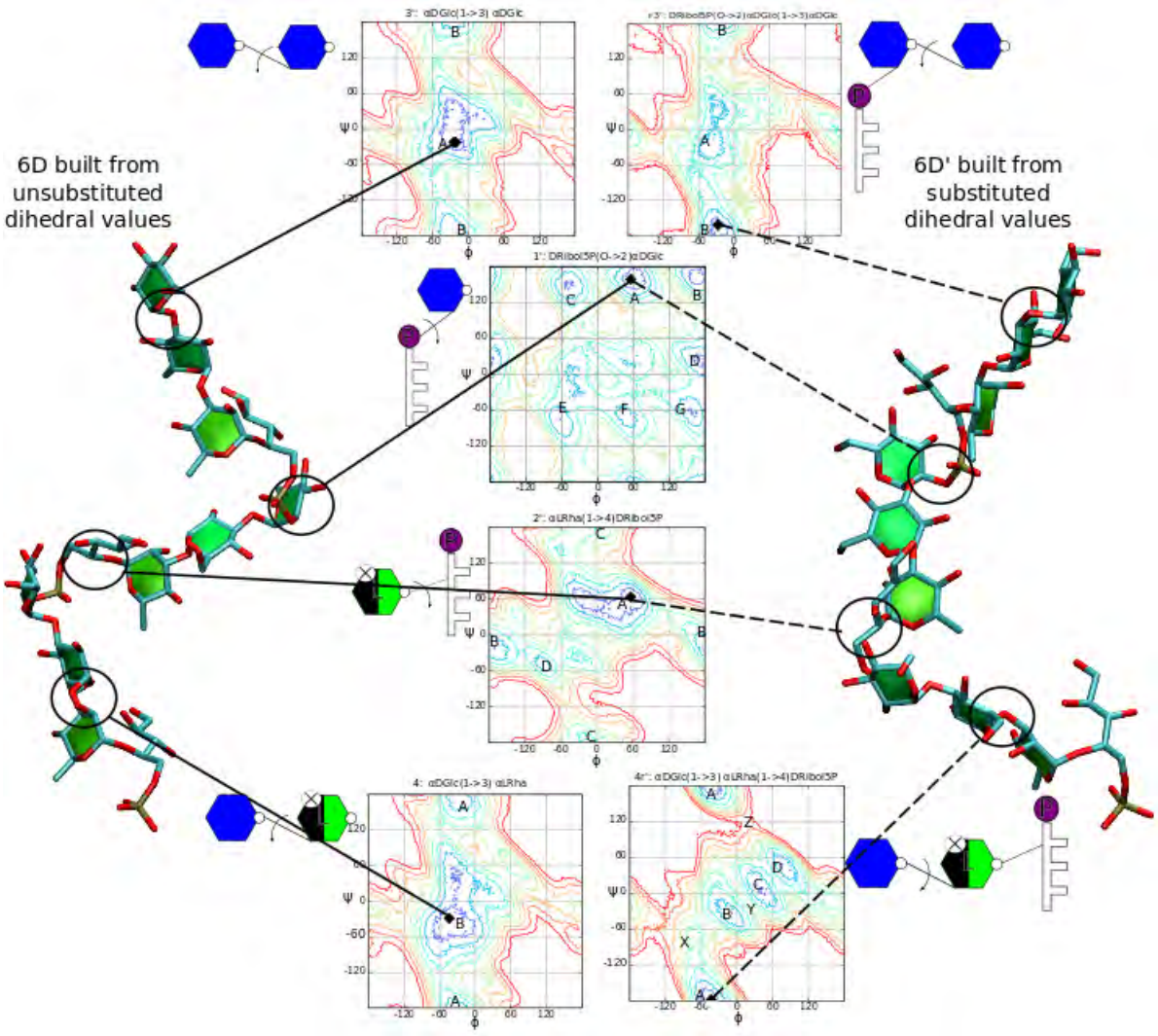


Figure 6.5: 3 repeating unit 6D ([→2)αD-Glc$p$(1→3)αD-Glc$p$(1→3)αL-Rha$p$(1→4)D-Ribol5P(O→]) oligosaccharides built from disaccharide conformations. The global minimum is indicated by the black diamonds.

6.2 Serotype 6A

Figure 6.6 details the glycosidic linkage conformations for residues in the middle repeating unit of serotype 6A ([→2)αDGal$p$(1→3)αDGlc$p$(1→3)αLRha$p$(1→3) DRibol5P(O→]) throughout the course of the 500 ns simulations. The two simulations started from 6A and 6A′ conformations, produced similar equilibrium distributions, indicating that the simulations have reached convergence.

6A′ populated the same conformational regions as 6A (Figure 6.6). The glycosidic linkages exclusively occupy the low energy regions of the PMF, with the flexible ribitol-phosphate linkages showing the greatest distribution (but remaining within the low energy wells calculated on the PMFs). This indicates that the PMF for the glycosidic linkages is a fairly reliable measure of the possible conformations of an oligosaccharide; the intermolecular interactions in the oligosaccharides do not make dramatic alterations. It is interesting to note that the A,B,C wells in linkage **1** (Figure 6.6) are not occupied in either simulation despite 6A starting from the A conformation. Rather, this linkage populates the D well (the conformation to which the linkage in 6A′ was set to avoid self intersection) and is distributed along ϕ from well D. Just as with the **3** linkage in 6A, the **r3** linkage of 6A′ also occupies the low energy regions of the calculated PMF. The low energy regions of the PMFs of **3** and **r3** are similar and as such the results are as expected. Likewise, the **4r** linkage in 6A′ occupies the similar region as **4** in 6A, that is the low energy regions of the unsubstituted linkage is occupied.

Conformational set:	1 (●) ϕ, ψ (°)	2 (■) ϕ, ψ (°)
αDGal$p$(1→3)αDGlc$p$	-27, -169	-29, 43
αDGlc$p$(1→3)αLRha$p$	-37, -27	-35, -21
αLRha$p$(1→3)DRibol5P	35, 1	33, 33
DRibol5P(O→2)αDGal$p$	-157, -35	27, -31

Table 6.1: Dihedral values used to build 20mers in Figure 6.7; values of the glycosidic linkages between the two structures are coloured in blue when they are located in different regions.

It is clear from the trajectories in Figure 6.6, that galactose linkages **1** and **3** have two different conformational preferences. Investigation of the trajectories revealed that these disaccharide conformations are paired, resulting in two distinct

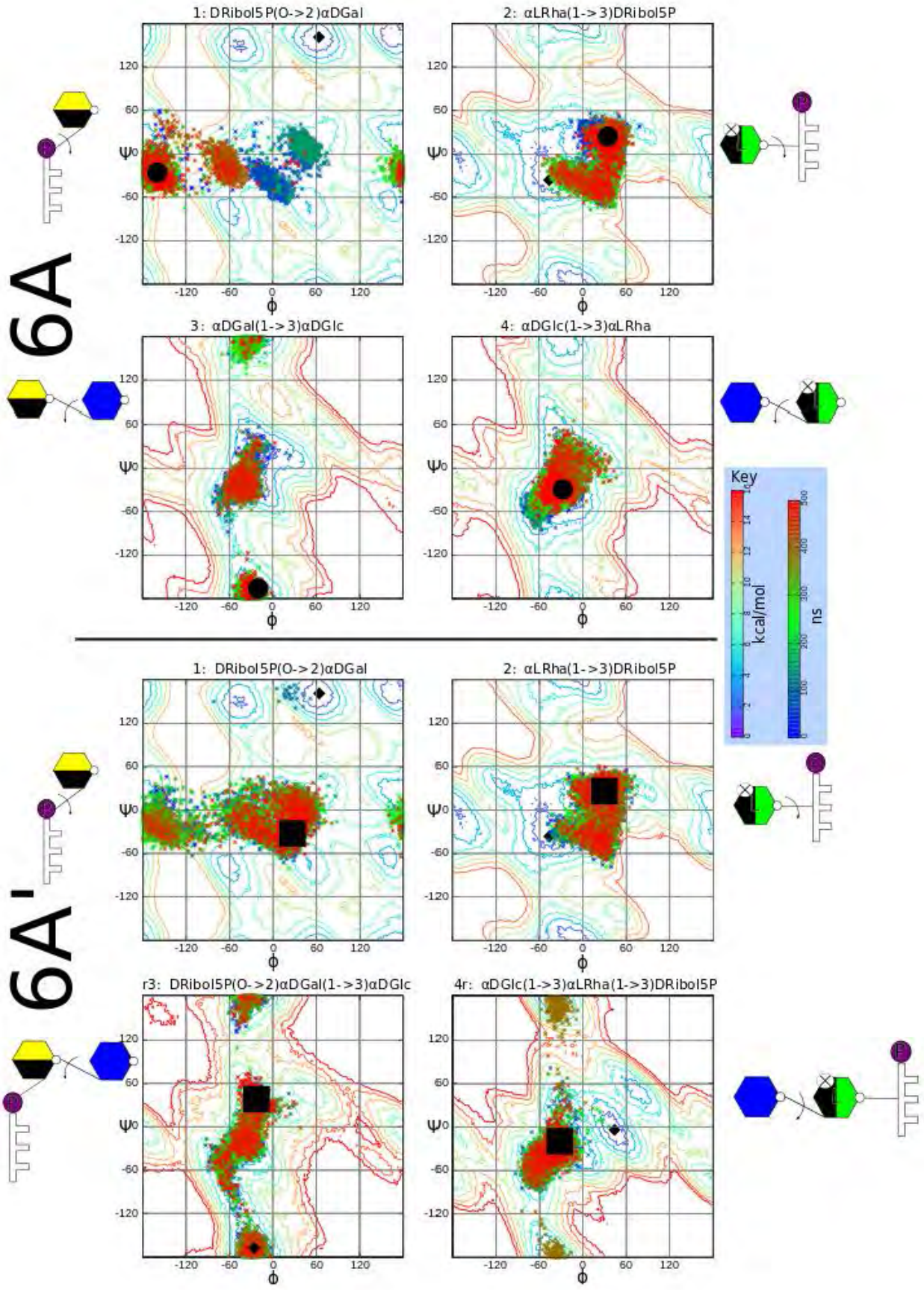


Figure 6.6: Progression of glycosidic linkages of the middle repeating unit of serotype 6A and 6A′. ● and ■ indicate the dihedrals of the two prevalent conformations of 6A.

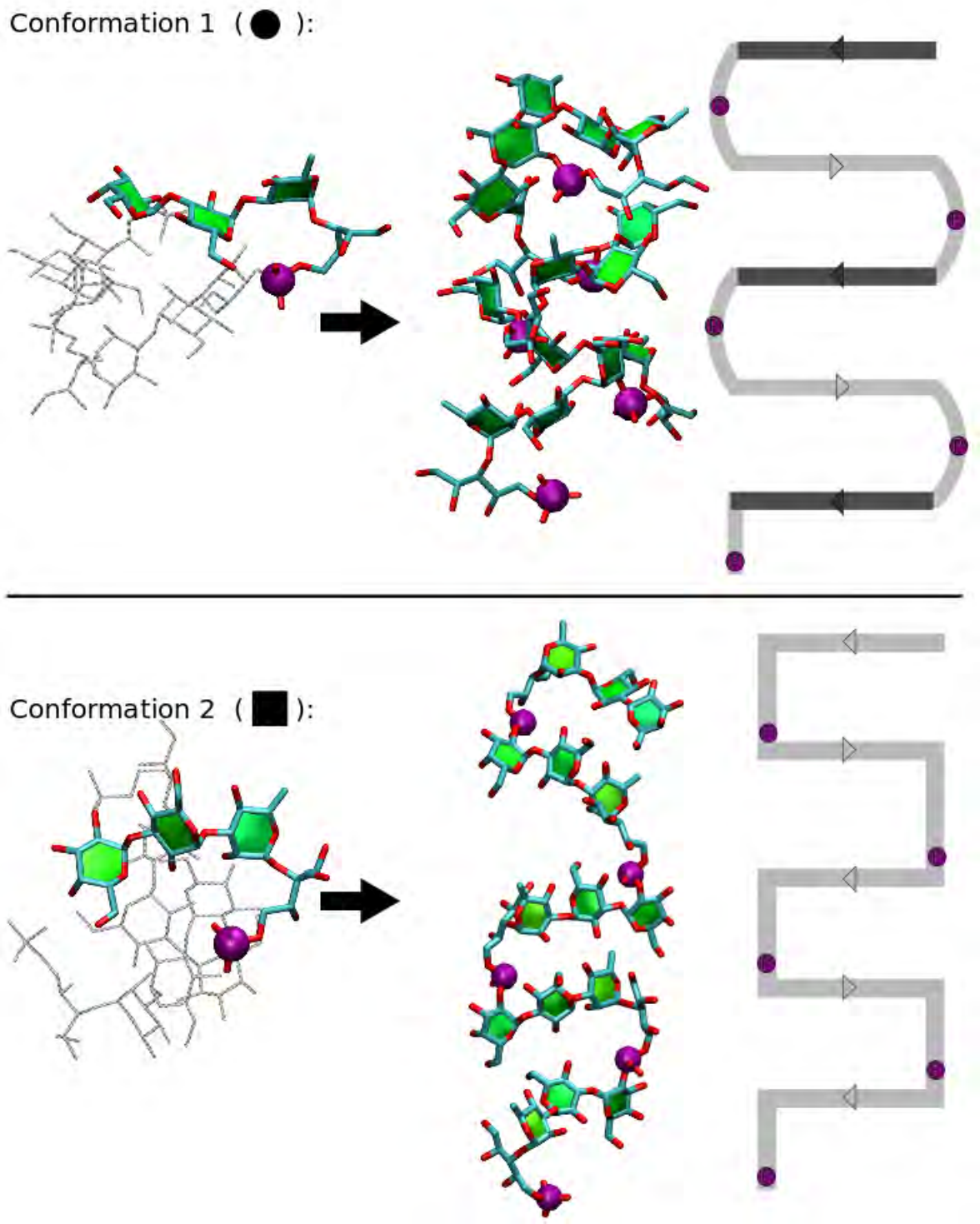


Figure 6.7: Snapshots of the two most prevalent conformations of the last 100 ns of the MD simulations of 6A (the ● and ■ regions on Figure 6.6), and the two possible 20mer structures of serotype 6A, built from the conformations. The darker segments indicate layers which are in the "front" plane while the lighter structures lie in the "back" plane.

prevalent structures for the 6A oligosaccharide: the ■ conformation with linkage **3** in a syn/syn conformation is paired with linkage **1** in a syn/syn position, and the ● conformation with linkage **3** in a syn/anti conformation is paired with linkage **1** in an anti/syn position. The dihedral values of the two resulting conformations are labeled on the PMF plots with ●s and ■s (Figure 6.6 and Table 6.1). From these most prevalent conformations from the last 100 ns two 20mer (5 repeating units) of serotype 6A were built and minimized and are shown in Figure 6.7 along with the corresponding ● and ■ 6A 20mers.

Conformations ● and ■ both represent stacked structures with the pyran residues of each repeating unit layered upon each other and with the ribitol-phosphate acting as a flexible joint between each of the layers. In the ● conformation, the resulting structure is more cramped, as the layers are not co-planar, but alternate between a front and back layer. The layers of the ■ structure are co-planar, with every other phosphate aligned on either side of the stack. Conformation ■ adopts a ladder like structure with each pyran layer as a rung. With both structures, the direction (non-reducing end to reducing end) of the pyran residues of each repeating unit switches every layer. This is indicated with the arrows on the schematic in Figure 6.7. For both conformations, the phosphates are on the outer edges of the structure making these charged motifs highly exposed.

6.3 Serotype 6B

Figure 6.8 shows plots for 6B ([→2)αDGal$p$(1→3)αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P(O→]) detailing the glycosidic linkages of the middle repeating unit of 6B serotypes through out the 500ns MD simulation. The two trajectories show that both structures converge, with their linkages occupying similar dihedral regions at the end of the simulation.

The linkages **3** and **r3** settle in similar conformational regions which sit in the primary energy well of both the PMFs. While the **4r′** linkage (6B′) converged away from the minimum on the PMF, the region of convergence is similar to that of linkage **4** which falls within the primary energy well. The ribitol linkages **1** and **2′** both converge to similar regions in both 6B and 6B′ starting conformations, however they have shifted away from their primary low energy well.

The higher concentration of the red points on the PMF plots of 6B′ shows that the oligosaccharides converged faster than the 6B oligosaccharide. While both oligosaccharides eventually converge to same regions (with linkage **1** and **3** in syn/syn conformations- labeled ■ in Figure 6.8) , it is clear from the trajectories, that an alternate configuration exists with linkage **1** in the anti/syn conformation and linkage **3** in a syn/anti conformation (labeled ● in Figure 6.8).

Two 20mer (5 repeating units) of serotype 6B were built and minimized with these two prevalent conformations (Table 6.2) and the structures are shown in Figure 6.9.

Conformational set:	1 (●) ϕ, ψ (°)	2 (■) ϕ, ψ (°)
αDGal*p*(1→3)αDGlc*p*	-21, 31	-33,-19
αDGlc*p*(1→3)αLRha*p*	-45,-39	-71,-65
αLRha*p*(1→4)DRibol5P	1,-41	25,-47
DRibol5P(O→2)αDGal*p*	41,1	41,1

Table 6.2: Dihedral values used to build 20mers shown in Figure 6.9; values of the glycosidic linkages between the two structures are coloured in blue when they are located in different regions.

The 6B ● conformation shows a ladder like structure, however with the ribitol-phosphates acting as rungs. However, the ■ conformation shows a somewhat helical structure interrupted by the linear ribitol-phosphates where the phosphates are almost aligned vertically and the repeating units projecting in opposite directions. The direction (non-reducing to reducing end) of the pyran residues of the repeating units for both conformations is maintained along a plane. Unlike 6A, the charged phosphates in 6B are in the centre of the structure making them more sheltered than the phosphates in 6A.

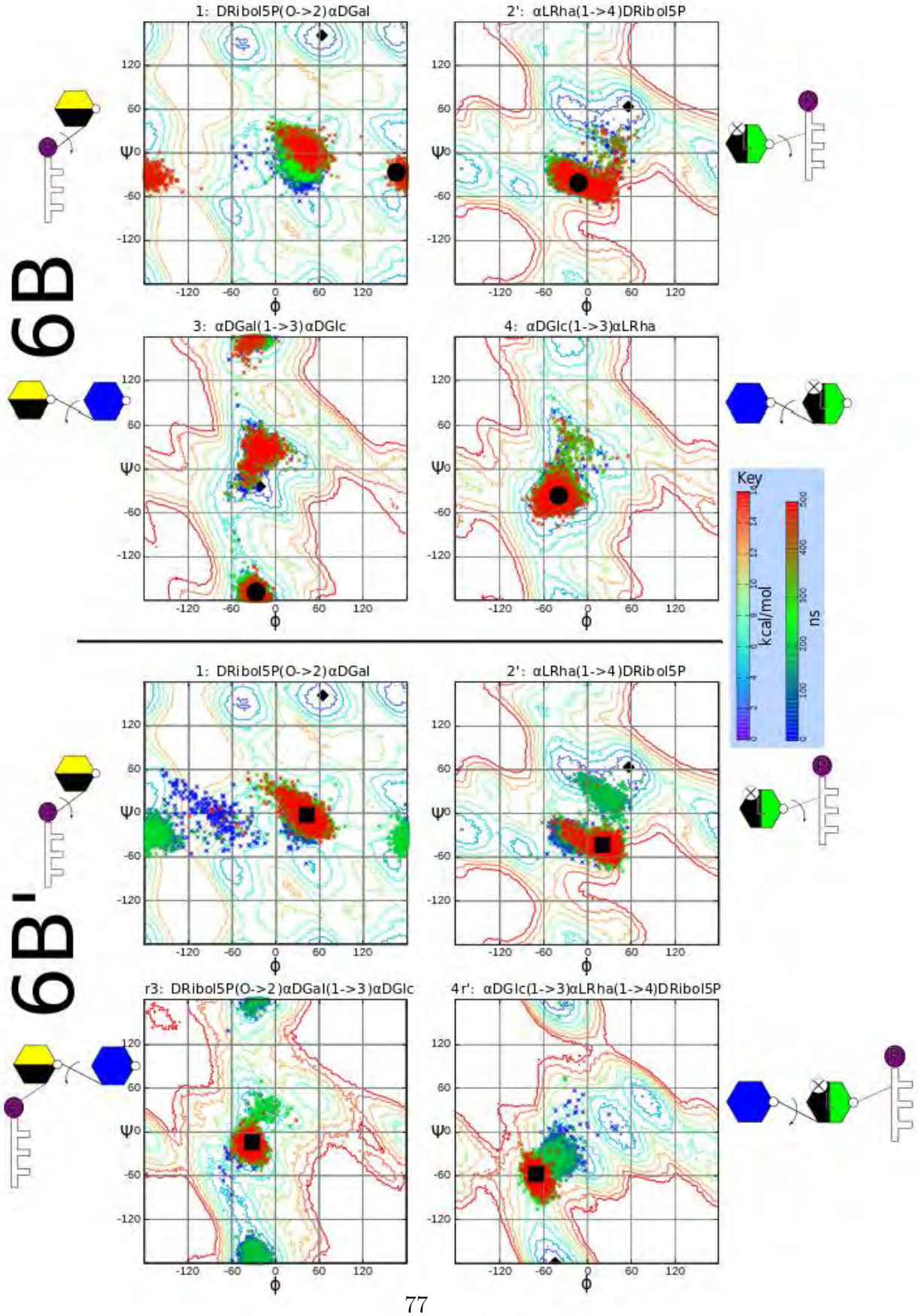


Figure 6.8: Progression of glycosidic linkages of the middle repeating unit of serotype 6B and 6B′. ● and ■ indicate the dihedrals of the two prevalent conformations of 6B.

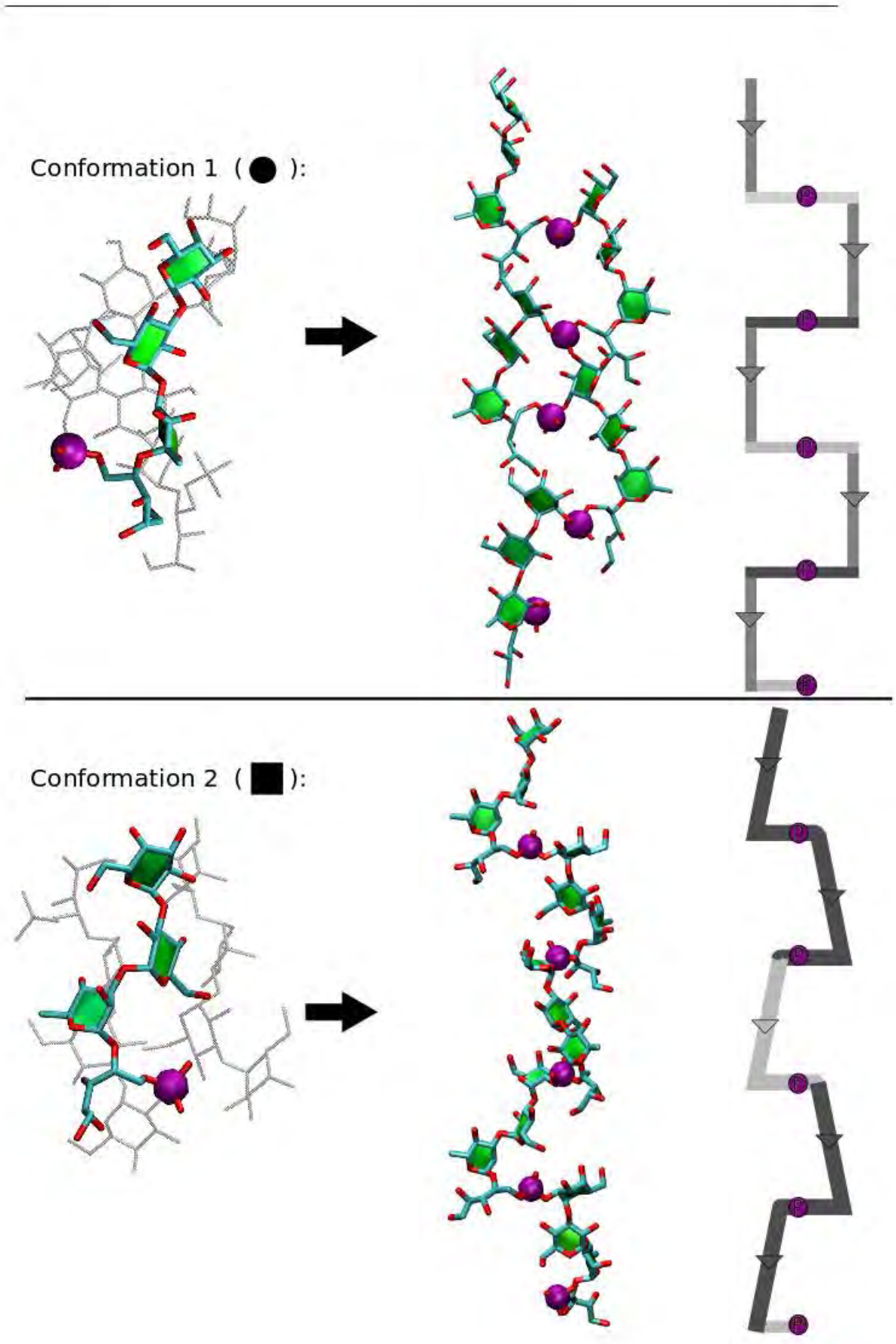


Figure 6.9: Two possible 20mer structures of serotype 6B, built from the most prevalent conformations of the last 100 ns of the MD simulations of 6B and 6B′ (the ● and ■ regions on Figure 6.8). The darker segments indicate layers which are in the "front" plane while the lighter structures lie in the "back" plane.

6.4 Serotype 6C

The glycosidic linkages of the middle repeating unit of 6C ([→2)αDGlc$p$(1→3)αDGlc$p$(1→3)αLRha$p$(1→3)DRibol5P(O→]) from both starting conformations 6C and 6C′ explore the same regions during the course of the simulation. However, not all of the corresponding linkages converge to the same regions (Figure 6.10).

Linkages **3′** (in 6C) and **r3′** (in 6C′) both explored the same ϕ/ψ regions, however they converged to different areas of the map. In the oligosaccharide built from the unsubstituted disaccharide values (6C), linkage **3′** converged to the primary syn/syn low energy well on the PMF, the corresponding linkage in the 6C′ oligosaccharide converged to the syn/anti region on the PMF. Linkages **4** and **4r** explored the same region and converged to similar regions by the end of the simulation. The ribitol linkages (linkage **1′** and **2**) explored regions away from the low energy regions calculated in Chapter 5, similar to what is seen with serotype 6A and 6B. In both 6C and 6C′, linkage **2** converged to similar regions. Linkage **1′** in 6C converges to a syn/syn region, while the corresponding linkage in 6C′ converges to an anti/syn region.

Conformational set:	1 (●) ϕ, ψ (°)	2 (■) ϕ, ψ (°)
αDGlc$p$(1→3)αDGlc$p$	-27,-167	-51,-25
αDGlc$p$(1→3)αLRha$p$	-43,-39	-29,-13
αLRha$p$(1→3)DRibol5P	33,23	27,31
DRibol5P(O→2)αDGlc$p$	-157,1	1,-45

Table 6.3: Dihedral values of the two most prevalent conformations which were used to build 20mers in Figure 6.11; values of the glycosidic linkages between the two structures are coloured in blue when they are located in different regions.

Like 6A, the trajectories of 6C in Figure 6.10 show that glucose linkages **1** and **3** have two separate conformational pair preferences: linkage **3** in a syn/syn conformation is paired with linkage **1** in a syn/syn position (the ■ conformation), while linkage **3** in a syn/anti conformation is paired with linkage **1** in a anti/syn position (the ● conformation). The dihedral values of the the two resulting conformations are labeled on the PMF plots with ●s and ■s (Figure 6.10 and Table 6.3).

The 20mer oligosaccharides were built and minimized using the most preva-

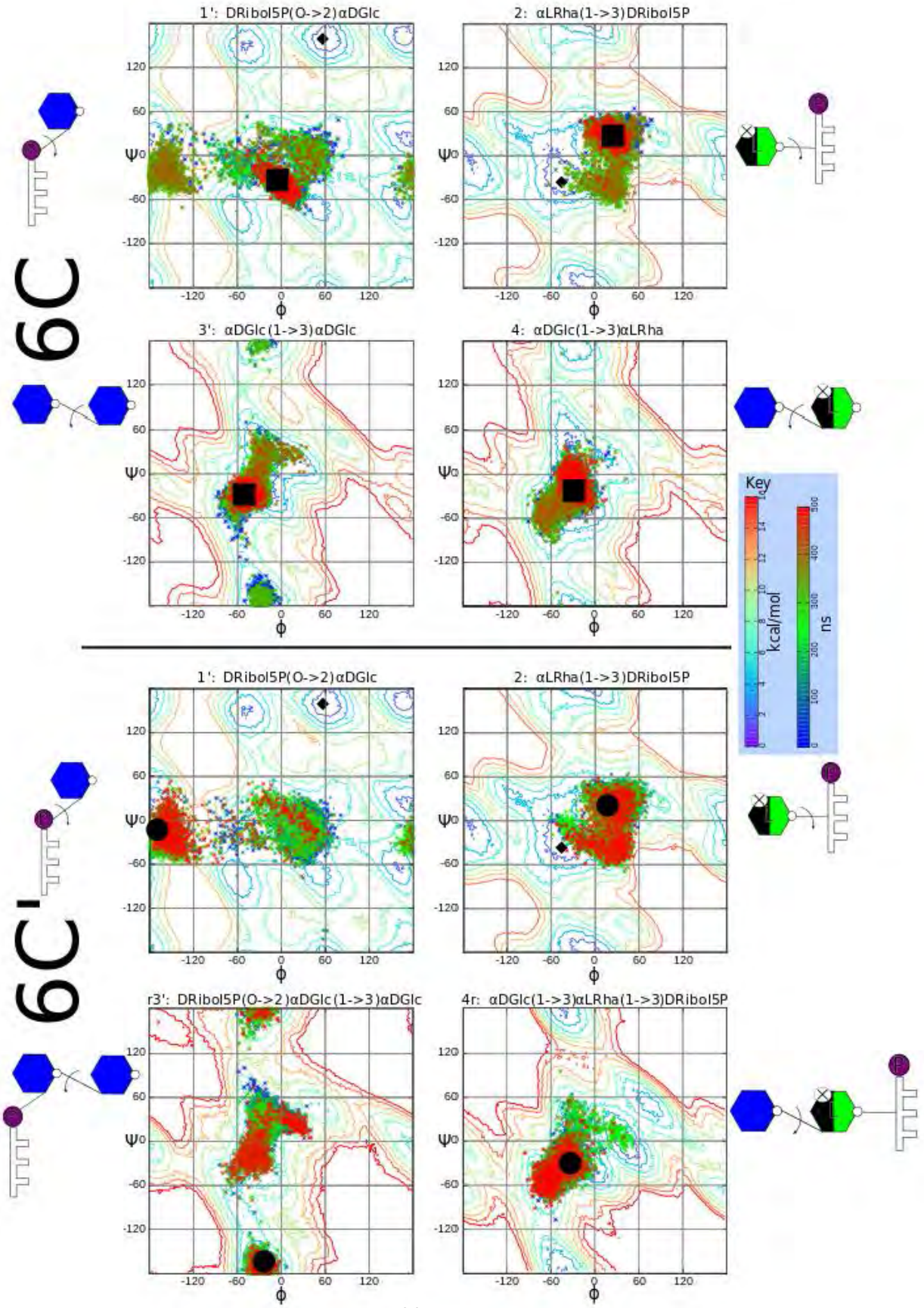


Figure 6.10: Progression of glycosidic linkages of the middle repeating unit of serotype 6C and 6C′. ● and ■ indicate the dihedrals of the two prevalent conformations of 6C.

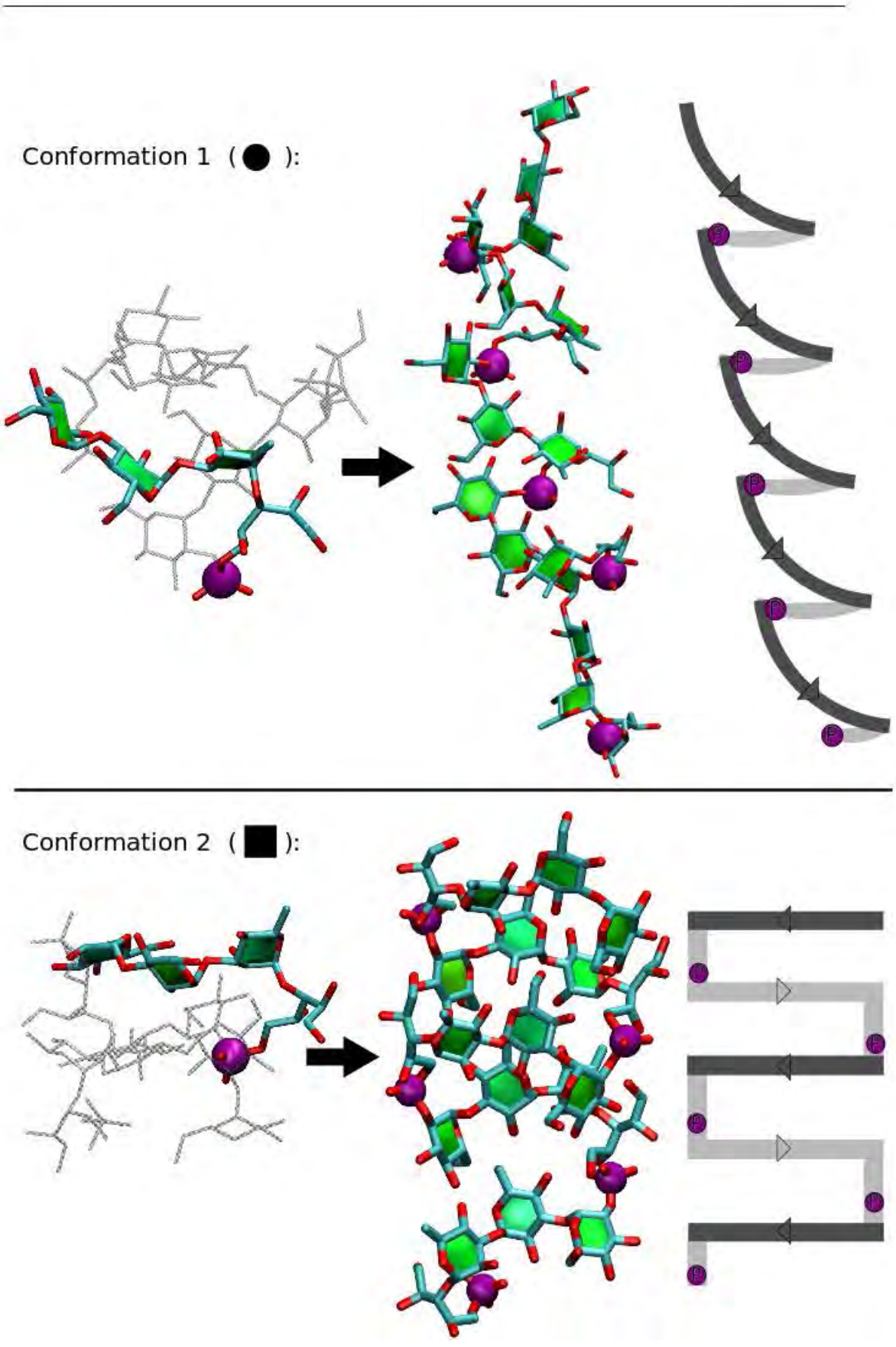


Figure 6.11: Snapshots of the two most prevalent conformations of the last 100 ns of the MD simulations of 6C (the ● and ■ regions on Figure 6.10), and the two possible 20mer structures of serotype 6A, built from the conformations. The darker segments indicate layers which are in the "front" plane while the lighter structures lie in the "back" plane.

lent conformations from the last 100 ns, the values of which are shown in Table 6.3. Figure 6.11 shows these structures; conformation ■ adopts a ladder stacked structure similar to a 6A structure, while the ● is slightly helical. Just as with the 6A structures, the 6C ■ structure has phosphates on the outside of the structure making the phosphates easily accessible. The 6C ● conformation has nearly helical structure with the phosphates in the middle of the helix. This is unusual for the (1→3) linked serotypes (6A and 6C) as the phosphates in this structure have now been sheltered.

6.5 Serotype 6D

Molecular dynamics was conducted on the final serotype 6D ([→2)αDGlc$p$(1→3)αDGlc$p$(1→3)αLRha$p$(1→4)DRibol5P(O→]) and the glycosidic linkages of the middle residue were plotted on the PMFs calculated in Chapter 5. Just as for the other three serotypes, both 6D and 6D′ starting conformations explored the same regions during the course of the simulation (Figure 6.12).

In 6D and 6D′, linkages **3′** and **r3′** both explored the same regions, however they converged to different areas of the map. For 6D, linkage **3′** converged to the primary syn/syn low energy well on the PMF, while the corresponding linkage in the 6D′ oligosaccharide converged to the syn/anti region on the PMF. On the other hand linkages **4** and **4r′** explored and converged to the same region by the end of the simulation. This region of convergence is removed from the low energy well in **4r′**, but they both lie in the syn/syn primary low energy well for the **4** glycosidic linkage. As usual, the ribitol linkages (linkage **1′** and **2′**) explored regions away from the low energy regions calculated in Chapter 5. In both 6D and 6D′, linkage **2′** converged to similar regions, which for the first time is the low energy well calculated in Chapter 5. Linkage **1′** in 6D converged to a syn/syn region, while the corresponding linkage in 6D′ converged to an anti/syn region.

Just as for the other serotypes the oligosaccharides have two prevalent conformations with linkage **1** and **3** adopting conformational pairs (Figure 6.12): syn/syn for linkage **1** with syn/syn for linkage **2** (■ conformation) and syn/anti for linkage **1** with anti/syn for linkage **2** (● conformation). The 20mer oligosaccharides were built using the most prevalent dihedral definitions from the middle

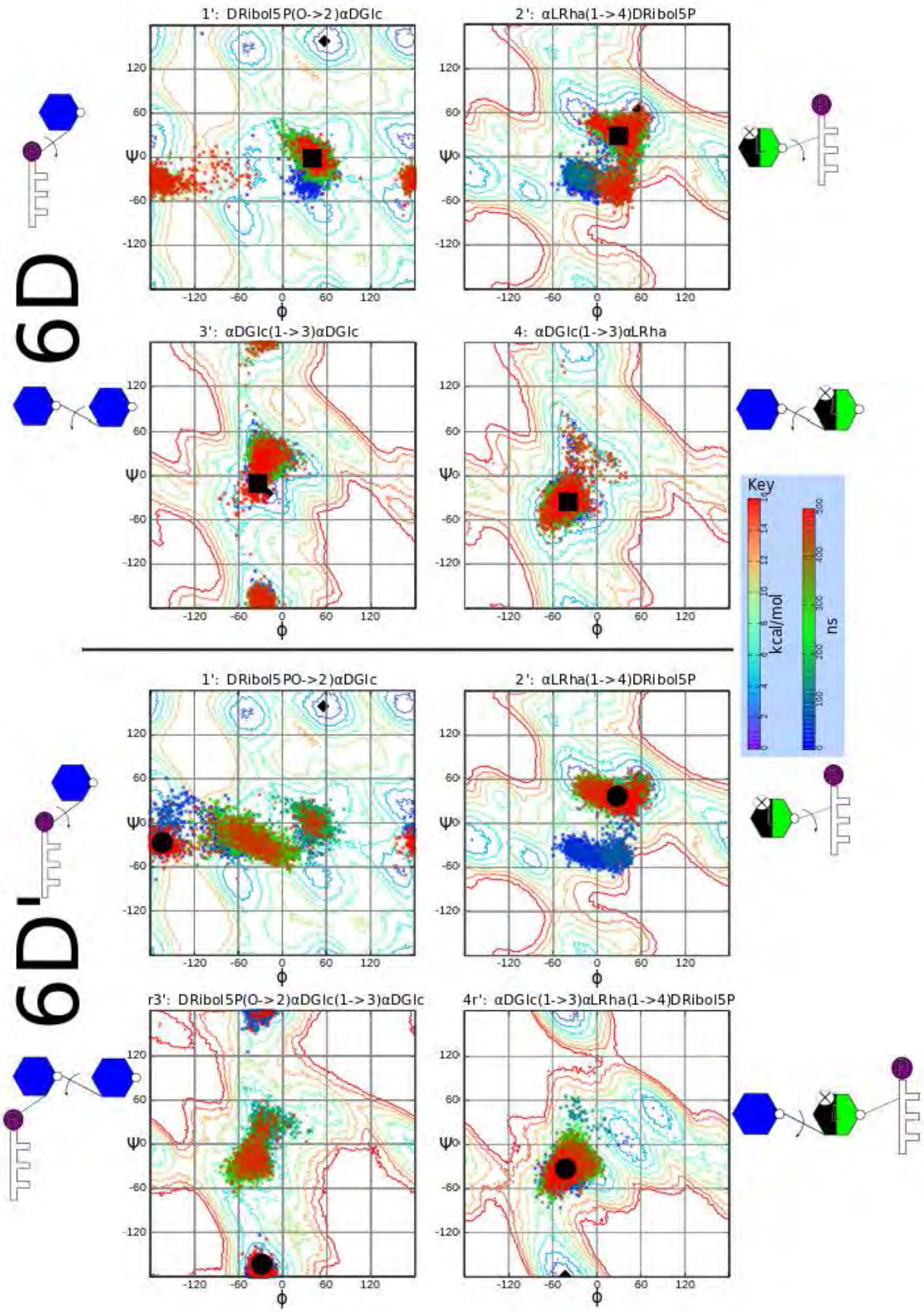


Figure 6.12: Progression of glycosidic linkages of the middle repeating unit of serotype 6C and 6C′. ● and ■ indicate the dihedrals of the two prevalent conformations of 6D.

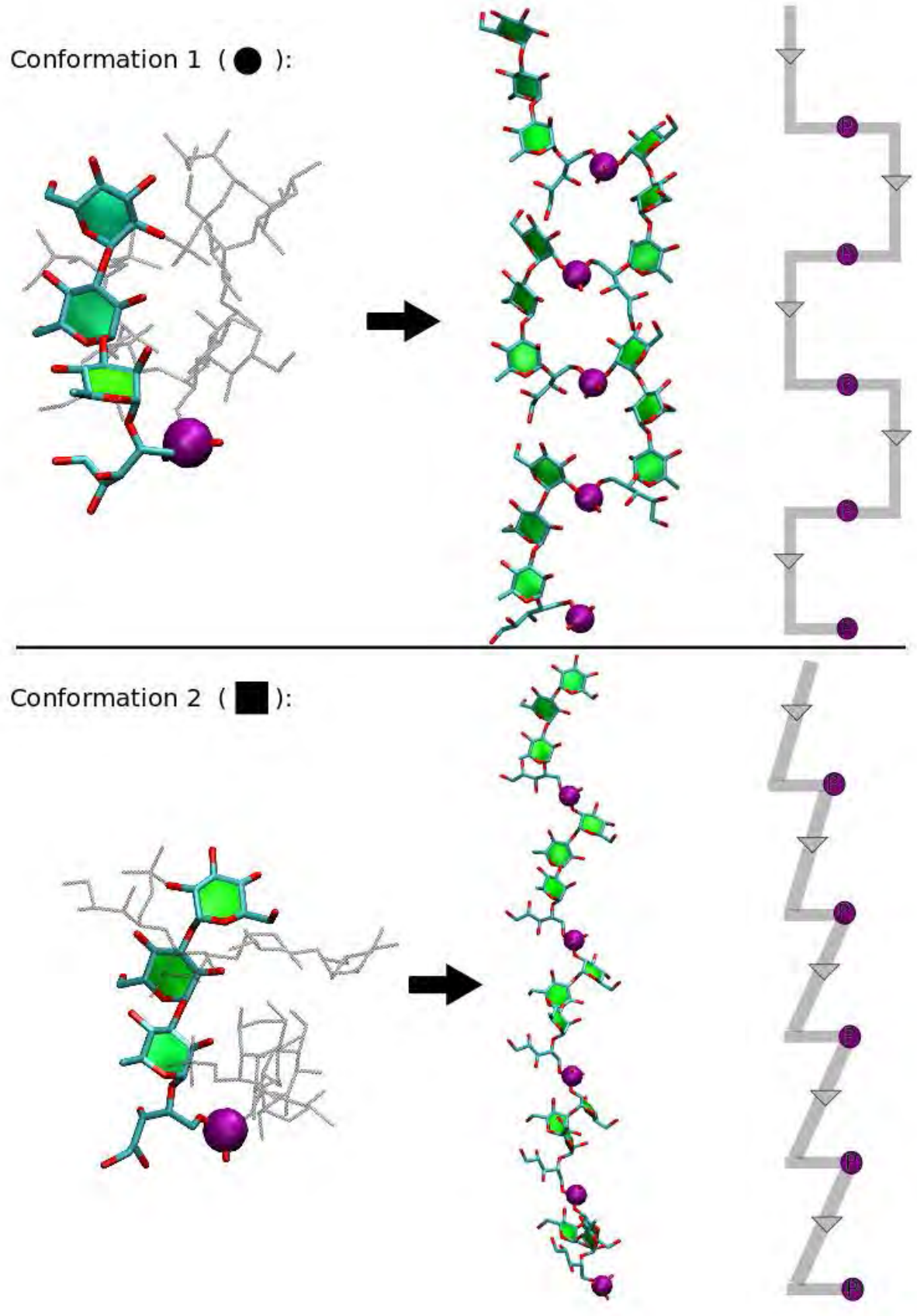


Figure 6.13: Snapshots of the two most prevalent conformations of the last 100 ns of the MD simulations of 6D (the ● and ■ regions on Figure 6.12), and the two possible 20mer structures of serotype 6A, built from the conformations. The darker segments indicate layers which are in the "front" plane while the lighter structures lie in the "back" plane.

Conformational set:	1 (●) ϕ, ψ (°)	2 (■) ϕ, ψ (°)
αDGlc*p*(1→3)αDGlc*p*	-31,-159	-37,1
αDGlc*p*(1→3)αLRha*p*	-51,-39	-39,-33
αLRha*p*(1→4)DRibol5P	33,29	35,33
DRibol5P(O→2)αDGlc*p*	-173,-27	43,1

Table 6.4: Dihedral values used to build 20mers in Figure 6.13; values of the glycosidic linkages between the two structures are coloured in blue when they are located in different regions.

repeating unit of the last 100 ns of the simulations (Table 6.4).

Figure 6.13 shows these molecules, which exhibit two possible co-planar structures for the 6D oligosaccharide. Conformation ● adopts a ladder-like layered structure similar to 6B, where the pyran residues of each repeating do not stack upon each other, but rather the ribitol-phosphate residues stack forming the "rungs" of the ladder. The ■ conformation is similar to ●, however the pyran residues are more skewed. Also like 6B, the direction (the non-reducing to reducing end) of the pyran residues do not change with each repeating unit (as indicated by the arrows on Figure 6.13).

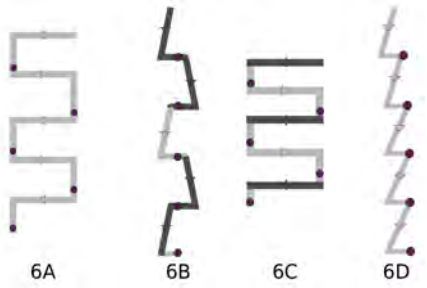
6.6 Discussion

In serotypes 6A and 6B a galactose residue is linked to two residues: αDGal*p* (1→3)αDGlc*p* and DRibol5P(O→2)αDGal*p*, which is replaced by a glucose residue in 6C and 6D. Nearly identical PMFs were calculated for the galactose and corresponding glucose counterpart disaccharides, and as such it was expected that the galactose/glucose substitution would have a minimal effect on the overall serotype structure.

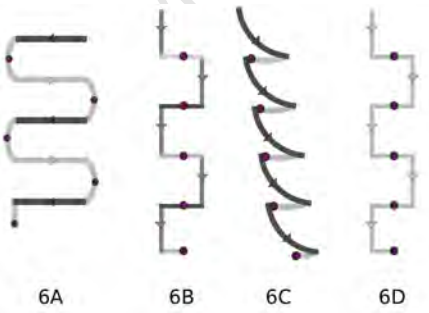
In order to observe the structural difference caused by a galactose/glucose difference, serotype pairs 6A/6C and 6B/6D were compared as the only point of difference within the pair is the galactose or glucose residue. From the results of the most populated conformations of the glycosidic linkages from the last 100 ns of the MD simulations, it would appear that the two galactose and glucose residues are likely to exist in two combinations:

The serotype pairs 6A/6C and 6B/6D show similar stacked structures. The

the ■ conformation:
αDGal$p$/αDGlc$p$(1→3)αDGlc$p$ (syn,anti) and
DRibol5P(O→2)αDGal$p$/αDGlc$p$ (anti,syn)



or the ● conformation:
αDGal$p$/αDGlc$p$(1→3)αDGlc$p$ (syn,syn) and
DRibol5P(O→2)αDGal$p$/αDGlc$p$ (syn,syn)



6A/6C pairs have pyran residues of each repeating unit layered upon each other. 6C shows an interesting alternate structure in the 6C-● conformation (not seen in 6A) with the helical structure and phosphates in the centre. The 6C/6D pair are the most similar of the serogroup: a ladder like stacked structure and an alternate helical-type structure.

The structural similarity of serotype pairs 6A/6C and 6B/6D shows that the biggest difference within the serogroup is the linkage between rhamnose and ribitol-phosphate. A rhamnose residue is linked to the ribitol-phosphate via an α(1→3) linkage in 6A and 6C and with an α(1→4) linkage in 6B and 6D. This difference can be seen in the stacking of the ladder like structures: pyran layers for the α(1→3) and ribitol-phosphate layers for the α(1→4) linkages. Serotypes 6B and 6D are also more extended structures (both stacked and helical structures) than 6A and 6C structures. The difference in saccharide extension can be explained by the extra length of the α(1→3) linkage backbone (from C3 to the P) which has greater conformational freedom to bend and thus forms a compressed structure when compared to the shorter backbone (C4 to P) of the α(1→4) link-

ages.

Ribitol-phosphate, also known as teichoic acid, is an important and common motif in bacterial polysaccharides. While the role of teichoic acids is not fully understood, it is clear that these anionic moieties play an important role in bacterial virulence, membrane integrity, and cell scaffolding and thus vaccine efficacy.[95–98] Therefore, the position of the phosphates in these serotypes is important in studying their structure. Serotypes 6A and 6C have the phosphates placed on the exterior of the structures making the anionic motif easily accessible, while in the 6B and 6D structures have less accessible phosphates in the interior of the structures. The anomaly is the alternate helical 6C structure which has sheltered phosphates. The limited number of conformational studies conducted on sugar phosphates show that the phosphate-sugar linkage is highly flexible.[99]

The distributions in conformation of the DRibol5P(O→2)αDGal$p$/αDGlc$p$ linkages during the course of these simulations shifted away from the global minima calculated in Chapter 4, but similar to that observed by H*öö*g et al. with α and ψ (in H*öö*g) corresponding to ϕ and ψ in the glycosidic linkage.[100] This difference can be explained by the fact that the phosphodiester linkages investigated by H*öö*g et al. involved two pyranosyl residues which are more rigid as opposed to one pyranose residue and one linear alditol residue (the disaccharides investigated in Chapter 5) as in the serogroup 6 saccharides. In the serotypes the interresidue effects reduce the flexibility of the ribitol therefore accounting for the distribution of the linkage conformation in the oligosaccharides.

What is clear from these results is that pairs 6A/6C and 6B/6D have similar structures and as such it is expected that serotypes 6A and 6B would cross protect against non-vaccine types 6C and 6D. This is consistent with current literature, however as PCV-13 is newly licensed (2010) and serotypes 6C and 6D are newly discovered, the clinical efficacy of this cross-protection has yet to be measured.[13,29,41,42,101] Serotype 6B does offer some cross-protection against 6A, however these results offer no obvious structural basis to explain this. Further simulations with explicit water and counter ions may shed light on this cross-protection.

These results are limited, as time constrains did not permit for the serotypes to be simulated in a water box with counter ions. The addition of counter ions

to the simulation would have been especially useful as the phosphate negative charges on these oligosaccharides are key to their behaviour in somatic environments. However the fact that both simulations (non-primed and primed for each serotype) explored similar regions regardless of the starting structure is encouraging. These simulations also give insight into ribitol-phosphate and phosphodiester modelling, on which limited previous work has been conducted.

Chapter 7

Conclusions and Future Work

The aim of this project was to structurally characterize the four serotypes of *Streptococcus pneumoniae* serogroup 6 and to identify the tertiary structural differences brought on by the slight differences in primary structure within the serogroup: a αDGal*p*/αDGlc*p* residue difference and a αLRha*p*-DRibol5P (1→3)/(1→4) linkage difference. The following questions were asked:

1. What effect does the αLRha*p*-DRibol5P (1→3)/(1→4) linkage change have on the structure of the serotypes?

2. What effect does a galactose/glucose residue change have on the structure of the serotypes?

3. Can a systematic approach to computational modelling of serogroup 6 provide insight into cross-protection observed between serotypes 6A-D?

The first step was to establish the low energy conformations of the disaccharide subunits of the serogroup. A metadynamics routine, using ϕ and ψ dihedral angles as collective variables, was used to produce potential of mean force landscapes that give insight into the low energy conformations of the disaccharides. Two classes of PMF landscapes were calculated: the first for disaccharides resulting from a naïve division of the serogroup, and the second, resulting from treating ribitol-phosphate as a substituent to adjacent residues rather than a residue itself.

These calculations showed that, within both disaccharide classes, the disaccharides with a αDGal*p*/αDGlc*p* residue difference showed nearly identical PMF

plots, indicating that this residue difference would have little effect on the overall structure of the oligosaccharides. The biggest change in energy landscapes was between residues with a differing αLRha*p*-DRibol5P (1→3)/(1→4) linkage, suggesting that this is the biggest point of difference within the serogroup.

Three repeating unit (12 residues) long oligosaccharides of serogroup 6 were built with torsion angles set to the conformations found for the global minima calculated from the disaccharide free energy calculations. Molecular Dynamics simulations with implicit solvation were conducted on these oligosaccharides and the dihedral angles of the glycosidic linkages from the middle repeating unit were plotted on their respective energy landscape. The conformations of the glycosidic linkages throughout the course of the MD simulations occupied the low energy regions on the energy landscapes, showing that the disaccharide plots are a good indicator of allowed conformations for dihedral linkages in an oligosaccharide. The DRibol5P(O→2)αDGal*p*/αDGlc*p* linkages occupied conformations away from the calculated global minima, however they did occupy low energy regions that agreed with other studies of similar linkages.

MD simulations on all four serotypes showed that the oligosaccharides had two conformational preferences. Five repeating-unit long oligosaccharides were built using these two conformational preferences and were subsequently minimized. These 20mers showed structural similarities between serotypes 6A and 6C and between 6B and 6D. Serotypes 6A and 6C have a αDGal*p*/αDGlc*p* difference but they both have a αLRha*p*(1→3)DRibol5P linkage. Both serotypes have a stacked ladder-like structure, with the phosphate motifs exposed on the outside of the structure. 6C also adopts an alternate helical structure not seen in 6A. Serotypes 6B and 6D have very similar structures and adopt either a regular ladder-like structure or an extended helical structure. In both cases the phosphates are more hidden than in the 6A/6C pair, being located in the middle of the "rungs" in the center of the structures. The results of the 20mer are in agreement with prediction from the disaccharide energy landscapes: the biggest points of difference within the serogroup is the αLRha*p*-DRibol5P (1→3)/(1→4) linkage. The structural information from this study suggests that serotype 6A will offer cross-protection against 6C and serotype 6B against 6D, as suggested by McEllistrem et. al.[29] However no structural relationships were seen between 6A and 6B to explain the

cross protection serotype 6B offers 6A.

However, this study has not considered the possible effects of solution and counter-ions on the conformations of serogroup 6. Now that low energy conformations have been identified for the four serotypes, the next step will be to undertake time consuming simulations of these molecules in solution.

References

[1] E David G McIntosh and Ralf R Reinert, "Global prevailing and emerging pediatric pneumococcal serotypes", *Expert Review of Vaccines*, vol. 10, no. 1, pp. 109–129, 2011. xi, 1, 18

[2] Hope L Johnson, Maria Deloria-Knoll, Orin S Levine, Sonia K Stoszek, Laura Freimanis Hance, Richard Reithinger, Larry R Muenz, and Katherine L O'Brien, "Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project", *PLoS Medicine*, vol. 7, no. 10, pp. e1000348, 2010. 1

[3] Ralf R Reinert, Peter Paradiso, and Bernard Fritzell, "Advances in pneumococcal vaccines: the 13-valent pneumococcal conjugate vaccine received market authorization in Europe", *Expert Review of Vaccines*, vol. 9, no. 3, pp. 229–236, 2010. 1

[4] Tirdad T Zangeneh, Gio Baracco, and Jaffar A Al-Tawfiq, "Impact of conjugate pneumococcal vaccines on the changing epidemiology of pneumococcal infections", *Expert Review of Vaccines*, vol. 10, no. 3, pp. 345–353, 2011. 1, 19

[5] Helena K Parsons and David H Dockrell, "The burden of invasive pneumococcal disease and the potential for reduction by immunisation", *International Journal of Antimicrobial Agents*, vol. 19, no. 2, pp. 85–93, 2002. 1

[6] Ping Jin, Fanrong Kong, Meng Xiao, Shahin Oftadeh, Fei Zhou, Chunyi Liu, Fiona Russell, and Gwendolyn L Gilbert, "First report of puta-

tive Streptococcus pneumoniae serotype 6D among nasopharyngeal isolates from Fijian children", *Journal of Infectious Diseases*, vol. 200, no. 9, pp. 1375–1380, 2009. 1, 4, 19

[7] Mark R Alderson, Jeff Maisonneuve, Lauren C Newhouse, and John W Boslego, "PATH pneumococcal vaccine project", *Pediatric Health*, vol. 4, no. 5, pp. 471–478, 2010. 1, 2, 3, 4

[8] Hong Le Duc, "News Brief", *Expert Review of Vaccines*, vol. 9, pp. 10, 2010. 1

[9] Lucio Toma, Laura Legnani, Anna Rencurosi, Laura Poletti, Luigi Lay, and Giovanni Russo, "Modeling of synthetic phosphono and carba analogues of N-acetyl-α-d-mannosamine 1-phosphate, the repeating unit of the capsular polysaccharide from Neisseria meningitidis serovar A", *Org. Biomol. Chem.*, vol. 7, pp. 3734–3740, 2009. 1

[10] E AlonsoDeVelasco, AF Verheul, J Verhoef, and H Snippe, "Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines.", *Microbiological Reviews*, vol. 59, no. 4, pp. 591–603, 1995. 1, 14, 16, 17

[11] Fikri Y Avci and Dennis L Kasper, "How bacterial carbohydrates influence the adaptive immune system", *Annual Review of Immunology*, vol. 28, pp. 107–130, 2010. 1, 14, 15

[12] Federica Compostella, Franca Marinone Albini, Fiamma Ronchetti, and Lucio Toma, "A theoretical study of the conformational behavior of analogues of α-l-rhamnose-1-phosphate", *Carbohydrate Research*, vol. 339, no. 7, pp. 1323–1330, 2004. 1

[13] David Cooper, Xinhong Yu, Mohinder Sidhu, Moon H Nahm, Philip Fernsten, and Kathrin U Jansen, "The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to *Streptococcus pneumoniae* serotypes 6C and 7A", *Vaccine*, vol. 29, no. 41, pp. 7207–7211, 2011. 1, 4, 5, 87

[14] Darren Taylor, "Vaccine to Save a Generation of South African Children", *Voice of America*, Nov 2010. 2

[15] Paul Roche, Vicki Krause, Mark Bartlett, David Coleman, Heather Cook, Megan Counahan, Craig Davis, Letitia Del Fabbro, Carolien Giele, Robyn Gilmore, et al., "Invasive pneumococcal disease in Australia, 2003", *Communicable Diseases Intelligence*, vol. 28, no. 4, pp. 441–454, 2004. 2

[16] Heather J Zar and Shabir A Madhi, "Pneumococcal conjugate vaccine–advancing child health in South Africa", *South African Journal of Child Health*, vol. 2, no. 3, pp. 94, 2008. 2

[17] Shabir A Madhi, "Introduction of the pneumococcal conjugate vaccine into the South African public immunisation programme: dawn of a new era?", *Southern African Journal of Epidemiology and Infection*, vol. 23, no. 4, pp. 5–9, 2009. 2

[18] "Pneumococcal Disease Poses Highest Risk In Africa South Africa is first African country to introduce vaccine", *Modern Ghana News*, March 2009. 2

[19] Keith P Klugman, "Contribution of vaccines to our understanding of pneumococcal disease", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1579, pp. 2790–2798, 2011. 2

[20] Saeyoung Park, Archana R Parameswar, Alexei V Demchenko, and Moon H Nahm, "Identification of a Simple Chemical Structure Associated with Protective Human Antibodies against Multiple Pneumococcal Serogroups", *Infection and Immunity*, vol. 79, no. 8, pp. 3472, 2011. 2, 14, 17

[21] Archana R Parameswar, In Ho Park, Rina Saksena, Pavol Kovćč, Moon H Nahm, and Alexei V Demchenko, "Synthesis, conjugation, and immunological evaluation of the serogroup 6 pneumococcal oligosaccharides", *ChemBioChem*, vol. 10, no. 18, pp. 2893–2899, 2009. 2, 17, 18

[22] Amanda M Salisburg, Ashley L Deline, Katrina W Lexa, George C Shields, and Karl N Kirschner, "Ramachandran-type plots for glycosidic linkages:

Examples from molecular dynamic simulations using the Glycam06 force field", *Journal of Computational Chemistry*, vol. 30, no. 6, pp. 910–921, 2009. 3, 11, 13, 34

[23] Halvor S Hansen and Philippe H Hünenberger, "A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers", *Journal of Computational Chemistry*, vol. 32, no. 6, pp. 998–1032, 2011. 3

[24] Michelle M Kuttel, "The conformational free energy of carbohydrates", *Mini-Reviews in Organic Chemistry*, vol. 8, no. 3, pp. 256–262, 2011. 3, 8, 13, 14

[25] Göran Widmalm, "A Perspective on the Primary and Three-dimensional Structures of Carbohydrates", *Carbohydrate Research*, 2013. 3, 13, 42

[26] Thomas Cherian, "WHO expert consultation on serotype composition of pneumococcal conjugate vaccines for use in resource-poor developing countries, 26–27 October 2006, Geneva", *Vaccine*, vol. 25, no. 36, pp. 6557–6564, 2007. 3

[27] Daniel J Isaacman, E David McIntosh, Ralf R Reinert, et al., "Burden of invasive pneumococcal disease and serotype distribution among Streptococcus pneumoniae isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines", *International Journal of Infectious Diseases*, vol. 14, no. 3, pp. e197–209, 2010. 3

[28] Preston E Bratcher, Kyung-Hyo Kim, Jin H Kang, Jung Y Hong, and Moon H Nahm, "Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization", *Microbiology*, vol. 156, no. 2, pp. 555–560, 2010. 4, 19

[29] M Catherine McEllistrem and Moon H Nahm, "Novel pneumococcal serotypes 6C and 6D: Anomaly or Harbinger", *Clinical Infectious Diseases*, vol. 55, no. 10, pp. 1379–1386, 2012. 4, 5, 87, 90

[30] Gerald Zon, SHOUSUN C Szu, WILLIAM Egan, JOAN D Robbins, and JB Robbins, "Hydrolytic stability of pneumococcal group 6 (type 6A and 6B) capsular polysaccharides.", *Infection and Immunity*, vol. 37, no. 1, pp. 89–103, 1982. 4

[31] Peter R Paradiso, "Advances in pneumococcal disease prevention: 13-valent pneumococcal conjugate vaccine for infants and children", *Clinical Infectious Diseases*, vol. 52, no. 10, pp. 1241–1247, 2011. 4, 5, 19

[32] Bin Chang, Taketo Otsuka, Atsushi Iwaya, Minoru Okazaki, Satoko Matsunaga, and Akihito Wada, "Isolation of Streptococcus pneumoniae serotypes 6C and 6D from the nasopharyngeal mucosa of healthy Japanese children", *Japanese Journal of Infectious Diseases*, vol. 63, pp. 381–383, 2010. 4, 5, 19

[33] In H Park, Matthew R Moore, John J Treanor, Stephen I Pelton, Tamara Pilishvili, Bernard Beall, Mark A Shelly, Barbara E Mahon, Moon H Nahm, et al., "Differential effects of pneumococcal vaccines against serotypes 6A and 6C", *Journal of Infectious Diseases*, vol. 198, no. 12, pp. 1818–1822, 2008. 4

[34] Michael R Jacobs, Caryn E Good, Saralee Bajaksouzian, and Anne R Windau, "Emergence of Streptococcus pneumoniae serotypes 19A, 6C, and 22F and serogroup 15 in Cleveland, Ohio, in relation to introduction of the protein-conjugated pneumococcal vaccine", *Clinical Infectious Diseases*, vol. 47, no. 11, pp. 1388–1395, 2008. 4, 18

[35] Moon H Nahm, Jisheng Lin, Jonathan A Finkelstein, and Stephen I Pelton, "Increase in the prevalence of the newly discovered pneumococcal serotype 6C in the nasopharynx after introduction of pneumococcal conjugate vaccine", *Journal of Infectious Diseases*, vol. 199, no. 3, pp. 320–325, 2009. 4

[36] Nikkol Melnick, Terry A Thompson, and Bernard W Beall, "Serotype-specific typing antisera for pneumococcal serogroup 6 serotypes 6A, 6B,

and 6C", *Journal of Clinical Microbiology*, vol. 48, no. 6, pp. 2311–2312, 2010. 4

[37] Preston E Bratcher, In H Park, Susan K Hollingshead, and Moon H Nahm, "Production of a unique pneumococcal capsule serotype belonging to serogroup 6", *Microbiology*, vol. 155, no. 2, pp. 576–583, 2009. 4

[38] Erik Mercado, Velusamy Srinivasan, Paulina Hawkins, Sopio Chochua, Theresa Ochoa, Bernard Beall, and Lesley McGee, "First report of Streptococcus pneumoniae serotype 6D in South America", *Journal of Clinical Microbiology*, vol. 49, no. 5, pp. 2080–2081, 2011. 4, 19

[39] Kwan Soo Ko, Jin Yang Baek, and Jae-Hoon Song, "Multidrug-resistant Streptococcus pneumoniae serotype 6D clones in South Korea", *Journal of Clinical Microbiology*, vol. 50, no. 3, pp. 818–822, 2012. 4

[40] JB Robbins, R Austrian, C-J Lee, SC Rastogi, G Schiffman, J Henrichsen, PH Mäkelä, CV Broome, RR Facklam, RH Tiesjema, et al., "Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups", *The Journal of Infectious Diseases*, vol. 148, no. 6, pp. 1136–1159, 1983. 5

[41] Elizabeth Miller, Nicholas J Andrews, Pauline A Waight, Mary PE Slack, and Robert C George, "Effectiveness of the new serotypes in the 13-valent pneumococcal conjugate vaccine", *Vaccine*, vol. 29, no. 49, pp. 9127–9131, 2011. 5, 87

[42] Julie M Skinner, Lani Indrawati, Jayme Cannon, Jeffrey Blue, Michael Winters, John MacNair, Narahari Pujar, Walter Manger, Yuhua Zhang, Joseph Antonello, et al., "Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model", *Vaccine*, vol. 29, no. 48, pp. 8870–8876, 2011. 5, 19, 87

[43] Anne Imberty and Serge Pérez, "Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental valida-

tions", *Chemical Reviews-Columbus*, vol. 100, no. 12, pp. 4567–4588, 2000. 8, 11, 12, 13

[44] Robert Eklund and Göran Widmalm, "Molecular dynamics simulations of an oligosaccharide using a force field modified for carbohydrates", *Carbohydrate Research*, vol. 338, no. 5, pp. 393–398, 2003. 8

[45] William A Bubb, "NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity", *Concepts in Magnetic Resonance Part A*, vol. 19, no. 1, pp. 1–19, 2003. 8, 12, 13

[46] Cristina S Pereira, David Kony, Riccardo Baron, Martin Müller, Wilfred F van Gunsteren, and Philippe H Hünenberger, "Conformational and dynamical properties of disaccharides in water: a molecular dynamics study", *Biophysical Journal*, vol. 90, no. 12, pp. 4337–4344, 2006. 8, 11, 12, 13, 40, 42, 53

[47] Lovorka Perić-Hassler, Halvor S Hansen, Riccardo Baron, Philippe H Hünenberger, et al., "Conformational properties of glucose-based disaccharides investigated using molecular dynamics simulations with local elevation umbrella sampling", *Carbohydrate Research*, vol. 345, no. 12, pp. 1781, 2010. 8, 12, 40, 42, 53, 64

[48] Elizabeth Hatcher, Elin Säwén, Göran Widmalm, and Alexander D MacKerell Jr, "Conformational properties of methyl β-maltoside and methyl α-and β-cellobioside disaccharides", *The Journal of Physical Chemistry B*, vol. 115, no. 3, pp. 597–608, 2010. 8

[49] Mark R Wormald, Andrei J Petrescu, Ya-Lan Pao, Ann Glithero, Tim Elliott, and Raymond A Dwek, "Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling", *Chemical Reviews*, vol. 102, no. 2, pp. 371–386, 2002. 11, 12, 13, 14

[50] Robert J Woods, "Computational carbohydrate chemistry: what theoretical methods can tell us", *Glycoconjugate Journal*, vol. 15, no. 3, pp. 209–216, 1998. 12, 13, 14

[51] Michelle M Kuttel and Kevin J Naidoo, "Ramachandran free-energy surfaces for disaccharides: trehalose, a case study", *Carbohydrate Research*, vol. 340, no. 5, pp. 875–879, 2005. 13, 34

[52] Norman WH Cheetham, Paramita Dasgupta, and Graham E Ball, "NMR and modelling studies of disaccharide conformation", *Carbohydrate Research*, vol. 338, no. 9, pp. 955–962, 2003. 13, 40, 53

[53] Claus-Wilhelm von der Lieth, *Experimental Methods for the Analysis of Glycans and Their Bioinformatics Requirements*, chapter 10, Wiley, 2009. 13, 14

[54] C Allen Bush, Manuel Martin-Pastor, and Anne Imbery, "Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides", *Annual review of Biophysics and Biomolecular Structure*, vol. 28, no. 1, pp. 269–293, 1999. 13, 14

[55] Anne Imberty, "Oligosaccharide structures: theory versus experiment", *Current Opinion in Structural Biology*, vol. 7, no. 5, pp. 617–623, 1997. 13

[56] JW Brady and RK Schmidt, "The role of hydrogen bonding in carbohydrates: molecular dynamics simulations of maltose in aqueous solution", *The Journal of Physical Chemistry*, vol. 97, no. 4, pp. 958–966, 1993. 14

[57] Yu S Ovodov, "Bacterial capsular antigens. Structural patterns of capsular antigens", *Biochemistry (Moscow)*, vol. 71, no. 9, pp. 937–954, 2006. 14

[58] D Goldblatt, "Conjugate vaccines", *Clinical & Experimental Immunology*, vol. 119, no. 1, pp. 1–3, 2000. 15

[59] Paolo Costantino, Rino Rappuoli, and Francesco Berti, "The design of semi-synthetic and synthetic glycoconjugate vaccines", *Expert Opinion on Drug Discovery*, vol. 6, no. 10, pp. 1045–1066, 2011. 15

[60] Marc P Girard, Marie-Pierre Preziosi, Maria-Teresa Aguado, and Marie Paule Kieny, "A review of vaccine research and development: meningococcal disease", *Vaccine*, vol. 24, no. 22, pp. 4692–4700, 2006. 16

[61] Caroline L Trotter and Mary E Ramsay, "Vaccination against meningococcal disease in Europe: review and recommendations for the use of conjugate vaccines", *FEMS Microbiology Reviews*, vol. 31, no. 1, pp. 101–107, 2007. 16

[62] G.R. Siber, K.P. Klugman, and P.H. Mäkelä, *Pneumococcal Vaccines: The Impact of Conjugate Vaccine*, American Society Mic Series. Amer Society for Microbiology, 2008. 16, 17

[63] Ener Cagri Dinleyici, "Current status of pneumococcal vaccines: lessons to be learned and new insights", *Expert Review of Vaccines*, vol. 9, no. 9, pp. 1017–1022, 2010. 18

[64] In Ho Park, David G Pritchard, Rob Cartee, Angela Brandao, Maria Cristina C Brandileone, and Moon H Nahm, "Discovery of a new capsular serotype (6C) within serogroup 6 of Streptococcus pneumoniae", *Journal of Clinical Microbiology*, vol. 45, no. 4, pp. 1225–1233, 2007. 18

[65] Dora Rolo, Asunción Fenoll, Carmen Ardanuy, Laura Calatayud, Meritxell Cubero, G Adela, and Josefina Liñares, "Trends of invasive serotype 6C pneumococci in Spain: emergence of a new lineage", *Journal of Antimicrobial Chemotherapy*, vol. 66, no. 8, pp. 1712–1718, 2011. 18

[66] JM Marimon, M Ercibengoa, M Alonso, G García-Medina, and E Pérez-Trallero, "Prevalence and molecular characterization of Streptococcus pneumoniae serotype 6C causing invasive disease in Gipuzkoa, northern Spain, 1990–2009", *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 29, no. 8, pp. 1035–1038, 2010. 18

[67] Robert Cohen, Corinne Levy, Edouard Bingen, Marc Koskas, Isabelle Nave, and Emmanuelle Varon, "Impact of 13-valent pneumococcal conjugate vaccine on pneumococcal nasopharyngeal carriage in children with acute otitis media", *The Pediatric Infectious Disease Journal*, vol. 31, no. 3, pp. 297–301, 2012. 19

[68] B Lachele Foley, Matthew B Tessier, and Robert J Woods, "Carbohydrate force fields", *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 4, pp. 652–697, 2012. 23, 24

[69] Sookhee N Ha, Ann Giammona, Martin Field, and John W Brady, "A revised potential-energy surface for molecular mechanics studies of carbohydrates", *Carbohydrate Research*, vol. 180, no. 2, pp. 207–221, 1988. 24

[70] Stephan Reiling, Michael Schlenkrich, and Jürgen Brickmann, "Force field parameters for carbohydrates", *Journal of Computational Chemistry*, vol. 17, no. 4, pp. 450–468, 1996. 24

[71] Michelle Kuttel, John W Brady, and Kevin J Naidoo, "Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations", *Journal of Computational Chemistry*, vol. 23, no. 13, pp. 1236–1243, 2002. 24, 38

[72] Olgun Guvench, Shannon N Greene, Ganesh Kamath, John W Brady, Richard M Venable, Richard W Pastor, and Alexander D Mackerell, "Additive empirical force field for hexopyranose monosaccharides", *Journal of Computational Chemistry*, vol. 29, no. 15, pp. 2543–2564, 2008. 25

[73] Elizabeth R Hatcher, Olgun Guvench, and Alexander D MacKerell Jr, "CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol", *Journal of Chemical Theory and Computation*, vol. 5, no. 5, pp. 1315–1327, 2009. 25

[74] Olgun Guvench, Sairam S Mallajosyula, E Prabhu Raman, Elizabeth Hatcher, Kenno Vanommeslaeghe, Theresa J Foster, Francis W Jamison, and Alexander D MacKerell Jr, "CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling", *Journal of Chemical Theory and Computation*, vol. 7, no. 10, pp. 3162–3180, 2011. 25

[75] Elizabeth Hatcher, Olgun Guvench, and Alexander D MacKerell Jr, "CHARMM additive all-atom force field for aldopentofuranoses, methyl-

aldopentofuranosides, and fructofuranose", *The Journal of Physical Chemistry B*, vol. 113, no. 37, pp. 12466–12476, 2009. 25

[76] Mallajosyula, Sairam S and Guvench, Olgun and Hatcher, Elizabeth and MacKerell Jr, Alexander D, "CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates", *Journal of Chemical Theory and Computation*, vol. 8, no. 2, pp. 759–776, 2012. 25, 37

[77] Alessandro Laio and Francesco L Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science", *Reports on Progress in Physics*, vol. 71, no. 12, pp. 126601, 2008. 31, 38

[78] Michelle Kuttel, Neil Ravenscroft, Michela Foschiatti, Paola Cescutti, and Roberto Rizzo, "Conformational properties of two exopolysaccharides produced by *Inquilinus limosus*, a cystic fibrosis lung pathogen", *Carbohydrate Research*, vol. 350, pp. 40–48, 2012. 34

[79] Wei Li, Steve W Cui, Qi Wang, and Rickey Y Yada, "Study of conformational properties of cereal β-glucans by computer modeling", *Food Hydrocolloids*, vol. 26, no. 2, pp. 377–382, 2012. 34

[80] Laercio Pol-Fachin, Rodrigo V Serrato, and Hugo Verli, "Solution conformation and dynamics of exopolysaccharides from *Burkholderia* species", *Carbohydrate Research*, vol. 345, no. 13, pp. 1922–1931, 2010. 34

[81] Keiko Kondo, Minoru Takeda, Tsukasa Mashima, Masato Katahira, Junichi Koizumi, and Kazuyoshi Ueda, "Conformational analysis of an extracellular polysaccharide produced by *Sphaerotilus natans*", *Carbohydrate Research*, 2012. 34

[82] Laura Legnani, Silvia Ronchi, Silvia Fallarini, Grazia Lombardi, Federica Campo, Luigi Panza, Luigi Lay, Laura Poletti, Lucio Toma, Fiamma Ronchetti, et al., "Synthesis, molecular dynamics simulations, and biology of a carba-analogue of the trisaccharide repeating unit of Streptococcus pneumoniae 19F capsular polysaccharide", *Organic & Biomolecular Chemistry*, vol. 7, no. 21, pp. 4428–4436, 2009. 34

[83] Lycknert, Kristina and Edblad, Malin and Imberty, Anne and Widmalm, Göran, "NMR and Molecular Modeling Studies of the Interaction between Wheat Germ Agglutinin and the β-d-Glc p NAc-(1→6)-α-d-Man p Epitope Present in Glycoproteins of Tumor Cells", *Biochemistry*, vol. 43, no. 30, pp. 9647–9654, 2004. 34

[84] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten, "Scalable molecular dynamics with NAMD", *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005. 37

[85] William Humphrey, Andrew Dalke, and Klaus Schulten, "VMD: visual molecular dynamics", *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996. 37

[86] Michelle M Kuttel and Kevin J Naidoo, "Free energy surfaces for the α (1→4)-glycosidic linkage: Implications for polysaccharide solution structure and dynamics", *The Journal of Physical Chemistry B*, vol. 109, no. 15, pp. 7468–7474, 2005. 38, 39

[87] VS Raghavendra Rao, *Conformation of Carbohydrates*, CRC PressI Llc, 1998. 40, 53

[88] Alfred D French, Anne-Marie Kelterer, Glenn P Johnson, Michael K Dowd, and Christopher J Cramer, "HF/6-31G* energy surfaces for disaccharide analogs", *Journal of Computational Chemistry*, vol. 22, no. 1, pp. 65–78, 2001. 40, 53

[89] Michael K Dowd, Jing Zeng, Alfred D French, and Peter J Reilly, "Conformational analysis of the anomeric forms of kojibiose, nigerose, and maltose using MM3", *Carbohydrate Research*, vol. 230, no. 2, pp. 223–244, 1992. 40, 53

[90] Karl-Heinz Ott and Bernd Meyer, "Molecular dynamics simulations of maltose in water", *Carbohydrate Research*, vol. 281, no. 1, pp. 11–34, 1996. 40

[91] Michelle Kuttel, Yue Mao, Göran Widmalm, and Magnus Lundborg, "CarbBuilder: An adjustable tool for building 3D molecular structures of carbohydrates for molecular simulation", in *2011 IEEE 7th International Conference on e-Science*. IEEE, 2011, pp. 395–402. 40

[92] Hugo Benjamin V. and Talbot Brandon J., "Heightmap Analyser", Tech. Rep., University of Cape Town, Computer Science, 2012. 50

[93] Norman WH Cheetham and Paramita Dasgupta, "Studies of Disaccharide Solvation: Molecular Dynamics versus HPLC Retention", *Australian journal of chemistry*, vol. 58, no. 11, pp. 803–809, 2006. 53

[94] Clas Landersjö, Roland Stenutz, and Göran Widmalm, "Conformational Flexibility of Carbohydrates: A Folded Conformer at the ϕ Dihedral Angle of a Glycosidic Linkage", *Journal of the American Chemical Society*, vol. 119, no. 37, pp. 8695–8698, 1997. 64

[95] Jonathan G Swoboda, Jennifer Campbell, Timothy C Meredith, and Suzanne Walker, "Wall teichoic acid function, biosynthesis, and inhibition", *Chembiochem*, vol. 11, no. 1, pp. 35–45, 2010. 87

[96] AR Archibald, "The structure, biosynthesis and function of teichoic acid", *Adv. Microb. Physiol*, vol. 11, pp. 53–95, 1974. 87

[97] James Baddiley, "Structure, biosynthesis, and function of teichoic acids", *Accounts of Chemical Research*, vol. 3, no. 3, pp. 98–105, 1970. 87

[98] D Bogaert, R De Groot, and PWM Hermans, "Streptococcus pneumoniae colonisation: the key to pneumococcal disease", *The Lancet Infectious Diseases*, vol. 4, no. 3, pp. 144–154, 2004. 87

[99] M Maestre and CS Pérez, "Conformational analysis of diribosylribitol phosphate by NMR spectroscopy and molecular dynamics", *Magnetic Resonance in Chemistry*, vol. 38, no. 2, pp. 123–125, 2000. 87

[100] Christer Höög, Aatto Laaksonen, and Göran Widmalm, "Molecular Dynamics Simulations of the Phosphodiester-Linked Repeating Units of the

Haemophilus i nfluenzae Types c and f Capsular Polysaccharides", *The Journal of Physical Chemistry B*, vol. 105, no. 29, pp. 7074–7079, 2001. 87

[101] Nurith Porat, Rachel Benisty, Ronit Trefler, Doreen Ozalvo, Noga Givon-Lavi, and Ron Dagan, "Baseline epidemiology and genetic structure of Streptococcus pneumoniae serotype 6D in southern Israel prior to the introduction of pneumococcal conjugate vaccines", *Journal of Clinical Microbiology*, 2013. 87