Structural Studies of O-antigen polysaccharides, Synthesis of $^{13}$C-labelled Oligosaccharides and Conformational Analysis thereof, using NMR Spectroscopy

Ulrika Olsson
To my daughter Tova
Abstract

In order to understand biological processes, to treat and diagnose diseases, find appropriate vaccines and to prevent the outbreak of epidemics, it is essential to obtain more knowledge about carbohydrate structures. This thesis deals with structure and conformation of carbohydrates, analysed by NMR spectroscopy and MD simulations.

In the first two papers, the structures of O-antigen polysaccharides (PS) from two different *E. coli* bacteria were determined using NMR spectroscopy. The O-antigenic PS from *E. coli* O152 (paper I) consists of branched pentasaccharide repeating units, built up of three different carbohydrate residues and a phosphodiester, whilst the repeating unit of the O-antigen from *E. coli* O176 (paper II) is built up of a linear tetrasaccharide consisting of two different monosaccharides.

In papers III and IV, the conformational analysis of different disaccharides is described. Conformational analysis was performed using NMR spectroscopy and MD simulations (paper IV). In paper III four different glucobiosides were studied using coupling constants and Karplus-type relationships. By use of specific $^{13}$C isotopically labelled derivatives, additional coupling constants were obtained and the number of possible torsion angles was reduced by half. In paper IV, we examine the conformations of two disaccharides that are part of an epitope of malignant cells. From NOE and T-ROE experiments, short proton-proton distances around the glycosidic linkage were estimated. Furthermore, interpretation of the extracted coupling constants using Kaplus relationships gave the values of the torsion angles. As in paper III, isotopically labelled compounds were synthesised in order to enhance the sensitivity of the analysis. Finally, MD simulations were performed and the results were compared with results from NMR data.
List of publications

This thesis is based on the following papers, which will be referred to their Roman numerals I-IV.

I  Structural analysis of the O-antigen polysaccharide from *Escherichia coli* O152

II Structural determination of the O-antigenic polysaccharide from the verotoxin producing *Escherichia coli* O176
   U. Olsson, A. Weintraub and G. Widmalm. Accepted for publication in *Carbohydr. Res.*

III Conformational analysis of β-linked glucobiosides based on hetero- and homonuclear couplings across the glycosidic linkage
   U. Olsson, A. S. Serianni and R. Stenutz. Accepted for publication in *J. Phys. Chem.*

IV Conformational flexibility and dynamics of a (1→6)-linked disaccharide related to an oligosaccharide epitope expressed on malignant tumor cells
   *In Manuscript*

Paper I was reprinted with the kind permission from the publisher.
Contents

1 Introduction .........................................................................................................................11
   1.1 General Introduction ..............................................................................................11
   1.2 Structure of carbohydrates ..................................................................................12
   1.3 Bacterial cell lipopolysaccharides ..........................................................................14
      1.3.1 Escherichia coli ..............................................................................................15

2 Synthesis of carbohydrates ................................................................................................17

3 Methods for structural analysis .....................................................................................19
   3.1 Component analysis ..............................................................................................19
   3.2 NMR spectroscopy ................................................................................................20

4 Methods for conformation analysis .............................................................................23
   4.1 Introduction ...........................................................................................................23
   4.2 Computer simulation ............................................................................................24
   4.3 NMR spectroscopy ..............................................................................................25
      4.3.1 Nuclear Overhauser effect ...........................................................................25
      4.3.2 Homo- and heteronuclear coupling constants ..............................................26

5 Structural analysis of the O-antigen polysaccharide from Escherichia coli O152 (paper I) ..........................................................................................................................31
   5.1 Introduction ...........................................................................................................31
   5.2 Result and discussion ............................................................................................31

6 Structural determination of the O-antigenic polysaccharide from the verotoxin producing Escherichia coli O176 (paper II) .........................................................35
   6.1 Introduction ...........................................................................................................35
   6.2 Result and discussion ............................................................................................35

7 Conformational analysis of β-linked glucobiosides based on hetero- and homonuclear couplings across the glycosidic linkage (paper III) ..........................41
   7.1 Introduction ...........................................................................................................41
   7.2 Synthesis ................................................................................................................41
   7.3 Analysis of NMR data ..........................................................................................43
   7.4 Results and discussion ..........................................................................................45
      7.4.1 β-D-Glcp-(1→2)-α-D-GlcpOMe, Methyl α-sophoroside (1) .........................46
      7.4.2 β-D-Glcp-(1→3)-α-D-GlcpOMe, Methyl α-laminaribioside (2) .................47
7.4.3 β-D-Glc-(1→4)-α-D-GlcOMe, Methyl α-cellobioside (3).........................47
7.4.4 β-D-Glc-(1→6)-α-D-GlcOMe, Methyl α-gentiobioside (4).......................48
7.5 Conclusions.............................................................................................................49

8 Conformational flexibility and dynamics of a (1→6)-linked disaccharide
related to an oligosaccharide epitope expressed on malignant tumour cells
(paper IV) ................................................................................................51
  8.1 Introduction..............................................................................................................51
  8.2 Synthesis.................................................................................................................52
  8.3 Results and discussion............................................................................................54
    8.3.1 α-D-ManOMe (25).....................................................................................54
    8.3.2 β-D-GlcNac-(1→6)-α-D-ManOMe (21).......................................................55
    8.3.3 α-D-Manp-(1→6)-α-D-ManpOMe (22).......................................................57

9 Concluding remarks ...............................................................................................59

Acknowledgements .................................................................................................61

References.................................................................................................................64
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgOTf</td>
<td>Silver trifluoromethanesulfonate (silver triflate)</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>ISPA</td>
<td>Isolated spin pair approximation</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular mechanics</td>
</tr>
<tr>
<td>NIS</td>
<td>N-iodosuccinimide</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>PS</td>
<td>Polysaccharide</td>
</tr>
<tr>
<td>rmsd</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>ROE</td>
<td>Rotating-frame Overhauser effect</td>
</tr>
<tr>
<td>TBAHS</td>
<td>Tetrabutylammonium hydrogen sulfate</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofuran</td>
</tr>
<tr>
<td>T-ROE</td>
<td>Transverse rotating-frame Overhauser effect</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 General Introduction

Carbohydrates have several important functions in life. For example, they are structural elements of plants and building blocks of cell surfaces, as well as an essential energy source. Other important functions to life are as markers in cell recognition and for information exchange within living organisms. They are present on cell surfaces as part of glycoproteins and glycolipids and occur with different glycan structures, glycoforms, which their function is dependent on, and the variation is almost infinite. For example, in their specific interplay with extracellular receptors, lectins or antibodies, the interaction is dependent on the carbohydrate structure, both primary and secondary. In order to understand different biological processes involving carbohydrates, it is essential to obtain more information about their structure and how they interact with other biomolecules.

Considerably efforts have been made to determine the conformations, structures and functions of proteins and peptides. Their three-dimensional structures, how the chains fold, are of great interest since their biological functions depends on it. Peptides, for example, are built from a limited set of amino acids in different combinations, but the linkage between the elements is always the same, a rigid and flat peptide bond. The folding of the chain is therefore the main issue, and this is dependent on both the amino acid sequence of the peptide and the possibility to form internal and external hydrogen bonds.

For oligo- and polysaccharides the picture is more complex. The diversity among the saccharides is almost infinite, depending on ring size, substituents, linkage pattern, branching, configuration, conformation and so on. The three dimensional structure of an oligo- or polysaccharide therefore depends on the identity of the different monomers, branching, the possibility to form hydrogen bonds, flexibility of the glycosidic linkages and the number of monomers in the molecule.

This thesis deals with the structure of carbohydrates, both their primary and secondary structure. In papers I and II, the structure of the repeating units of O-antigen polysaccharides (PS) from Escherichia coli (E. coli) bacteria are investigated using component analysis and nuclear magnetic resonance.
(NMR) spectroscopy. The topic of papers III and IV is the three-dimensional (3D) structure of different disaccharides. Specific $^{13}$C labelled disaccharides have been synthesized and conformational analysis performed thereon, using $J_{\text{HH}}$, $J_{\text{CH}}$, and $J_{\text{CC}}$ measured by NMR spectroscopy. The results are compared with molecular dynamics (MD) simulation data (paper IV).

1.2 Structure of carbohydrates

The elementary formula of a carbohydrate is $C_n(H_2O)_y$. Glucose for example has the formula $C_6(H_2O)_6$. A monosaccharide can contain one or more functional groups and may be linked together with other monosaccharides to form larger structures. An oligosaccharide has 2 to 10 monomers linked together whereas larger polymers are called polysaccharides.

Monosaccharides can exist as furanose- or pyranose rings, which are in equilibrium with the open-chain structure (hydrate and keto form). The diversity of the natural occurring carbohydrates is further complicated by having different configurations and conformations. The absolute configuration of a carbohydrate is determined by looking at the stereochemistry of the highest numbered stereogenic centre in the structure (Figure 1.1).

![Figure 1.1](image)

Figure 1.1 (a) D-glucose and (b) L-glucose with the highest numbered stereogenic centre at C5, which is the reference atom for determination of the absolute configuration of the molecule.

The configuration at the anomeric centre is determined by how OH1 in aldoses (OH2 in ketoses) is orientated relative to the reference atom (the stereocentre that determines the absolute configuration) in the Fischer projection. In $\alpha$ anomeric carbohydrates these hydroxyl groups are on the same side in the Fischer projection whilst in $\beta$ compound, the groups are on opposite sides. In aldohexoses in the $^4C_1$ conformation, the $\alpha$ anomer has an axial hydroxyl group at the anomeric centre and the $\beta$ anomer an equatorial hydroxyl group, as illustrated in Figure 1.2.
For reducing carbohydrates, there is an equilibrium between the α- and β-pyranose ring forms and the open structures, the hydrate and the keto form, as well as an equilibrium with α- and β-furanose rings. The anomeric carbon (C1) in aldoses can be seen as a masked carbonyl group which explains the high reactivity of the anomeric carbon compared to other carbons in the structure (see Figure 1.3).

The axial free electron pair (n) of the ring oxygen has the ability to donate electron density into the empty σ* orbital of the C1-O1 bond when trans to it; denoted as an n → σ* interaction. A consequence of this is a stabilizing effect on the α-configuration at the anomeric centre. This phenomenon is referred to as the anomeric effect and is pictured in Figure 1.4.

The hyperconjugation between the free electron pair on the exocyclic anomeric oxygen, O1 and the endocyclic C-O bond is referred to as the exo-anomeric effect (Figure 1.5). As a consequence of this, the torsion angle φ1 has a preference to adopt certain values (section 4.1).
Adjacent electron pairs and/or polar bonds in a structure tend to adopt a gauche arrangement which is referred to as the Gauche effect. Having electron negative groups anti to an electron rich bond results in a $\sigma \rightarrow \sigma^*$ interaction, as in the anomeric effect, which is favourable (Figure 1.6). For example, the torsion angle $\omega$ prefers to adopt certain values as a consequence of this effect (see section 4.1).

![Figure 1.6 The Gauche effect. (a) Electron withdrawing groups (EWGs) anti to each other is energetically unfavourable, whereas (b) the Gauche arrangement is favourable.](image)

### 1.3 Bacterial cell lipopolysaccharides

Bacterial cells can be divided into Gram-positive and Gram-negative. Cells consisting of a single inner cell membrane covered by a thick peptidoglycan layer are denoted Gram-positive whilst bacteria having both an inner and an outer membrane, with a thin peptidoglycan layer in between, are Gram-negative. Both types of cells may be covered by capsular polysaccharides (CPS). Lipopolysaccharides (LPS) are linked to the outer membrane of Gram-negative bacterial cells. The LPS consists of the O-antigen PS, the outer and inner core oligosaccharides and lipid A (Figure 1.7). The LPS is essential for the function of the outer membrane and as a structural
component of the cell. It may play important roles in Gram-negative bacterial infections.\textsuperscript{13}

![The schematic structure of LPS.](image)

**Figure 1.7** The schematic structure of LPS.

\subsection{1.3.1 *Escherichia coli*}

*Escherichia coli* (*E. coli*) is a Gram-negative bacterium that is part of the normal human intestinal flora. Most strains of *E. coli* are quite harmless and only a few of the *E. coli* strains are capable of causing human illness (pathogens). *E. coli* that can cause inflammation of the stomach and bowels are termed enterovirulent *E. coli* and are divided into five classes: enteroinvasive *E. coli* (EIEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC) and enteroaggregative *E. coli* (EAEC).\textsuperscript{14}

The bacterial species is subdivided into serotypes that are designated O:K:H, where O is the somatic O-antigen, K is the capsular antigen and H the flagella antigen. The O-antigen PS is linked to the outer end of the core in the LPS, and can be seen as the “identity” of an *E. coli* bacterium. Knowing the structure of the O-antigen PS, it is possible to track outbreaks of epidemics and their source as well as finding proper pharmaceuticals for their treatment.
2 Synthesis of carbohydrates

The main challenge of carbohydrate synthesis is to control the stereoselectivity of glycosylation reactions as well as finding proper protecting groups that can be specifically removed and are participating or non-participating in the chemical reaction. The glycosylation reaction is performed using a glycosyl donor carrying a leaving group (LG) at the anomeric centre, which is replaced by a glycosyl acceptor, often an alcohol. The general route for the glycosylation reaction is illustrated in Figure 2.1, yielding α- and β-glycosides.

Figure 2.1 The general glycosylation reaction, with exclusion of stereoselectivity.

To achieve stereoselective glycosylation reactions, the choice of protecting groups is important. The stereoselectivity is greatly influenced by the presence of an acyl protecting group at C2, whose electron rich carbonyl oxygen can stabilize the positive charge at the anomeric centre. As shown in Figure 2.2, in the acyloxonium ion intermediate, one of the faces is shielded against a nucleophilic attack. In this way the stereoselectivity of the reaction is controlled.

P=Protecting group
LG=Leaving group
Protecting groups can also influence the reactivity of the reactants. Generally, electron donating protecting groups, for example ethers, increase the reactivity of a donor whilst they decrease the reactivity of an acceptor. The opposite is true when having acyl groups or other electron withdrawing substituents. Additionally, bulky protecting groups can be a steric hinder in the reaction, preventing a nucleophilic attack. Furthermore, the choice of LG and promoter is dependent on the desired reactivity and anomeric configuration as well as how stable they are to the rest of the reaction conditions. Common LG-promoter systems are for example thioglycosides promoted by N-iodosuccinimide (NIS) and silver trifluoromethanesulfonate (AgOTf)15 or glycosyl bromides and chlorides activated by silver salts.16 Time, temperature and solvents are other important parameters when controlling the outcome of a chemical reaction. For example, in a reversible reaction a polar solvent can stabilize a charged intermediate, slowing down the reaction and thereby directing the outcome towards the thermodynamic product. Similarly, an increase of the temperature may speed up the reaction and give more of the kinetic product. However, there are several parameters to have in mind when planning a synthetic route and the outcome is not always easy to predict, every reaction is unique.
3 Methods for structural analysis

3.1 Component analysis

A structural investigation of a PS includes identification of the different residues and determination of their anomeric and absolute configurations, followed by analysis of glycosylation positions as well as the substitution pattern. The study often starts with a component analysis, which gives elementary information, before analysis by NMR spectroscopy.

Sugar analysis is a method that gives the identity and relative quantities of the monomers that build up a PS. The procedure used is as follows: degradation the PS by acid hydrolysis, reduction of the anomeric centres and acetylation of the resulting alditols. The retention times for the formed alditol acetates are then compared with retention times of synthetic standards by gas chromatography (GC). Co-injection with standards can be performed to verify the identity of the carbohydrates. To obtain relative quantities of the identified monomers, the peak areas in the chromatogram are compared.

To determine whether the monosaccharides have the D- or the L-configuration, the PS is hydrolysed, glycosylated with (+)-2-butanol, acetylated and then analysed by GC. The retention times of the acetylated (+)-2-butyl glycosides depends on the absolute configuration (D/L) and identity of the monomers. Results are compared with standards (acetylated (+)-2-butyl glycosides and a mixture of acetylated (+/-)-2-butyl glycosides). Co-injection with standards may be performed to make sure of the identities.

Methylation analysis is a method that gives the substitution sites of the monosaccharides within the PS. In the first step, the non-substituted sites of the PS become methylated. After hydrolysis, reduction and acetylation (as in the sugar analysis), the remaining positions are acetylated. Analysis by gas chromatography-mass spectrometry (GC-MS) gives characteristic fragmentation patterns for the different alditols, depending on the location of the acetyl and methyl groups. Instead of methylation analysis it is possible to determine the glycosylation positions by comparing NMR chemical shifts (see section 3.2).
3.2 NMR spectroscopy

NMR spectroscopy is an important and non-destructive technique used to, for example, analyse and determine the structures of organic molecules. A sample tube is placed in a strong external magnetic field and a target isotope (a nucleus with the right magnetic properties, for example $^1$H and $^{13}$C) within the sample experience this external field. The outcome of the experiment gives information of the surroundings, for example, neighbouring nuclei, bonds, angles and so forth. But of course, much more information is available from this technique and the outcome is dependent on the experiment carried out and the question asked.

A range of different 1D and multi-dimensional ($n$D) experiments are available for the assignment of the primary structure of a molecule. For simple problems it is often enough to perform 1D $^1$H and $^{13}$C experiments to obtain the chemical shifts and coupling constants. When analysing larger structures, the chemical shifts have a higher degree of overlap in the 1D spectrum and it becomes necessary to use advanced 1D and two dimensional (2D) hetero- and homonuclear experiments to determine the structure.

In carbohydrate research, it can be useful to know that in a 1D $^1$H-NMR spectrum for common saccharides, the signals from the anomeric protons appear in the range 4.3-5.6 ppm whilst ring proton resonances appear in the range of 3.2-4.5 ppm. Furthermore, chemical shifts from methyl groups of deoxy carbohydrates and acetyl groups appear in the range 1.1-2.2 ppm. Of course there are deviations from the expected chemical shifts, depending on the substitution pattern and other factors. Homo- ($J_{HH}$) and heteronuclear coupling constants ($J_{CH}$) give information about the number and identity of neighbouring atoms, conformation and bond angles within the molecule. The coupling constants also reveal the anomeric configuration. For example, the size of $J_{H1,H2}$ for $\alpha$-D-pyranoses with the gluco-configuration in the $\delta$C1 chair form is about 4 Hz and for the $\beta$ compound it is around 8 Hz. Likewise, the size of $J_{C1,H1}$ is around 170 and 160 Hz for the $\alpha$ and $\beta$ configurations, respectively. Information about glycosylation positions within a carbohydrate residue is obtained by comparing chemical shifts with an analogous unsubstituted reference compound.$^{17}$

The assignment of the different nuclei in an oligo- or polysaccharide is performed by combining a set of different homo- and heteronuclear 2D NMR experiments.$^{18}$ For example, correlations between adjacent protons are obtained in $^1$H,$^1$H-COSY and $^1$H,$^1$H-TOCSY experiments. Increasing the mixing time in the $^1$H,$^1$H-TOCSY experiment allows the magnetization to travel further in the spin system and it is possible to correlate all the protons within the residue. Both intra- and inter-residual cross relaxation correlations from protons close in space are obtained in the $^1$H,$^1$H-NOESY
experiment, giving information about sequence and conformations. Likewise, information concerning the linkage pattern can be revealed from the $^1$H,$^{13}$C-HMBC experiment through \textit{inter-} and \textit{intra-residue} long-range correlations two to three bonds away. In the $^1$H,$^{13}$C-HSQC spectrum, protons are correlated to their corresponding carbon and if $^1J_{C,H}$ couplings are desired, a coupled $^1$H,$^{13}$C-HSQC experiment can be performed.
4 Methods for conformation analysis

4.1 Introduction

A molecule in solution is a dynamic system; it moves, rotates and vibrates. These flexible parameters are dependent on the shape and size of the molecule and the temperature and the viscosity of the solvent. The conformations of a molecule describe its 3D structure. For carbohydrates, the conformations at the glycosidic linkage, determined by the glycosidic torsion angles $\phi$ (phi) and $\psi$ (psi), are the most important. For 6-linked carbohydrates the torsion angle $\omega$ (omega) also has to be considered. $\phi_H$ is defined as H1'-C1'-OX-CX, $\psi_H$ as C1'-OX-CX-HX and $\omega$ as O5-C5-C6-O6 (Figure 4.1).

![Figure 4.1](image.png)

**Figure 4.1** The torsion angles $\phi$, $\psi$ and $\omega$ in a 6-linked disaccharide.

As the conformations have different energies, some of them will be more populated than others. In general, the value of $\phi$ is determined by the exo-anomeric effect (see section 1.2) and $\psi$ by steric effects and the ability to form internal hydrogen bonds. Therefore, $\phi_H$ values are around -60° for $\alpha$-D and $\beta$-L hexoses and about ±60° for $\alpha$-L and $\beta$-D hexoses, whereas $\psi_H$ values usually lie between -50° and +50°. The three staggered rotamers of $\omega$ are shown in Figure 4.2 and are referred to as $gt$ ($\omega \approx +65°$), $gg$ ($\omega \approx -65°$) and $tg$ ($\omega \approx 180°$). The value of $\omega$ is dependent on steric and/or electronic effects where the electronic effects are explained by the Gauche effect (section 1.2).
Common techniques used for conformation analysis of carbohydrates are X-ray crystallography, NMR spectroscopy and computer simulations. Unfortunately, crystallographic data only provides information about the structures in the solid state, whilst biological processes involving carbohydrates are usually dynamic systems in solution (water). To include the flexibility and dynamics, a combination of NMR spectroscopy and computer simulations may be used, which is a powerful combination for this purpose.

### 4.2 Computer simulation

Molecular mechanics (MM) is a computer simulation method where a force field is applied to a system in order to find low energy conformations. The total energy function ($E_{tot}$) is the sum of a number of different energy terms, for example bond length, bond angle and van der Waals forces. The molecule is given initial coordinates representing a conformation, which then are adjusted to minimize the calculated energy. The simulation proceeds as long as the newly calculated conformation has a lower energy than the previous one. The MM simulation can be performed with a systematic method (grid search) or a random search method (Metropolis Monte Carlo).

However, it is worth noting that biological systems are dynamic. Therefore, a molecular dynamics (MD) simulation may be the method of choice. This method gives a supply of kinetic energy to the system and the movements of the molecule are recorded. The simulation can be performed in vacuum, with implicit water (continuum solvent model), or explicit water (with water molecules). Simulations in vacuum or with implicit water demand less computer resources and are therefore less time consuming. In order to simulate actual biological conditions it is, however, preferable to perform the study with explicit water. The result is a time trajectory of the different conformations (see Figure 8.6) which can be converted to a population distribution function.
4.3 NMR spectroscopy

4.3.1 Nuclear Overhauser effect

For conformational analysis, a few NMR methods are preferred. The nuclear Overhauser effect (NOE) is often used for this purpose, where short $^1$H-$^1$H distances are measured via dipolar direct trough-space magnetic interactions.\textsuperscript{22, 23} The result of a NOE experiment depends on the orientation and distances of the protons and provides an indication about the molecular 3D structure.

The NOE is the change in spin population leading to decrease or increase of signal intensity when a nearby proton is stimulated to saturation (steady-state experiment) or inversion (transient experiment). The size and sign of the NOE is dependent on not only conformation but also on the correlation time ($\tau_C$), the time it takes for a molecule to rotate one radian in any direction. For medium size molecules the NOE can be zero and no correlations are detected.\textsuperscript{24} To overcome the problem, the temperature or the strength of the magnetic field can be changed. However, the rotating-frame Overhauser Effect (ROE) experiment always remains positive and provides an alternative to NOE. Unfortunately, the ROE experiment has a drawback: TOCSY transfer during the applied spin-lock gives rise to false ROE signals.\textsuperscript{25} The transverse ROE (T-ROE) experiment is developed to eliminate the problem with TOCSY artefacts and is often preferred over the ROE experiment, even though it gives a lower signal to noise ratio.\textsuperscript{26}

![Figure 4.3](image)

**Figure 4.3** The transient NOE effect ($\eta_{\text{transient}}$) plotted against $\omega \tau_C$ for transient NOE, ROE and T-ROE experiments, where $\tau_C$ is the correlation time of the molecule and $\omega$ the Larmor frequency of a proton nucleus.\textsuperscript{27}

For the measurement of proton-proton distances, one proton resonance is selectively excited in an array of experiments with different mixing times. Data from the experiments give build-up curves (see Figure 8.5), where the
initial slope corresponds to $\sigma$ (build-up rate or cross-relaxation rate). The distance between two nuclei, $r_{ij}$, is then calculated from $\sigma$ using the isolated spin pair approximation (ISPA)

$$r_{ij} = r_{\text{ref}} \left( \frac{\sigma_{\text{ref}}}{\sigma_{ij}} \right)^6$$  \hspace{1cm} (4.1)

where $\sigma_{ij}$ is the NOE build-up rate between nuclei $i$ and $j$, $r_{\text{ref}}$ is a known reference distance within the molecule and $\sigma_{\text{ref}}$ is the build-up rate of the reference atoms.28, 29

**4.3.2 Homo- and heteronuclear coupling constants**

Scalar couplings (couplings through bonds) can give conformational information via Karplus-type relationships.30,31 For example, in hexopyranoses, the size of the three bond H5-H6 coupling constant is correlated with the highly flexible hydroxymethyl torsion angle ($\omega$). The relationship is provided by Equations 4.2 and 4.3, which are also illustrated in Figure 4.4.32 The conformation of $\omega$ is interpreted as population distributions, namely as the three staggered rotamers $gt$, $gg$ and $tg$ (see Figure 4.2). Calculation of the population distributions from $^3J_{\text{H5,H6}}$ is further discussed in Chapter 7.

$$^3J_{\text{H5,H6,R}} = 5.08 + 0.47\cos(\omega) + 0.90\sin(\omega) - 0.12\cos(2\omega) + 4.86\sin(2\omega)$$  \hspace{1cm} (4.2)

$$^3J_{\text{H5,H6,S}} = 4.92 - 1.29\cos(\omega) + 0.05\sin(\omega) + 4.58\cos(2\omega) + 0.07\sin(2\omega)$$  \hspace{1cm} (4.3)

![Figure 4.4](image-url) **Figure 4.4** Karplus-type equations where $\omega$ is plotted as a function of $^3J_{\text{H5,H6,R}}$ (▲) and $^3J_{\text{H5,H6,S}}$ (■), respectively. One value of $J$ corresponds to up to four different values of the torsion angle.

Unfortunately, due to the periodicity of Karplus-type equations, one value of $J$ corresponds to up to four different values for a torsion angle. Hence, additional experiments are needed in order to reduce the number of possible torsion angles in the analysis. A heteronuclear long-range experiment gives...
heteronuclear coupling constants, two or three bonds away. For example, it is possible to measure $^{3}J_{C4,H6pro-S}$ and $^{3}J_{C4,H6pro-S}$ which show a Karplus dependency on $\omega$ as well (shown in Equations 4.4 and 4.5, respectively). The equations are shown in Figure 4.5.

$$^{3}J_{C4,H6pro-S} = 3.60 + 0.50\cos(\omega) + 0.06\cos(2\omega) - 0.13\sin(\omega) - 3.46\sin(2\omega)$$  \hspace{1cm} (4.4)

$$^{3}J_{C4,H6pro-R} = 3.58 + 0.11\cos(\omega) + 3.50\cos(2\omega) + 0.35\sin(\omega) - 0.57\sin(2\omega)$$  \hspace{1cm} (4.5)

**Figure 4.5** Illustration of the Karplus-type dependence of $^{3}J_{C4,H6pro-R}$ (▲) and $^{3}J_{C4,H6pro-S}$ (■).

Another result that can be obtained from a long-range experiment is the value of the trans-glycosidic $^{3}J_{COCH}$ couplings, which have a Karplus-type relationship to the glycosidic torsion angles $\phi$ and $\psi$ as shown below in Equation 4.6 and Figure 4.6.

$$^{3}J_{COCH} = 7.49 \cos^{2} \theta - 0.96 \cos \theta + 0.15$$  \hspace{1cm} (4.6)

**Figure 4.6** The Karplus-relationship for the trans-glycosidic long-range $^{3}J_{COCH}$ and the value of a torsion angle ($\phi$ or $\psi$).

Apart from the Karplus relationships discussed above, there is still a desire for more data to enhance the reliability of the conformation analysis. In this
respect, specifically $^{13}$C isotopically labelled saccharides may be used. An advantage of this approach is the additional $J_{C,H}$ and $J_{C,C}$ coupling obtained, which can be measured directly from 1D spectra. The outcome is, of course, controlled by the choice of labelling positions, but more importantly, a number of different Karplus equations have been reported for these different carbon-carbon and carbon-proton coupling constants. For example, the relationship between the $\omega$ torsion and the two bond C4-C6 coupling constant is shown below in Equation 4.7 (see Figure 8.3).\textsuperscript{33}

$$2J_{C4,C6} = 1.36 + 1.03\cos(\omega) + 2.26\sin(\omega) \quad (4.7)$$

Furthermore, Equation 4.8 describes the Karplus dependency of different $^{3}J_{COCC}$ which are sensitive to the torsion angles $\phi_{C}$ and $\psi_{C}$ (Figure 4.7).\textsuperscript{37}

$$^{3}J_{COCC} = 3.49 \cos^{2} \theta + 0.16 \quad (4.8)$$

Figure 4.7 Karplus-relationship for the size of trans-glycosidic long-range $^{3}J_{COCC}$ and the value of a torsion angle ($\phi_{C}$ or $\psi_{C}$).

In Figure 4.8, examples of different $J_{C,C}$ are demonstrated, obtained by using isotopic labelling at different positions in disaccharides.
Figure 4.8 Examples of additional $J_{C,C}$ obtained when (a) a (1→3)-linked disaccharide is [1'-13C]-labelled or (b) [2'-13C] labelled, (c) when a (1→6)-linked disaccharide is [1'-13C]-labelled or (d) [6-13C]-labelled.

In conclusion, a number of additional hetero- and homonuclear coupling constants are obtained by using $^{13}$C-labelled compounds. This can be used to enhance the reliability of the conformational analysis.
5 Structural analysis of the O-antigen polysaccharide from *Escherichia coli* O152 (paper I)

5.1 Introduction

*E. coli* O152 is a Gram negative enterovirulent *E. coli* classified as enteroinvasive *E. coli* (EIEC), which is known to cause small outbreaks of dysentery. *E. coli* antigens are often closely related or identical to *Shigella* O-antigens and Cheasty *et al.* reported the antigenic relationship between *E. coli* O152 and the provincial *Shigella* serovar 3341.55; they revealed identical cross-reactivity. The parent *Shigella* O-antigen was later defined as *Shigella dysenteriae* 12.

5.2 Result and discussion

The structure of the repeating unit of the O-antigen PS from *E. coli* O152 was analysed using component analysis and NMR spectroscopy. The $^1$H and $^{13}$C chemical shifts of the PS were assigned using 1D NMR spectroscopy together with a combination of homo- and heteronuclear 2D NMR experiments.

Sugar and absolute configuration analysis revealed that the repeating unit consisted of three different monosaccharides: D-glucose (D-Glc), N-Acetyl-2-amino-2-deoxy-D-glucopyranose (D-GlcNAc) and L-rhamnose (L-Rha) in the ratio of 30:44:21. This is in agreement with information from the $^1$H-NMR spectrum, showing five anomeric protons having chemical shifts of $\delta_{\text{H}}$ 5.37 (A), 5.27 (B), 4.91 (C), 4.75 (D) and 4.71 (E) (Figure 5.1).

In the $^1$H,$^{13}$C-HSQC spectrum, the five anomeric proton signals are correlated to their corresponding carbon signals. The anomic $^{13}$C chemical shifts are in the range 95.0-102.2 ppm. The resonance at $\delta_{\text{C}}$ 17.7 is correlated to the peak at $\delta_{\text{H}}$ 1.33 which is a doublet with a coupling constant of 6.1 Hz in the 1D $^1$H-NMR spectrum and assigned as the methyl signal of H6 in the Rha residue. The four resonances at $\delta_{\text{C}}$ 23.3, 23.5, 174.9 and 175.2 ppm originate from the two N-acetyl groups in the structure. Finally, besides the
annotated carbon resonances above, there are 25 peaks in the $^{13}$C NMR spectrum identified as ring carbons of the PS. The total assignment of the resonances and conclusions about the ring size was performed analyzing $^1$H,$^1$H-COSY, $^1$H,$^1$H-TOCSY and $^1$H,$^1$H-COXS spectrum as well as the $^1$H,$^1$H-HMBC spectrum.

Figure 5.1 The $^1$H NMR spectrum of the O-antigen PS from Escherichia coli O152 with the anomeric proton resonances denoted as A-E.

The question of the anomeric configurations of the monomers was solved by looking at anomeric $^1$H and $^{13}$C chemical shifts as well as the size of $^3$J$_{H1,H2}$, measured in the proton spectrum, and $^1$J$_{C1,H1}$, obtained from the coupled $^1$H,$^1$C-HSQC spectrum. To sum up: $^3$J$_{H1,H2}$ for residue A is 3.4 Hz and for B 3.8 Hz indicating $\alpha$-configuration. This assumption was confirmed by the size of $^1$J$_{C1,H1}$ being 178 and 176 Hz respectively. Residues D and E are recognised as $\beta$-anomeric carbohydrates having the $J$ values of 8.2/167 and 7.7/165 Hz. L-Rha, residue C, is identified as $\beta$, having $^1$J$_{C1,H1}$=164 Hz. The $^3$J$_{H1,H2}$ coupling constant was not resolved but the signal width at half-height is 2.6 Hz.

Summarising the information so far, the identity of the residues in the repeating unit are in agreement with $\alpha$-D-GlcNAc (A), $\alpha$-D-Glcp (B), $\beta$-L-Rhap (C), $\beta$-D-GlcpNAc (D) and $\beta$-D-Glcp (E). The anomeric proton of residue A has an extra coupling constant of 7.1 Hz, which is the size of a $^3$J$_{H1,P}$ coupling constant. The 1D $^{31}$P NMR experiment reviled a resonance at $\delta$P = -1.1, a chemical shift characteristic of a phosphodiester.

Information about the glycosylation positions was obtained solely via NMR analysis. That is, by comparing the $^{13}$C chemical shifts of the residues in the PS with known chemical shifts of unsubstituted monosaccharides (Table 5.1). Residue A, where $\Delta$C3 and $\Delta$C4 is $\sim$4 ppm, is herein reported as 3- and 4-disubstituted. Likewise, B is 6-substituted ($\Delta$C6=4.4), D is 3-
substituted ($\Delta \delta_{C3}=4.7$) and E is 2-substituted ($\Delta \delta_{C2}=3.5$). Residue C, $\beta$-L-Rhap, is assumed to be terminal because no major glycosylation shifts are observed for C2 to C4.

Table 5.1 $^{13}$C NMR chemical shifts of the O-antigen PS from *E. coli* O152. Differences compared to reference monosaccharides are reported in parentheses.

<table>
<thead>
<tr>
<th>Carbohydrate residue</th>
<th>$^{13}$C</th>
</tr>
</thead>
<tbody>
<tr>
<td>→3,4)-α-D-GlcNAc-(1→ A</td>
<td>95.0 54.2 75.5 75.7 72.4 61.2</td>
</tr>
<tr>
<td>(3.2) (−0.8) (3.8) (4.4) (−0.1) (−0.6)</td>
<td></td>
</tr>
<tr>
<td>→6)-α-D-GlcNAc-(1→ B</td>
<td>98.9 72.4 73.5 69.9 71.6 65.2</td>
</tr>
<tr>
<td>(5.9) (−0.1) (−0.3) (−0.8) (−0.8) (3.4)</td>
<td></td>
</tr>
<tr>
<td>$\beta$-L-Rhap-(1→ C</td>
<td>102.2 71.1 73.5 73.0 73.0 17.7</td>
</tr>
<tr>
<td>(7.8) (−1.1) (−0.3) (0.2) (0.2) (0.1)</td>
<td></td>
</tr>
<tr>
<td>→3)-β-D-GlcNAc-(1→ D</td>
<td>101.0 56.9 79.5 69.3 76.2 61.9</td>
</tr>
<tr>
<td>(5.2) (−1.0) (4.7) (−1.8) (−0.6) (0.1)</td>
<td></td>
</tr>
<tr>
<td>→2)-β-D-GlcNAc-(1→ E</td>
<td>101.6 78.7 75.7 70.7 76.8 61.6</td>
</tr>
<tr>
<td>(4.8) (3.5) (−1.1) (0.0) (0.0) (−0.2)</td>
<td></td>
</tr>
</tbody>
</table>

Via heteronuclear trans-glycosidic long-range correlations, visible in the $^{1}$H, $^{13}$C-HMBC spectrum, the substitution pattern was revealed and then further confirmed by the $^{1}$H,$^{1}$H-NOESY experiment. There are no trans-glycosidic NOEs or long-range correlations observed for residue A. The explanation for that is the phosphodiester linked at the anomeric position. On the other hand, residue B has correlations to E2, residue C is linked to residue A proven by cross-peaks to A4, residue D correlates to A3 and finally, E has correlations to D3. From this, it seems like the phosphodiester is linked in between residue A and B. This was further proven by the $^{1}$H,$^{31}$P-COSY experiment where the phosphorus resonance shows cross-peaks to H1A, H2A and H6B (Figure 5.2).

Figure 5.2 $^{1}$H,$^{31}$P-COSY spectrum of the O-antigen PS from *E. coli* O152.
Furthermore, $^{13}$C, $^{31}$P couplings between phosphorus and C1A, C2A, C5B and C6B were observed in the 1D $^{13}$C NMR spectrum, which further corroborates the proposed structure of the PS. To conclude, the structure of the repeating unit of the O-antigenic PS from *E. coli* O152 is:

$$\text{C}$$

$\beta$-l-Rhap-(1→4)

$\rightarrow$3-$\alpha$-d-GlcpNac-(1-P→6)-$\alpha$-d-Glcp-(1→2)$\beta$-d-Glcp-(1→3)$\beta$-d-GlcpNac-(1→

\[\text{A} \quad \text{B} \quad \text{E} \quad \text{D}\]

Figure 5.3 The structure of the repeating unit PS from *E. coli* O152.
6 Structural determination of the O-antigenic polysaccharide from the verotoxin producing Escherichia coli O176 (paper II)

6.1 Introduction

*E. coli* O176 is classified as an *enterohemorrhagic E. coli* (EHEC), also called Shiga toxin producing *E. coli* (STEC) or Verotoxigenic *E. coli* (VTEC). This highly pathogenic class of *E. coli*, with low infective dose, causes watery and bloody diarrhoea, homolytic uremic syndrome (HUS), thrombotic thrombocytopenic purpura (TTP) or even death. The infection spreads via contaminated food and water or person-to-person.\(^{40,61}\)

6.2 Result and discussion

Investigation of the repeating unit of the O-antigen PS from *E. coli* O176 included component analysis and NMR spectroscopy. The component analysis showed *d*-mannose (*d*-Man) and *N*-Acetyl-2-amino-2-deoxy-*d*-galactopyranose (*d*-GalNAc), in the ratio of 4.76:1, as components of the PS. Furthermore, peaks arising from known core residues were identified in the sugar analysis GC chromatogram: glucose (6%), galactose (2%), 2-deoxy-*N*-acetyl-glucose (2%) and heptose (2%).\(^{42}\)

The 1D \(^1\)H NMR spectrum of the PS shows four distinct signals in the anomeric region, having the chemical shifts 5.35 (A), 5.33 (B), 5.06 (C) and 4.81 (D) ppm, indicating that the repeating unit is a tetrasaccharide. In the \(^1\)H,\(^{13}\)C-HSQC NMR spectrum, the anomeric protons show cross-peaks to their corresponding carbons (Figure 6.1). Besides the anomeric signals and resonances from the ring atoms, there is a cross-peak at \(\delta_H 4.31/\delta_C 49.97\), showing that an *N*-acetylated aminosugar is present in the structure.
Figure 6.1 The $^1$H, $^{13}$C-HSQC NMR spectrum of the repeating unit of the O-antigen PS from *E. coli* O176. The residues are labelled A-D with respect to their anomeric chemical shifts, starting with the anomer having the highest $\delta_H$.

The chemical shifts of the tetrasaccharide were assigned using 1D $^1$H and $^{13}$C NMR experiments in combination with different 2D $^1$H and $^{13}$C homo- and heteronuclear NMR techniques (see Table 6.1). Observations show that the residues A, C and D have the *manno*-configuration whilst residue B has *galacto*-configuration, which is consistent with results from the component analysis. Finally, anomeric homo- and heteronuclear coupling constants, $^3J_{H1,H2}$ and $^1J_{C1,H1}$, established the anomeric configurations of the monomers. $^3J_{H1,H2}$ of the *manno*-residues were not resolved at the current field and therefore not measurable. On the other hand, $^1J_{C1,H1}$ shows that residue A and C are $\alpha$-linked since they have values around 175 Hz. Moreover, D is $\beta$-linked and has a coupling constant of 160 Hz. The aminosugar, B, has corresponding coupling constants of 3.5 and 177 Hz and is therefore denoted as an $\alpha$-pyranoside.

The substitution positions were determined solely by identifying $^{13}$C glycosylation shifts in the polymer (Table 6.2). Both residues A and D have significant glycosylation shifts of 7.27 and 4.60 Hz, respectively, at position 2 and are therefore 2-substituted. Residue B has a $\Delta\delta_C$ of 9 Hz which means that it is 3-glycosylated whilst residue C has a glycosylation shift of 6.67 ppm at C4.
Table 6.1 $^1$H and $^{13}$C chemical shifts (ppm) from the O-antigen PS of *E. coli* O176. $^3J_{HH}$ and $^3J_{HC}$ are given in Hz in square brackets and braces, respectively.

<table>
<thead>
<tr>
<th>Carbohydrate residue</th>
<th>$^1$H</th>
<th>$^1$C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>α-D-Manp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5.35</td>
<td>4.12</td>
</tr>
<tr>
<td></td>
<td>100.46</td>
<td>78.96</td>
</tr>
<tr>
<td></td>
<td>[n.r.][175]</td>
<td>[n.r.][177]</td>
</tr>
<tr>
<td>α-D-GalpNac$^\ddagger$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5.33</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td>99.18</td>
<td>49.97</td>
</tr>
<tr>
<td>α-D-Manp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5.06</td>
<td>4.01</td>
</tr>
<tr>
<td></td>
<td>102.69</td>
<td>71.15</td>
</tr>
<tr>
<td>β-D-Manp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4.81</td>
<td>4.01</td>
</tr>
<tr>
<td></td>
<td>102.20</td>
<td>76.73</td>
</tr>
</tbody>
</table>

*n.r.* = not resolved. $^\ddagger$Chemical shifts for NAc are $\delta_H$ 2.06, $\delta_C$ 22.90 and 174.84.

Table 6.2 Chemical shift differences, $\Delta\delta_C$, for the O-antigen PS from *E. coli* O176, as compared to reference monosaccharides.

<table>
<thead>
<tr>
<th>Carbohydrate residue</th>
<th>$\Delta\delta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>→2)-α-D-Manp (1→</td>
<td>A</td>
</tr>
<tr>
<td>→3)-α-D-GalpNac (1→</td>
<td>B</td>
</tr>
<tr>
<td>→4)-α-D-Manp (1→</td>
<td>C</td>
</tr>
<tr>
<td>→2)-β-D-Manp (1→</td>
<td>D</td>
</tr>
</tbody>
</table>

Trans-glycosidic three-bond heteronuclear coupling correlations from the $^1$H,$^{13}$C-HMBC NMR experiment were used to determine the sequence of the carbohydrate residues in the PS. Residue A has long-range correlations to residue D, residue B to C, C to A, and finally, D shows cross-peaks to residue B. In conclusion, the O-antigen tetrasaccharide repeating unit from *E. coli* O176 has the structure

→4)-α-D-Manp (1→2)-α-D-Manp (1→2)-β-D-Manp (1→3)-α-D-GalpNac (1→ C A D B

which was confirmed by the $^1$H,$^1$H-NOESY experiment, showing corresponding cross-peaks at the glycosidic linkages (Figure 6.2).
Figure 6.2 The anomeric region in the $^1$H, $^1$H-NOESY ($\tau_{\text{mix}}=50\text{ms}$) spectrum of the O-antigen PS from *E. coli* O176.

As pictured in Figure 6.2, there is a strong NOE correlation between H-1 in A and H-5 in residue C, which means that they are close in space. Figure 6.3 (below) shows parts of the anomeric region in the $^1$H,$^1$H-TOCSY spectrum ($\tau_{\text{mix}}=30\text{ ms}$). At a certain point, low intensity cross-peaks show up, having similar chemical shifts and cross-peaks to residues A and C and are denoted $A'$ and $C'$, respectively (see figure 6.3). Consequently, $A'$ and $C'$ resonances are assumed to arise from the terminal residues of the PS.

Figure 6.3 Part of the anomeric region of the $^1$H, $^1$H-TOCSY spectrum ($\tau_{\text{mix}}=30\text{ ms}$). $H_1A'/H_2A'$ and $H_1C'/H_2C'$ are cross-peaks from terminal residues of the O-antigenic PS from *E. coli* O176.

Identification of the terminal residues reveals the number of repeating units in the PS. Integration of the A and A' resonances in the $^1$H NMR spectrum showed that the PS consists of about 10 units. Finally, the repeating unit of the O-antigen PS from *E. coli* O176 is given as:

$$C'-A'-D'-B'-[C-A-D-B]_9$$

Interestingly, there are PS structures from other *E. coli* bacteria that resemble the structure of the PS from O176. A striking example of this is the structure of the O-antigenic PS from enteropathogenic *E. coli* strain 73-1, isolated from a young child with diarrhoea in Chile.\textsuperscript{43} Referring the structure of O176, the only difference between these two structures is a $\alpha$-D-Glep residue linked to residue D:
Another similar structure found is an O-antigenic PS from the enteroaggregative *E. coli* O44:H18, found in a Peruvian infant with diarrhoea:\(^\text{44}\)

\[
\begin{align*}
\alpha-D-Glc p(1\rightarrow4) \rightarrow &-\alpha-D-Man p(1\rightarrow2)-\alpha-D-Man p(1\rightarrow2)-\beta-D-Man p(1\rightarrow3)-\alpha-D-Gal pNAc(1\rightarrow6)
\end{align*}
\]

In this structure, the differences lie in residue C being 6-substituted and in residue B having a *gluco*-configuration instead of *galacto*-configuration. As in the previous example, residue D is substituted with a $\alpha-D-Glc p$. Wang *et al.* studied a group of *E. coli* O-antigens with the same backbone as for *E. coli* O44:H18.\(^\text{45}\) The authors explain that all of the *E. coli* in the study shares a common gene cluster which encodes for the biosynthesis of the O-antigens. The small differences between the structures are due to genetic variations in their respective O-antigen gene clusters.

Additionally, Vingradov *et al.* solved the structure of the repeating unit PS from O-specific *Salmonella cerro* a few years ago which shows to be identical with the O-antigenic structure from *E. coli* O176.\(^\text{46}\)

*Figure 6.4* The structure of the O-antigenic repeating unit PS from *Salmonella cerro* and *E. coli* O176.
7 Conformational analysis of β-linked glucobiosides based on hetero- and homonuclear couplings across the glycosidic linkage (paper III)

7.1 Introduction

In this project, the focus has been on analysing the torsion angles around the glycosidic linkage of four β-linked glucobiosides, using two different methods, described below. The theory behind the two methods is based on homo- and heteronuclear coupling constants. The analysed compounds are methyl α-sophoroside (1, β-D-Glc p-(1→2)-α-D-Glc pOMe, g2g), methyl α-laminaribioside (2, β-D-Glc p-(1→3)-α-D-Glc pOMe, g3g), methyl α-cellobioside (3, β-D-Glc p-(1→4)-α-D-Glc pOMe, g4g) and α-gentiobioside (4, β-D-Glc p-(1→6)-α-D-Glc pOMe, g6g). The four glucobiosides were synthesized in three different sets each; unlabelled disaccharides, and specifically 13C labelled at either the C1’ or C2’ positions. The unlabelled derivatives were used in the assignment of the NMR chemical shifts for each structure as well as for extraction of the proton-proton couplings. From the isotopically labelled disaccharides, different \( J_{C,H} \) and \( J_{C,C} \) were measured from 1D \( ^1H \) and \( ^13C \) NMR spectrums.

7.2 Synthesis

The syntheses of disaccharides 1-4 were performed by coupling of the donor (5) in a straightforward stereoselective glycosylation reaction, promoted by NIS/AgOTf, with acceptors 6, 7, 8 and 9, respectively (Figure 7.1).

![Figure 7.1](image-url) Figure 7.1 The acceptors 6, 7, 8 and 9 used in the synthesis of glucobiosides 1-4.
Due to the presence of the participating acetyl group in position 2 of the donor, the outcome of the glycosylation reactions is shifted towards the β-product. Another advantage of having the acetyl group in position 2 is the possibility of performing selective deprotection in a future synthesis of trisaccharides, with 5 as the central sugar unit.

The donor 5 was prepared from D-glucose (10), [1-13C]- or [2-13C]-D-glucose, respectively, as starting material (Scheme 7.1). To achieve the β-anomeric donor 5, the first two steps in the synthetic route was a stereoselective acetylation giving 11, followed by thioglycosylation, using BF3·OEt2 as a promoter, giving 12 as the product. Furthermore, selective protection of the other positions in the sugar was performed.

![Scheme 7.1](image)

**Scheme 7.1** a. NaOAc, Ac2O, 150 °C; b. EtSH, BF3·OEt2, CH2Cl2, 0-25 °C; c. NaOMe, MeOH; d. PhCH(OMe)2, D-(+)-10-camphorsulfonic acid, DMF, 60 °C; e. 1. [CH3(CH2)3]2SnO, MeOH, Δ, 2. CsF, BnBr, DMF; f. Pyridine, Ac2O.

Acceptors 6 and 7 were made in the same phase-transfer reaction from methyl 4,6-benzylidene α-D-glucopyranoside (16) (Scheme 7.2). The organic catalyst, tetrabutylammonium hydrogen sulfate (TBAHS), transports the alkoxide to the organic phase where it becomes benzylated by benzyl bromide (BnBr). The regioisomeric product ratio between 6 and 7 was 1:3.

![Scheme 7.2](image)

**Scheme 7.2** a. BnBr, TBAHS, NaOH (aq), reflux.
Acceptors 8 and 9 were synthesised from the starting material 17\textsuperscript{54} by detritylation under acidic conditions (Scheme 7.3). Compound 9 was the expected product but there is a tendency for acetyl groups to migrate from secondary to primary positions,\textsuperscript{55} which in this case resulted in formation of 8. This was taken advantage of because both 8 and 9 could be used in the subsequent bioside synthesis. Due to the formation of acyl migration product the reaction was not run to completion.

\begin{center}
\begin{tikzpicture}
\node (a) at (0,0) {17};
\node (b) at (2,0) {8};
\node (c) at (4,0) {9};
\node (d) at (2,1) {32\%};
\node (e) at (4,1) {17\%};
\draw[->] (a) -- node[above] {a} (b);
\draw[->] (b.east) -- node[above] {32\%} (c);
\end{tikzpicture}
\end{center}

\textbf{Scheme 7.3} a. HOAc, H\textsubscript{2}O, 80 °C.

Deprotection of the obtained disaccharides was performed by hydrogenation to remove the benzylidene and benzyl protecting groups, followed by deacetylation under basic conditions. The synthetic route for a disaccharide (2) is illustrated in Scheme 7.4. The biosides were purified by size exclusion chromatography and subsequently lyophilised from D\textsubscript{2}O before analysis by NMR spectroscopy.

\begin{center}
\begin{tikzpicture}
\node (a) at (0,0) {5};
\node (b) at (2,0) {7};
\node (c) at (4,0) {18};
\node (d) at (2,1) {79\%};
\node (e) at (4,1) {82\%};
\draw[->] (a) -- node[above] {a} (b);
\draw[->] (b) -- node[above] {79\%} (c);
\draw[->] (b) -- node[above] {b} (c);
\end{tikzpicture}
\end{center}

\textbf{Scheme 7.4} a. AgOTf, NIS, CH\textsubscript{2}Cl\textsubscript{2}; b. 1. H\textsubscript{2}, Pd(OH)\textsubscript{2}/C, EtOH:EtOAc:HOAc 4:2:1, 2. NaOMe, MeOH.

\subsection*{7.3 Analysis of NMR data}

A spin simulation program (PERCH) was used to extract the coupling constants from the \textsuperscript{1}H NMR spectrum of the different \textsuperscript{13}C isotopically labelled β-biosides.\textsuperscript{56,57} The initial chemical shifts and \(J\) values were
optimized so that the difference between the experimental and simulated spectra was minimized. The extracted $^{3}J_{\text{H,H}}$ and $^{3}J_{\text{C,H}}$ from the $^{1}$H spectra and $^{3}J_{\text{C,C}}$ from $^{13}$C spectra were then used in the analysis. The analysis is based on two methods. In method A, the root mean squared deviation ($rmsd$), between the different experimental and calculated coupling constants, is considered (Equation 7.1).

$$rmsd = \sqrt{\sum x^2}$$ (7.1)

$x$ = the deviation between the experimental and the calculated value of the coupling constants.

A small $rmsd$ value means a higher possibility for a torsion angle to adopt the value in question. On the other hand, in this approach, the torsions are assumed to be rigid and a large $rmsd$ value (>0.5 Hz) may be caused by flexibility. The flexibility may be added to the analysis by fitting a population distribution. This is true in method B where a population density function of the motion around a specific torsion angle, $\theta$, is calculated (Equation 7.2).

$$P(\theta) = \frac{1}{Z} \times e^{\frac{1-cos(\theta-\theta_0)}{\sigma^2}}$$ (7.2)

$\theta_0$ is the mean torsion angle, $\sigma^2$ is the variance (the square of the standard deviation) and $Z$ is the normalizing factor.

Theoretical coupling constants for each torsion angle were computed from Karplus-type equations. After multiplication by the population density function, the sum of the weighted $J$ values was compared with experimental data. Conclusively, in both method A and method B two $rmsd$ minima were found for each torsion angle.

The population probability distribution of the hydroxy methyl rotamer torsion angles, $\omega$, was given by solving a linear equation system:

$$<^{3}J_{15,165}> = P(gt) \times ^{3}J_{15,165}(gt) + P(gg) \times ^{3}J_{15,165}(gg) + P(tg) \times ^{3}J_{15,165}(tg)$$ (7.3)

$$<^{3}J_{15,165}> = P(gt) \times ^{3}J_{15,165}(gt) + P(gg) \times ^{3}J_{15,165}(gg) + P(tg) \times ^{3}J_{15,165}(tg)$$ (7.4)

$$1 = P(gt) + P(gg) + P(tg)$$ (7.5)
The theoretical $^3J_{H5,H6}$ values used in the calculation are computed from fixed values of $\omega$, namely the three staggered rotamers ($\omega_{(gt)} = +65$, $\omega_{(gg)} = -65$ and $\omega_{(tg)} = 180$), and a Karplus equation.

### 7.4 Results and discussion

Different hetero- and homonuclear coupling constants, which are relevant for the torsion angles in the target $\beta$-biosides, were collected from 1D $^1$H and 1D $^{13}$C NMR spectra and related to their corresponding Karplus-type equations. From the Karplus relationship the torsion angles $\phi_H$ and $\psi_H$ were estimated using method A (see section 7.3) and are shown in Table 7.1.

#### Table 7.1 Method A. Conformational analysis results for $\phi_H$ and $\psi_H$ of the four glucobiosides 1-4. Values of the torsion angles are given in degrees and $rmsd$ in Hz.

<table>
<thead>
<tr>
<th>$\phi_H$</th>
<th>syn</th>
<th>rmsd</th>
<th>anti</th>
<th>rmsd</th>
<th>$\psi_H$</th>
<th>syn</th>
<th>rmsd</th>
<th>anti</th>
<th>rmsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>g2g (1)</td>
<td>+41</td>
<td>0.2</td>
<td>-127</td>
<td>0.6</td>
<td>g2g (1)</td>
<td>-36</td>
<td>0.3</td>
<td>+133</td>
<td>0.5</td>
</tr>
<tr>
<td>g3g (2)</td>
<td>+40</td>
<td>0.3</td>
<td>-129</td>
<td>0.7</td>
<td>g3g (2)</td>
<td>-24</td>
<td>0.8</td>
<td>+140</td>
<td>1.2</td>
</tr>
<tr>
<td>g4g (3)</td>
<td>+39</td>
<td>0.3</td>
<td>-130</td>
<td>0.6</td>
<td>g4g (3)</td>
<td>-27</td>
<td>0.3</td>
<td>+140</td>
<td>0.6</td>
</tr>
<tr>
<td>g6g (4)</td>
<td>+38</td>
<td>0.1</td>
<td>-132</td>
<td>0.4</td>
<td>g6g (4)*</td>
<td>-3</td>
<td>0.5</td>
<td>+173</td>
<td>1.1</td>
</tr>
</tbody>
</table>

* The psi torsion angle is defined as $\psi_C$ (C5-C6-O6-C1').

The $rmsd$ value reported for each torsion angle is an indication of the probability that a torsion angle adopt this value. A lower $rmsd$ value gives a higher probability and vice versa. The flexibility at the $\phi_H$ torsion angle can be considered to be rather restricted for all four biosides because of the low $rmsd$ values. Methyl $\alpha$-gentiobioside, 4, which is a 6-linked disaccharide and therefore should be the compound with most flexibility around the glycosidic linkage, has the lowest $rmsd$ values for $\phi_H$, which was surprising. On the other hand, comparison of our results with X-ray data for $\beta$-gentiobioside and amygdalin (see Section 7.4.4) indicates that the analysis was successful. In general, there is a higher degree of flexibility at the $\psi_H$ torsion angle than around $\phi_H$. To take the flexibility into account, both method A (Table 7.1) and method B (Table 7.2) are used in the analysis of $\psi$ (see section 7.4.4).

#### Table 7.2 Method B. Conformational analysis results for $\psi_H$ ($\psi_C$ for 4) of the four glucobiosides 1-4. Values of the torsion angles are given in degrees and $rmsd$ in Hz

<table>
<thead>
<tr>
<th>$\psi$ (syn)</th>
<th>$\sigma$</th>
<th>$rmsd$</th>
<th>$\psi$(anti)</th>
<th>$\sigma$</th>
<th>$rmsd$</th>
</tr>
</thead>
<tbody>
<tr>
<td>g2g (1)</td>
<td>-35</td>
<td>11</td>
<td>0.3</td>
<td>+133</td>
<td>15</td>
</tr>
<tr>
<td>g3g (2)</td>
<td>-9</td>
<td>29</td>
<td>0.5</td>
<td>+159</td>
<td>44</td>
</tr>
<tr>
<td>g4g (3)</td>
<td>-26</td>
<td>11</td>
<td>0.2</td>
<td>+142</td>
<td>21</td>
</tr>
<tr>
<td>g6g (4)</td>
<td>+4</td>
<td>21</td>
<td>0.2</td>
<td>-170</td>
<td>33</td>
</tr>
</tbody>
</table>
Results for the \( \omega \) torsion angle are shown in Table 7.3 and are presented as probability distributions, i.e. as \( \psi_t \), \( \psi_g \) and \( \psi_{tg} \) staggered rotamers.

**Table 7.3** Hydroxymethyl group conformations of compound 1-4. Populations (P) are given in % and \( J \)-values in Hz.

<table>
<thead>
<tr>
<th>( g^2g ) (1)</th>
<th>( \alpha )-Glc ( \rho )</th>
<th>5.5</th>
<th>2.1</th>
<th>-12.2</th>
<th>49</th>
<th>44</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )-Glc ( \rho )</td>
<td>5.7</td>
<td>2.4</td>
<td>-12.3</td>
<td>50</td>
<td>40</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>( g^3g ) (2)</td>
<td>( \alpha )-Glc ( \rho )</td>
<td>5.3</td>
<td>2.2</td>
<td>-12.4</td>
<td>46</td>
<td>46</td>
<td>9</td>
</tr>
<tr>
<td>( \beta )-Glc ( \rho )</td>
<td>6.0</td>
<td>2.2</td>
<td>-12.4</td>
<td>54</td>
<td>38</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>( g^4g ) (3)</td>
<td>( \alpha )-Glc ( \rho )</td>
<td>4.8</td>
<td>2.0</td>
<td>-12.3</td>
<td>41</td>
<td>52</td>
<td>6</td>
</tr>
<tr>
<td>( \beta )-Glc ( \rho )</td>
<td>5.9</td>
<td>2.0</td>
<td>-12.4</td>
<td>54</td>
<td>40</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>( g^6g ) (4)</td>
<td>( \alpha )-Glc ( \rho )</td>
<td>4.8</td>
<td>2.0</td>
<td>-11.6</td>
<td>41</td>
<td>52</td>
<td>6</td>
</tr>
<tr>
<td>( \beta )-Glc ( \rho )</td>
<td>6.1</td>
<td>2.2</td>
<td>-12.3</td>
<td>55</td>
<td>37</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

The distributions are computed using Equations 7.3, 7.4 and 7.5. From Table 7.3 one can draw the conclusion that the \( \omega \) population distribution does not change when the 6-positon is glycosylated (4).

### 7.4.1 \( \beta \)-D-Glc\( \rho \)-(1\( \rightarrow \)2)-\( \alpha \)-D-Glc\( \rho \)OMe, Methyl \( \alpha \)-sophoroside (1)

The conformational analysis of 1 gave a single conformer where \( \phi_H \) is +41° and \( \psi_H \) is -36° (Table 7.1), which is consistent with a syn conformation around the glycosidic linkage. Flexibility is of course an issue and the anti conformation is possible, but the higher rmsd value of the anti conformation indicates that the proposed syn conformation is the preferred one. This is also in line with results from X-ray diffraction data where \( \phi_H = +41° \) and \( \psi_H = -21° \).\(^{58}\) Comparison of 1 with \( \beta \)-D-Gal\( \rho \)-(1\( \rightarrow \)3)-\( \beta \)-DGal\( \rho \)OMe (19) (Figure 7.2) shows that the arrangement around the glycosidic linkage is the same and therefore should adopt similar conformations. Results calculated by Stenutz et al. shows that this is indeed the case.\(^{59}\) They measured \( J_{C,C} \) in \( ^{13}C \) isotopically labelled 19 and from this, the torsion angles were found to be -31° (\( \psi_H \)) and around +60° (\( \phi_H \)), which supports the syn conformation, shown in Figure 7.2.

![Figure 7.2](Image)  
**Figure 7.2** Comparison of the syn arrangements around the glycosidic linkage of disaccharides 1 and 19.
7.4.2 β-D-Glcp-(1→3)-α-D-GlcpOMe, Methyl α-laminaribioside (2)

A syn arrangement, $\phi_H = +40^\circ$ and $\psi_H = -24^\circ$, around the glycosidic linkage of disaccharide 2 was found when considering the calculated rmsd values in Table 7.1 (Figure 7.3). As expected, the flexibility at the $\psi_H$ torsion angle can be considered to be larger than the flexibility at the $\phi_H$ torsion, which is established by the high rmsd of the $\psi_H$ syn conformation and the low difference between the rmsd of the $\psi_H$ syn and anti conformers (Table 7.2). Measurements in the solid state by X-ray diffraction shows that the torsion angle $\phi_H$ adopts values in the range of +28° to +43° and $\psi_H$ of -38° to -52°, which is consistent with our results.

In the syn conformation, O4 is pointing towards the ring oxygen O5’ and hydrogen bonding OH-4 to O5’ is possible (Figure 7.3). This is thought to give a stabilizing effect and to favour the proposed syn conformation. A hydrogen bond has been reported for methyl β-D-lactoside (20, β-D-Galp-(1→4)-β-D-GlcpOMe) where the OH-3 hydroxy group is pointing towards O5’, in the same manner as OH-4 and O5’ in 2.62, 63

![Figure 7.3](image)

**Figure 7.3** The syn arrangement around the glycosidic linkage in structures 2 and 20. Possible hydrogen bonds are drawn as dotted lines.

7.4.3 β-D-Glcp-(1→4)-α-D-GlcpOMe, Methyl α-cellobioside (3)

Similarly to the results for methyl α-laminarabioside (2, Figure 7.3), the analysis results for 3 are consistent with a syn arrangement around the glycosidic linkage (Figure 7.4). Hydrogen bonds are possible when the equatorial hydroxyl group, OH-3, of the reducing residue is pointing towards the ring oxygen O5’ in the non reducing residue. The results of the analysis of 3 indicate a relatively rigid structure with a syn conformation where $\phi_H = +39^\circ$ and $\psi_H = -27^\circ$ (Table 7.1) and the proposed hydrogen bond can be expected to have a stabilizing effect.

Our results for 3 are consistent with data reported for lactose and cellobiose. Solid state data for these compounds show that $\phi_H = +24^\circ - +52^\circ$ and $\psi_H = -48^\circ - -7^\circ$. The results are further confirmed by NOE data reported for lactose in water where $\phi_H = +50^\circ$ and $\psi_H = +0^\circ$, as well as from...
measurements of dipolar couplings where $\phi_H = +49^\circ$ and $\psi_H = +13^\circ$. Measurements of $^3J_{C,H}$ and $^3J_{C,C}$ gave $\phi_H = +40^\circ$ and $\psi_H = -15^\circ$. All these data support the results we observed for 3.

Larsson et al. observed a small population of the anti conformer of 3 ($\psi_H = +140^\circ$) from MD data. Flexibility is, of course, an issue, but due to the margin of error in the Karplus equation this could not be established in our analysis. Further NMR experiments may be performed to resolve this question.

![Figure 7.4](image)

**Figure 7.4** The syn conformation of 3 with a proposed hydrogen bond between O5’ and O3 marked with a dotted line.

### 7.4.4 β-D-Glc-(1→6)-α-D-GlcpOMe, Methyl α-gentiobioside (4)

Methyl α-gentiobioside (4) differs from the other biosides in this study. The linkage to a primary position gives a more flexible glycosidic linkage and an additional torsion angle, $\omega$, to consider in the conformational analysis. The results from our analysis indicate a syn conformation around the glycosidic linkage where $\phi_H = +38^\circ$ and $\psi_C = -3^\circ$, but taking the flexibility into account, the anti conformation ($\psi_C = -170^\circ$) is the most possible arrangement (Tables 7.1 and 7.2).

Analysis of $\omega$ shows almost equal populations for the $gt$ and $gg$ rotamers and a minor population of the $tg$ rotamer (Table 7.3). This is in agreement with the distribution found for all hydroxymethyl rotations of all biosides in this investigation. In principle, it shows that the substitution by a carbohydrate residue in the 6-position does not influence the $\omega$ torsion angle to a large extent. In contrast, the results concerning β-D-GlcNAc-(1→6)-α-D-ManpOMe (21) contradict the results for 4.

![Figure 7.5](image)

**Figure 7.5** Structure 4 in its anti arrangement around the glycosidic linkage.
β-gentiobiose and amygdalin ((R)-1-cyano-1-(phenylmethyl)-β-D-glucopyranosyl-(1→6)-β-D-glucopyranoside) are structures similar to 4. The published solid state data of β-gentiobiose are $\phi_H = 63^\circ$, $\psi_C = -156^\circ$ and $\omega = -54^\circ$. This differs from our results as well as from the solid state data of amygdalin, where $\phi_H = 26^\circ$, $\psi_C = 155^\circ$ and $\omega = -74^\circ$. By taking the average of the X-ray data for β-gentiobiose and amygdalin ($\phi_H = 44^\circ$, $\psi_C = \pm 180^\circ$ and $\omega = -64^\circ$) a better agreement with the anti conformation proposed by us is found.

7.5 Conclusions

The approach of using coupling constants, obtained from specifically $^{13}$C isotopically labelled carbohydrates, in combination with the analytical methods described above, has been proved to be successful. Generally, both methods A and B give acceptable solutions for each torsion angle. The results discussed above show that for a rigid bond, having a preference for only a single conformer, method A is the preferred approach. This is often true for $\phi$, which is related to less flexibility than $\psi$ because of the stabilizing exo-anomeric effect. On the other hand, the flexibility around $\psi$ can not be excluded in the analysis and method B may be used.
Conformational flexibility and dynamics of a (1→6)-linked disaccharide related to an oligosaccharide epitope expressed on malignant tumour cells (paper IV)

8.1 Introduction

The trisaccharide \(N\)-Acetyl-2-amino-2-deoxy-\(\beta\)-D-glucopyranose (1→6)-\(\alpha\)-D-mannopyranose-(1→6)-\(\beta\)-D-mannopyranose (\(\beta\)-D-GlcNAc(1→6)-\(\alpha\)-D-Manp-(1→6)-\(\beta\)-D-Manp) is part of glycoproteins occurring on human cell surfaces (Figure 8.1). It is known that an increased \(\beta\)(1→6)-branching of terminal carbohydrates of cell surface glycoproteins is associated with cancer. Unfortunately, little is known about how the alteration influences molecular recognition processes, interactions that are often very specific and dependent on both the primary and secondary structure of the carbohydrate. Consequently, to study these mechanisms more closely, conformational flexibility and dynamics of the carbohydrates have to be taken into account. The highest degree of freedom in carbohydrates is found around the glycosidic linkage and generally, 6-linked carbohydrates are more flexible than others.

The aim of this project was to analyse the conformations around the two glycosidic linkages in the trisaccharide \(\beta\)-D-GlcNAc-(1→6)-\(\alpha\)-D-Manp-(1→6)-\(\beta\)-D-Manp, highlighted in Figure 8.1 (above). To simplify the task, the project is divided into two, where disaccharides \(\beta\)-D-GlcNAc-(1→6)-\(\alpha\)-
D-ManpOMe (21) and α-D-Manp-(1→6)-α-D-ManpOMe (22) each represent one glycosidic linkage of the parent trisaccharide. The analysis is performed using NMR spectroscopy and MD simulations. To enhance the sensitivity of the NMR analysis and to reduce the problem with overlap in the proton spectrum, specifically 13C labelled disaccharides have been used in the analysis.

8.2 Synthesis

Both the unlabelled and the [6-13C]-labelled derivative of disaccharides 21 and 22 were synthesised. Compound 21 was made from thioglycoside donor 23 together and acceptor 24 in a stereoselective glycosylation reaction (Scheme 8.1). The first step in the preparation of acceptor 24, the methyl α-D-mannoside (25) was made by Fischer glycosylation from D-mannose and [6-13C]-D-mannose, respectively with methanol as the nucleophile. Selective tritylation of the primary hydroxyl group in 25, benzylolation and finally, detritylation under acidic conditions, gave acceptor 24.

The glycosylation reaction gave the fully protected disaccharide 26 exclusively as the β-product in excellent yield. The amino group of the donor was protected as N-tetrachlorophthalimide (NTCP) and AgOTf/NIS promoted the reaction. The NTCP group was then removed with ethylenediamine, yielding the free amine which then was acetylated before the removal of benzyl- and acetyl groups to achieve the desired disaccharide 21. Before analysis by NMR spectroscopy, the disaccharide was purified on a P2 size exclusion column, deionised with CHELEX, lyophilized from D2O and degassed.
The glycosylation reaction to form the precursor of disaccharide 22 was performed, using AgOTf as promoter, at –40 °C (Scheme 8.2).\textsuperscript{84,85,86} The glycosyl bromide 27\textsuperscript{87} was coupled to the free primary hydroxyl group of acceptor 24 yielding the protected disaccharide 28 solely with the α-configuration. From 28 the target product 22 was obtained after debenzylation and debenzoxylation, in a total yield of 58% (from 27). Disaccharide 22 was purified and lyophilized with D\textsubscript{2}O before analysis.

Scheme 8.1 a. NIS, AgOTf, CH\textsubscript{2}Cl\textsubscript{2}; b. 1. Ethylenediamine, MeCN:EtOH:THF 2:1:1, 60 °C 2. Ac\textsubscript{2}O, Pyridine; c. 1. NaOMe, MeOH/CH\textsubscript{2}Cl\textsubscript{2}, 2. H\textsubscript{2}, Pd(OH)\textsubscript{2}/C.

Scheme 8.2 a. AgOTf, collidine, CH\textsubscript{2}Cl\textsubscript{2}, –40 °C; b. 1. NaOMe, MeOH/CH\textsubscript{2}Cl\textsubscript{2}, 2. H\textsubscript{2}, Pd(OH)\textsubscript{2}/C.
8.3 Results and discussion

Results from the conformational analysis of the glycosidic linkages in the (1→6)-linked disaccharides 21 and 22 are described below. As a reference, a conformational distribution analysis of the hydroxymethyl group in α-D-ManpOMe (25) was performed and the result is compared to the results of 21 and 22.

![Figure 8.2 The two analysed 6-linked disaccharides 21 and 22. Monosaccharide 25 is used as a reference in the conformational analysis of the hydroxymethyl group.](image)

The main focus of the project has been disaccharide 21 where conformations around the glycosidic linkage were studied in greater detail than for disaccharide 22. In both cases, the ω torsion angles were studied and in 21 also φ and ψ torsion angles.

8.3.1 α-D-ManpOMe (25)

As a reference, the hydroxymethyl population distribution of 25 was analysed and is reported as 52% gt, 38% gg and 9% tg, calculated from $^1J_{H5,H6}^{pro-R}$=5.9 Hz and $^3J_{H5,H6}^{pro-S}$=2.3 Hz. From the MD simulation (40 ns) the population distribution is found to be 47:36:17. Furthermore, from the 1D $^1$C NMR spectrum $^2J_{C4,C6}$ was measured and found to be 0 Hz. Because of the low population of tg, this finding confirms the reported population equilibrium between gt and gg rotamers (see Figure 8.3).
Figure 8.3 Visualization of Equation 4.7. A value of $^{2}J_{\text{C4,C6}} = 0$ in compound 25 indicates a population equilibrium between $g_{r}$ and $g_{g}$ rotamers.

8.3.2 β-d-GlcNAc-(1→6)-α-d-ManpOMe (21)

A closer examination of 21 shows that the H$_{6}^\text{pro-R}$ signal overlaps with ring resonances in the $^1$H NMR spectrum. An advantage of using the $^{13}$C-6 enriched derivative of 21 is the split of the H6 signals by $^{1}J_{\text{C,H}}$ in the $^1$H NMR spectrum (Figure 8.4). Importantly, by using this approach it is possible to measure $^{3}J_{\text{H5,H6}^\text{pro-R}}$ and thereby reveal information considering the $\omega$ torsion angle. Similarly, by the appearance of the parent signal in the $^1$H NMR spectrum NOE effects could be observed.

Figure 8.4 Part of the $^1$H NMR spectrum of 21, showing the H6 protons split by the one bond coupling to C6, respectively.

The distance between protons around the glycosidic linkage gives valuable conformational information. The measurements of short proton-proton distances ($r_{ij}$) around the linkage was performed using the [6-$^{13}$C]-labelled compound in the NMR analysis. Selective excitation of H1’ in the 1D NOE and T-ROE experiments, using five different mixing times (50, 80, 110, 150 and 180 ms), revealed correlations to the two H6 protons. From 1D NOE (Figure 8.5) and T-ROE build-up curves, where the initial slope corresponds
to the cross-relaxation rate ($\sigma_{ij}$), $r_{ij}$ was calculated for H6$_{\text{pro-R}}$, H6$_{\text{pro-S}}$ and H3’ using ISPA (Equation 4.1). The distances are reported in Table 8.1.

![Figure 8.5](image)

**Figure 8.5** 1D NOE build-up curves showing the build-up rates ($\sigma$) of H1’ to ($\bullet$) H6$_{\text{pro-R}}$, (■) H3’ and (♦) H6$_{\text{pro-S}}$.

Calculated effective correlation times ($\tau_{\text{eff}}$) are around 120 ps for all three protons. This means that the intramolecular H1’-H3’ distance, obtained from the MD simulation, may be used as a reference distance in the analysis.

A 25 ns long MD simulation was performed and the results were compared to NMR data (Table 8.1). It shows that the $r_{ij}$ values from the two techniques are closely related and conclusively, the force field gives a realistic picture of the reality and that MD data obtained may further on be used to describe the conformations of 21.

**Table 8.1** Calculated cross-relaxation rates ($\sigma$) and distances obtained upon excitation of H1’ in the [6-$^{13}$C]-labelled derivative of compound 21, using 1D NOE and T-ROE experiments.

<table>
<thead>
<tr>
<th>Proton</th>
<th>$\sigma_{\text{NOE}}$</th>
<th>$\sigma_{\text{T-ROE}}$</th>
<th>$r_{ij}$ (NMR)</th>
<th>$r_{ij}$ (MD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H6$_{\text{pro-R}}$</td>
<td>6.67</td>
<td>10.1</td>
<td>2.33</td>
<td>2.39</td>
</tr>
<tr>
<td>H3’</td>
<td>4.01</td>
<td>6.06</td>
<td>2.54[a]</td>
<td>2.54</td>
</tr>
<tr>
<td>H6$_{\text{pro-S}}$</td>
<td>3.08</td>
<td>4.51</td>
<td>2.67</td>
<td>2.74</td>
</tr>
</tbody>
</table>

[a] Reference distance from the MD simulation of 25 ns.

Furthermore, the resulting time trajectory from the MD simulation shows that the value of $\phi_{H1}$ is around 50° and the more flexible $\psi_C$ adopts an average value close to 180° (Figure 8.6). Measured $^3J_{H1',C6} = 4.2$ Hz and $^3J_{C6,C2'} = 2.9$ Hz corresponds to a $\phi_{H1}$ value in agreement with the proposed value from the MD experiment (see Figure 4.7).
The most flexible torsion angle, \( \omega \), shows two main conformations, corresponding to \( gt \) and \( gg \) rotamers. From the MD simulation the percentage distribution of the three rotamers \( gt \), \( gg \) and \( tg \) are 67, 28 and 5 %, respectively. From the \(^1\)H NMR spectrum, \( ^3J_{\text{H5,H6pro-R}} = 6.4 \) Hz and \( ^3J_{\text{H5,H6pro-S}} = 1.9 \) were extracted, giving the corresponding population distribution of 60:35:5. This result is equal to previous results reported by Lycknert et al.\(^90\) and in agreement with the results for compound 25. As for 25, the finding of \( ^2J_{\text{C4,C6}}=0 \) indicates that there is a population equilibrium between \( gt \) and \( gg \) rotamers, which further confirms the suggested results of the hydroxymethyl population distribution.

To sum up, the value of \( \psi \) corresponds to an antiperiplanar arrangement of \( C1' \) and \( C5 \). The NOE data showed that there is close contact between the \( C1' \) and the \( H6 \) protons which indicates that there is a syn-arrangement around the glycosidic linkage, as discussed for 4 in section 7.4.4.

### 8.3.3 \( \alpha-\text{D-Manp-(1} \rightarrow 6)\)-\( \alpha-\text{D-ManpOMe (22)} \)

The conformational analysis of 22 is based on NMR data and crystallographic data from the Protein Data Bank (PDB).\(^91\) Conventionally, the hydroxymethyl population distribution of 22 was calculated from \( ^3J_{\text{H5,H6}} \) values, which were found to be rather small: \( ^3J_{\text{H5,H6pro-R}} = 3 \) Hz and \( ^3J_{\text{H5,H6pro-S}} < 2 \) Hz. The measured coupling constants correlate with a population distribution of 21% \( gt \), 72% \( gg \) and 7% \( tg \). In the same manner as discussed in section 8.3.1, the finding of \( ^2J_{\text{C4,C6}} = 0 \) Hz in 22 verifies the proposed conformations of \( \omega \). Spronk et al. reported values of \( ^3J_{\text{H5,H6pro-R}} = 2.2 \) Hz and

---

**Figure 8.6** The time dependence of glycosidic torsion angles as a result of a 25 ns long MD simulation of 21.
$^3J_{H5;H6pro-S} = 1.9$ Hz, which corresponds to 96% gg.\(^{92}\) This is not in agreement with our results and to answer the question more experiments have to be performed. A search for the structural element $\alpha$-D-Manp-(1→6)-$\alpha$-D-Manp in the PDB gave 49 hits (chain type=N-glycan, resolution=2.5Å and experimental method=X-ray). Examination of the $\omega$ conformations of the structures from PDB gave 23% gt, 71% gg and 6% tg, which perfectly matches our results from NMR data. Unfortunately, about 30% of the recorded anomeric configurations in the database may be incorrect.\(^{93}\) So, in context, the anomeric preference has to be excluded from the analysis. A new search for $\alpha$-D-Manp-(1→6)-D-Manp resulted in 302 hits and the distribution became slightly different: 40% gt, 55% gg and 5% tg, but still matches our results.
9 Concluding remarks

The aim for new and cheap vaccines and treatments against bacterial and viral infections is always current. This is especially true for developing countries, where an infection outbreak may lead to serious consequences for large populations. This thesis describes attempts to contribute to a better understanding of carbohydrate primary and secondary structures. Hopefully, it will increase the understanding of the biological systems that carbohydrates are involved in.

In the first two projects discussed, the structures of two O-antigenic PS from *E. coli* bacteria were successfully determined, using NMR spectroscopy. Their similarities to other O-antigens are also discussed. The last two chapters deal with the 3D structure of different disaccharides and the advantages of using specific $^{13}$C-labelled saccharides in conformational analysis. For this purpose, two statistical methods have been developed and are presented in this work. It is important to emphasize that a combination of different techniques, methods and experiments enhances the accuracy of the analysis.

Finally, knowledge of the structure, conformations and dynamics of poly- or oligosaccharides may facilitate future research, for example, studies of carbohydrate-protein interactions.
Acknowlegements

There is a bunch of people that supported me in different ways through the years, whom I couldn’t have done this work without!

*I would especially like to thank…*

…my excellent supervisor Professor Göran Widmalm for being a fantastic mentor! Thanks for all your patience and time and for sharing your knowledge with me! I’m very pleased that you accepted me as one of your students when Roland decided to go elsewhere..!

…Docent Roland Stenutz, without whom this thesis wouldn’t have been written! Thanks for accepting me as a PhD student and for the first three years as a graduate student!

…Professor emeritus Per J. Garegg and Professor Jan-E. Bäckvall for their kind interest in this work.

…Professor Pino Pilotti for graduate courses in Advanced Organic Chemistry and Physical Organic Chemistry.

…Docent Ian Cumstey for linguistic improvements.

…Elin Säwén, Hanna Jonsson, Jens Landström, Magnus Lundborg and Mona Svensson for comments on this thesis.

…the past and present members of the GW-group that have contributed to the fantastic atmosphere in our lab and office: Andy, Cecilia, Classe, Elin, Hannaj, Jennie, Jens, Lucas, Mange F., Magnus L., Mona, Matthias, Robban, Smulan, Tara and Tina. Work is work, but I really believe that you are a lot more motivated if you have fun at work – and I have been motivated from the beginning to the bitter end and all this thanks to you!!
…Pia Seffers and Kristina Romare for all the help with instruments and stuff. Well, the truth is: no doctoral thesis without perfectly working analytical equipments!

…Britt Eriksson and Olle Bristrand for everything! I really hope you know how important you are!

…Torbjörn Astlind at the Department of Biochemistry and Biophysics for the technical support with the 600 MHz NMR spectrometer!

…Jenny Adrian-Meredith och Linnéa Borén: vi bestämde oss för att ”go for it” och det har inte bara varit en dans på rosor, det har varit kämpigt också! Men mindre kämpigt än förväntat tack vare alla fikningar, luncher, pratstunder (”tissel och tassel”), tårar och skratt!

…the rest of the past and present colleges and friends at the Department of Organic Chemistry. Thanks for discussions, lunches, coffee brakes and parties. My time at the department wouldn’t have been the same without nice company and crazy persons!

…L&E Kinanders stipendiestiftelse, C.F. Liljevalchs J:ors stipendiestiftelse, Knut och Alice Wallenbergs stiftelse, AstraZeneca och Hilda Rietz’s stiftelse för de generösa stipendier jag fått till resor och dator. Jag vill också tacka Ernst och Karl Tryggers utvecklingsfond för de generösa bidrag jag fått under hela min grundutbildning.

Jag vill också tacka alla er som indirekt bidragit till att jag klarat av mitt arbete:

Jag är världens rikaste människa som har så många nära vänner, sådana som man kan lita på i vått och torrt och som ger mig allt självförtroende man kan behöva för att ta sig igenom en forskarutbildning. Jag tackar er ALLA, ingen är nämnd och ingen glömd, för att ni alltid talar om för mig att jag är så jäkla bra! Ni är också helt fantastiska och utan er support så hade jag lagt ner för länge sedan…

Tack Hjalmar för gott samarbete kring Tova! Och Monika för allt stöd och stora engagemang!
Stort tack till mina föräldrar, mamma Barbro och pappa Roger som alltid ställer upp när det kniper på ett eller annat vis. Tack också för att ni orkar med er, ibland något, lynniga dotter..! Och till lillebror Kent, mina småsyster Louice och Veronica för att ni finns till! Tack också mormor Aline och farmor Stina som säger att jag är så duktig!

Tack Tova, mitt hjärta, den underbaraste lilla varelsen i hela universum – du är helt fantastisk och livet är så underbart med dig! Du är den klokaste och snällaste människan jag känner och du gör mig alltid så glad, det är du som är min drivkraft! Mamma älskar dig obegränsat 🌹!
References

1 Sharon, N.; Lis, H. Scientific American, 1993, 74-81.
4 Varki, A. Glycobiology 1993, 3, 97-130.


Stenutz, R.; Vercillo, M.; Kubish, M.; Serianni A. *Unpublished results*.


