Project Proposal

Developing methodologies to elucidate carbohydrate conformation of pneumococcal antigens.

Neann Mathai
Department of Computer Science
University of Cape Town
nmathai@cs.uct.ac.za

Supervisor: Dr. Michelle Kuttel*
Co-supervisor: Assoc. Prof Neil Ravenscroft†

June 17, 2011

Contents

1 Introduction 1
1.1 Invasive Pneumococcal Disease (IPD) ......................... 1
1.2 Pneumococcal Conjugate Vaccines .......................... 2
1.3 Carbohydrate Force Fields .................................. 2
1.4 Serogroup 6 ................................................. 3

2 Related Work .................................................. 3

3 Research Questions .............................................. 4

4 Research Plan .................................................... 5
4.1 Probable Conformation of Disaccharides ................... 5
4.2 Oligosaccharides and Molecular Dynamics ................ 6
4.3 Risks ......................................................... 7
4.4 Time-Line .................................................... 7
4.4.1 Milestones ............................................... 8
4.5 Outputs ...................................................... 8
4.6 Success Factors ............................................. 8

* mkuttel@cs.uct.ac.za (Department of Computer Science, UCT)
† neil.ravenscroft@uct.ac.za (Department of Chemistry, UCT)
1 Introduction

Invasive pneumococcal disease (IPD) is a leading cause of death in children under 5, killing nearly 800,000 children each year [12, 22]. Over the years, vaccine development against *S.pneumoniae* has come far, Figure 1, and the latest conjugate vaccines have been shown to be effective in preventing disease in high risk population groups. While conjugate vaccines are effective, they do not protect against all the serotype causing IPDs.

In order to expand the range of serotypes covered by a conjugate vaccine, the structure of the capsular polysaccharides of these serotypes must be understood. Understanding the three dimensional structure of carbohydrates is a complex problem as experimental data is unclear. Computational simulations are thus used to complement these experiments. Molecular dynamics (MD) simulations of molecules are conducted in well-defined empirical force fields which aims 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.

During the course of this project a thoroughly systematic study will be conducted to understand the conformation of the oligosaccharide repeating units of the four serotypes of serogroup 6. This will first begin by exploring the entire conformational space for all the disaccharide substructures present, to gain information on the preferred conformations of these subunits. The disaccharides, in their preferred subunits, will then be pieced together into the four oligosaccharides. MD simulations of these oligosaccharides will then be conducted in a vacuum, water, and ion solution, the results of which will then be used to explain the cross protectivity observed in this group.

An implication of this project is the creation of a well-defined methodology to study oligosaccharides. In addition to this, insights on serogroup 6 of *S.pneumoniae* will also be gained.

1.1 Invasive Pneumococcal Disease (IPD)

Nearly all the fatalities, caused by IPDs, occur in developing countries in Asia and Africa [1]. In addition to pneumonia, *Streptococcus pneumoniae* also causes other diseases, such as: meningitis, sepsis, otitis media and other mucosal diseases [1, 11].

*S. pneumoniae* are surrounded by a protective polysaccharide capsule. Polysaccharides, also known as carbohydrates, are long chains built from monomers, which can join together in a variety of possible combinations. There are also a variety of linkages that are possible between two monosaccharides, and variation is further increased by the fact that these linkages have great flexibility, which allows the molecule to adopt a variety 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 [28, 3].

On the basis of the structure of the capsular polysaccharide (PS), pneumococci are classified into over 90 different serotypes. Although these bacteria share many similarities, just a handful of serotypes cause most of the IPDs in the world [12].

The history of pneumococcal vaccines can be seen in Figure 1. The original pneumococcal vaccines were whole cell vaccines, first introduced in 1918, administered to the patient in two doses [25, 3]. These vaccines had very strong side effects, and as a result, the 1920s saw the development of PS-based vaccines.

In 1920, the capsular PSs and nucleoproteins of the bacteria were chemically isolated from bacterial cultures. When animals were immunized with the capsular PS, 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 not protective while the type specific antibodies did elicit protection. This lead to the development of vaccines, with the aim of combining the different PSs to from a polyvalent vaccine which would protect against a range of IPD causing serotypes.[25]

The first of these polyvalent vaccines, a four-valent vaccine, showed good efficacy, protecting against
serotypes 1, 2, 5, and 7 [25, 19]. Despite its efficacy, the then newly-developed hexavalent PS vaccine was withdrawn due to the wide use of antibiotics such as penicillin [3]. The eventual rise of antibody resistant pneumococci renewed the interest in PS vaccines [25, 3]. From the late 1960’s this interest led to the successful 14-valent and then 23-valent PS vaccines [25]. However, while PS 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 [3, 19].

Figure 1: History of Pneumococcal Vaccines

1.2 Pneumococcal Conjugate Vaccines

It was found that coupling pneumococcal PSs 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 difficult, production of conjugate vaccines is expensive, and conjugate vaccines cater for a limited number of serotypes [20].

One approach to improving conjugate vaccines is to cater for more serotypes [20, 19]. This requires knowledge of which PSs might elicit antibodies. Understanding the three dimensional structure of biomolecules, plays a huge part of understanding their properties and functions [24]. The extreme flexibility of carbohydrates makes it difficult to experimentally determine their conformation, and consequently only a few studies have looked at the structure of the PSs.

With the appearance of more powerful computational technology, molecular mechanics has been one of the techniques used to study molecular structure. While one can obtain structures of carbohydrates using X-ray crystallography, these results are for static structures and require crystals. NMR data can also provide structural information of carbohydrates in solution; however this data is often an average of all the populated conformation states [8, 14]. As a result of these difficulties, molecular modelling has emerged to compliment these experiments.

1.3 Carbohydrate Force Fields

In order to carry out molecular simulations, one needs to have force fields which accurately model experimental conditions [9]. The validity of the results of a molecular simulation, is highly dependent on the quality of the force field chosen [10]. Due to the structural diversity, designing carbohydrate force fields is a difficult task, however significant efforts are continuously made to accurately produce more accurate force fields.
fields, and as a result produce more accurate results from computational experiments [8, 7]. Two important questions always need to be asked when such studies are conducted [5]. These are:

1. Which force field is best suited for the problem?
2. Can the results produced by the force fields be trusted?

Common force fields that have been developed for carbohydrates include: the GLYCAM series (for AMBER), OPLS-AA, GROMOS, and CSFF (for CHARMM). Studies have shown that these force fields give reasonable results and predict conformational equilibria consistently. [14]

1.4 Serogroup 6

Serogroup 6 is the focus of this project, and it is composed of four serotypes: 6A, 6B, 6C, and 6D. It has been calculated that serotypes 6A and 6B account for 14% - 18% of IPD across the different regions of the world [12]. With the introduction of pneumococcal conjugate vaccine Prevnar/Prevenar (PCV7), it was discovered that the carriage and IPD due to serotype 6B was nearly eliminated [6]. This is to be expected as 6B is included in PCV7. In addition to this, it was found that IPD due to 6A has reduced, suggesting that there is a cross protection provided for against 6A by 6B [6]. In 2007, a new serotype, 6C was identified, originally indistinguishable from 6A [6]. It has been observed that 6B provides no cross protection for 6C. In fact, since the introduction of PCV7, the number of cases of 6C have increased [27]. In an attempt to protect against 6C, the potential of a newly synthetically developed serotype, 6D, is being investigated [4, 18]. Naturally occurring 6D has also recently been identified [4, 18]. The serotype replacement that is seen is yet another motivation to increase coverage of vaccines, and in order to reduce costs, identifying motifs that will protect against more than one serotype is key. There is limited knowledge about the conformation of the linkages between the monosaccharides units, however, the monosaccharide sequence of the four group 6 serotypes, is known and are:

- 6A: [-2)-α-D-Galp-(1-3)-α-D-Glcp-(1-3)-α-L-Rhap-(1-3)-D-Rib-ol-(5(P-]
- 6B: [-2)-α-D-Galp-(1-3)-α-D-Glcp-(1-3)-α-L-Rhap-(1-2)-D-Rib-ol-(5(P-]
- 6C: [-2)-α-D-Glcp-(1-3)-α-D-Glcp-(1-3)-α-L-Rhap-(1-3)-D-Rib-ol-(5(P-]
- 6D: [-2)-α-D-Glcp-(1-3)-α-D-Glcp-(1-3)-α-L-Rhap-(1-2)-D-Rib-ol-(5(P-]

2 Related Work

Molecular dynamics (MD) is a simulation technique which involves solving classical equations of motion, in a step-by-step fashion, for each of the atoms in a system [2]. It is used to calculate the time dependent behaviour, such as structure, dynamics and thermodynamics, of a system within a given environment [26]. There have been relatively few studies that have evaluated the structure of carbohydrates; this is due to the large variety of conformations and configurations that a polysaccharide can adopt. In 2009, Legnanin et. al investigated *S.pneumoniae* serotype 19F, and a 19F carba-analogue [16]. The aim of this study was to find molecules which mimicked the biological functions of 19F, but had a greater stability, potency or efficacy. The study was conducted by first carrying out MD simulations on disaccharide substructures of 19F and the 19F analogue, first in a vacuum and then in water. MD simulations were then conducted on the 19F and 19F analogue in a vacuum. It was shown that carba-analogue did show almost the same conformational behaviour as 19F, a strong case for further study into a longer synthetic oligomer as a possible vaccine against 19F. [16]
Pereira et. al, conducted a study on the behaviour of different types of glycosidic linkages, of the disaccharides of D-Glc, in water. They carried out MD simulations on eight glucose disaccharides, showcasing the possible glycosidic linkages. The disaccharide, nigerose (Glcα(1-3)Glc), studied by Pereira et. al, is of particular interest to this project as it is found in both serotypes 6C and 6D. [21]

The use of Ramachandaran type plots to describe carbohydrate conformation is increasing in popularity. Kuttel et. al, used Ramachandaran type free energy surfaces to investigate the disaccharide trehalose. The authors produced free-energy surfaces to study the conformation of the glycosidic linkages of trehalose, both in a vacuum and an aqueous solution. The results of the free energy plots showed that the low energy conformation of simulations corresponded to experimental crystal structure and NMR data. [15]

Salisburg et. al also present an extension of a traditional Ramachandran Plot to describe glycosidic phi, psi and omega angles. The authors look at a variety of disaccharides, conducting minimizations and MD simulations. The plots were then generated using histograms of the conformations occupied. [24]

These plots were also used by Lycknert et. al, who studied the interaction of Wheat Germ Agglutinin and disaccharide β-D-GlcpNAC-(1-6)-α-DManp. In order to study this interaction they first analysed the conformation of the disaccharide, by producing energy maps as a function of the glycosidic linkages. They used these energy maps to determine that this disaccharide could comfortably exist in a few different conformations, which was then used for the rest of their study. [17]

Despite being such an important class of biomolecules, there is still much work to be done to fully understand the structure of polysaccharides.

### 3 Research Questions

When looking at serogroup 6, the following questions are asked:

1. **What is the difference between 6A and 6B?**
   Unlike 6B, serotype 6A is unstable, and as a result it has not been included in the conjugate vaccines [25, 23]. What structural difference between 6A and 6B explains 6A’s instability?

2. **Why does 6B offer cross protection against 6A?**
   6B provides some cross protection against IPD’s caused by 6A. Since the only difference between 6A and 6B is a α-L-Rhap-(1-3)-D-Rib-ol and α-L-Rhap-(1-2)-D-Rib-ol linkage, what is it about these structures that explains the cross protection offered by 6B?

3. **What is the feasibility of 6D protecting against 6C?**
   6C and 6D are similar to 6A and 6B. The serotype 6D is found to be stable while 6C is not. As the discovery of 6D in nature is recent [4], there is no information yet on whether it offers cross protection against 6C.

4. **What is the feasibility of α-D-Glcp-(1-3)-α-L-Rhap protecting against the whole serogroup?**
   α-D-Glcp-(1-3)-α-L-Rhap is found in all the members of serogroup 6. Is it feasible to elicit protection against all of serogroup 6 using this disaccharide?

The aim of this project is to develop a methodology to investigate the conformation of the glycosidic linkages found in these oligosaccharides, and in doing so, answer the questions above. The protocol to obtain this conformational information will first be tested using a model disaccharide for which a considerable amount of experimental data on its conformation already exists. The information obtained from the protocol will then compared to that from the experimental data, which serve to validate the methodology used for this project.
4 Research Plan

The process of how this research will be carried out can be seen in the figure below:

Figure 2: Methodology to be followed for this study

4.1 Probable Conformation of Disaccharides

For each serotype, every glycosidic linkage will be investigated. This will be done by conducting analysis of each disaccharide component. The serotypes are broken down into their disaccharide components, as disaccharides are the smallest molecules that possess all the rotational degrees of freedom present in oligo- and polysaccharides. Understanding the conformation of the disaccharides is an important first step towards understanding the configuration of an oligosaccharide [21].

Figure 3: Schematic of an oligosaccharide and its disaccharide substructures
We will first begin by looking for a structure of the disaccharide, in a PDB format, a file format that contains the coordinates of the atoms of a molecule. Often this might not be available, and in such cases a structure will be produced using the molecule building application CarbBuilder [29]. In addition to this, we also need to look for conformational data, crystallographic or NMR, for the glycosidic linkages, which will give insight at a good starting conformation.

The starting conformation of a disaccharide will also be determined by exploring the entire conformational space incrementally. At each increment, the disaccharide will be minimised and the enthalpic energy of the disaccharide will be recorded. Once all the conformations have been explored, an enthalpic plot, the plot of the enthalpy at the different conformations, will be produced. From this plot, the conformations with low enthalpies can be seen, and thus a good starting conformation can be chosen. An example of such a plot is shown in Figure 4, which shows the preferred conformations of the analysed disaccharide.

![Disaccharide conformation of mannan](image)

**Figure 4: Disaccharide conformation of mannan [13]**

### 4.2 Oligosaccharides and Molecular Dynamics

Once a good starting structure of each disaccharide has been determined, the disaccharide components will be pieced together to build the oligosaccharide. It is important that the serotype be built with good disaccharide starting structures, as this will considerably reduce the computational time of the molecular dynamics conducted on the oligosaccharide.

Once the oligosaccharides have been built, Molecular dynamics will then be conducted. First the simulations will be carried out in a vacuum, then in a solution and finally with ions.

Having completed the MD simulations, analysis and comparisons of the serotype structures will then be conducted; with a particular focus on specific segments:

- D-Rib-ol-(5(P)→2)-α-D-Galp-(1→)3) (in 6A and 6B)
- D-Rib-ol-(5(P)→2)-α-D-Gclp-(1→)3) (in 6C)
- L-Rhap-(1→3)-D-Rib-ol-(5P) (in 6A)
- L-Rhap-(1→3)-D-Rib-ol-(5P) (in 6B)
- D-Glpa-(1→3)-α-D-Glc (in 6A and 6B)
- D-Glcp-(1→3)-α-D-Glc (in 6C)
- α-D-Glcp-(1→3)-α-L-Rhap (in 6A, 6B and 6C)

4.3 Risks

The potential risks of this project are described in the table below:

<table>
<thead>
<tr>
<th>Identified Risk</th>
<th>Likelihood of occurrence</th>
<th>Impact of occurrence</th>
<th>Actions to reduce likelihood</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force field not accurate</td>
<td>Low</td>
<td>High</td>
<td>Research appropriate forcefields</td>
<td>Work with a different forcefield</td>
</tr>
<tr>
<td>NMR and X-Ray data of disaccharide not found</td>
<td>High</td>
<td>Low</td>
<td>Continue on without data</td>
<td>Continue without data</td>
</tr>
<tr>
<td>PDB structures not are found</td>
<td>High</td>
<td>Low</td>
<td>Build PDBs using CarbiBuilder</td>
<td></td>
</tr>
<tr>
<td>Methodology is incorrect</td>
<td>Low</td>
<td>High</td>
<td>Adopt fast fail and test using known disaccharide early on</td>
<td>Redesign method</td>
</tr>
<tr>
<td>Project scope is too large</td>
<td>Medium</td>
<td>Low</td>
<td>Meet with supervisor weekly to reassess progress and scope</td>
<td>Re-evaluate the scope; reduce the number of oligosaccharide</td>
</tr>
<tr>
<td>Project not completed in appropriate time</td>
<td>Medium</td>
<td>High</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Potential risks of the project

4.4 Time-Line

The expected timeline of this project is presented in the Gantt Chart shown in Figure 5.

Figure 5: Gantt Chart
4.4.1 Milestones

The milestones of this project are as follows:

- **June 2011** Research proposal
- **August 2011** Calculate enthalpic plot for model disaccharide
- **September 2011** Disaccharides enthalpy plots calculated
- **October 2011** Background Chapter
- **October 2011** Oligosaccharides in vacuum complete
- **November 2011** Design/Methodology Chapter
- **December 2011** Oligosaccharides in solution with ions complete
- **April 2012** Preparation and submission of a paper to a journal
- **May 2012** Completed first draft of thesis
- **August 2012** Final submission of thesis

4.5 Outputs

The primary goal of this project is to set up a research methodology in order to thoroughly perform conformational analysis of carbohydrates. We will analyse the four oligosaccharides that make up serogroup 6 of *Streptococcus pneumoniae*, to gain insight into cross protection that occurs within this group.

In addition to the above, we aim to produce a paper of this research, for submission to an international scientific journal.

4.6 Success Factors

The key success factor, of this project, will be the conformational knowledge of the investigated oligosaccharides gained from the proposed methodology. This will serve in not only understanding the investigated oligosaccharides, but also aid in setting up a protocol for further studies on the structure of carbohydrates.

In addition, we hope that this information can be used to explain the cross reactivity seen in serogroup 6, which will contribute to the knowledge used for vaccine development against *S.pneumoniae*. 
References


