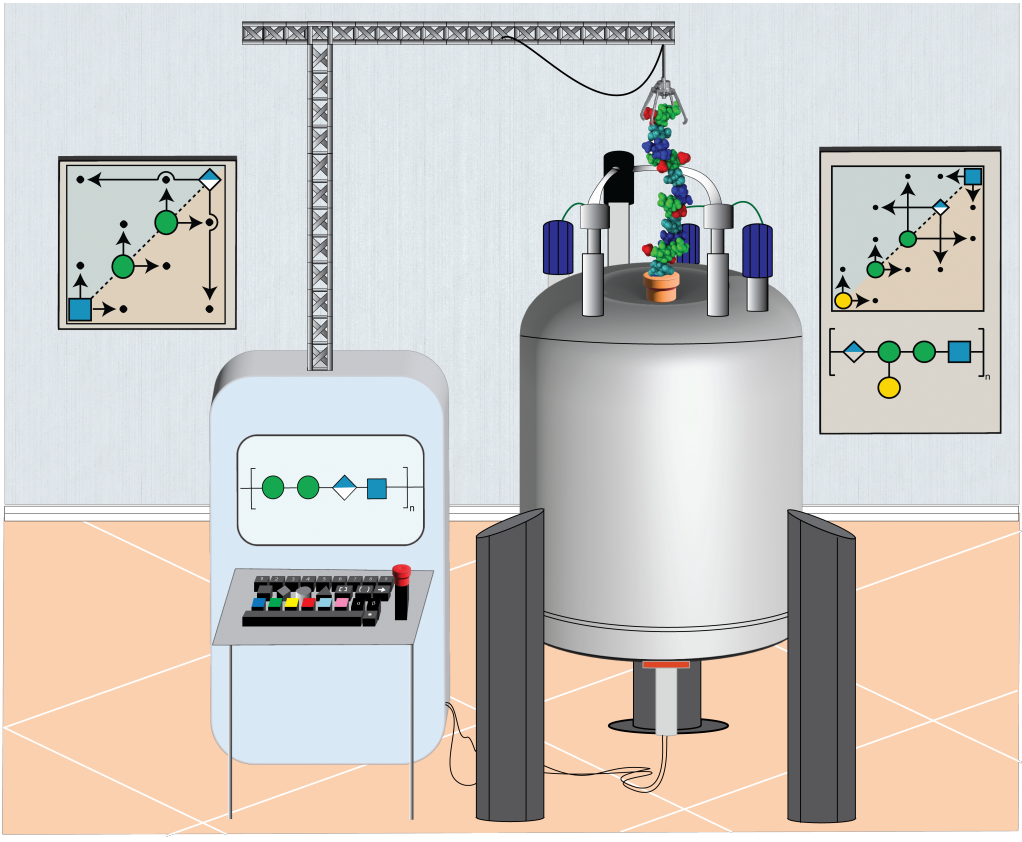
Stockholm University

Structural and Conformational Analysis of Bacterial Polysaccharides using NMR Spectroscopy

Axel Furevi



Doctoral Thesis in Organic Chemistry at Stockholm University, Sweden 2022

Structural and Conformational Analysis of Bacterial Polysaccharides using NMR Spectroscopy

Axel Furevi

Academic dissertation for the Degree of Doctor of Philosophy in Organic Chemistry at Stockholm University to be publicly defended on Friday 22 April 2022 at 10.00 in Magnélisalen, Kemiska övningslaboratoriet, Svante Arrhenius väg 16 B.

Abstract

Carbohydrates is one of the three classes of biomolecules found in nature. It is the most common one in comparison to the other two classes, lipids and proteins. However, this simple categorization does not reflect the reality since carbohydrates often are covalently linked to *e.g.*, proteins, so-called glycoproteins where, for example, N-glycans are used as markers of quality control during the process of protein folding. Another example is lipopolysaccharides, which cover the cell surfaces of gram-negative bacteria and which contain both a lipid moiety (Lipid A) and a carbohydrate chain. The outer part of the carbohydrate chain is a polysaccharide, also called O-antigen, as it interacts with the immune system of the host. The polysaccharide has, like a polymer, a repeating unit consisting of 2-7 monosaccharides. The repeating unit varies between different bacteria. Determining the structure of these polysaccharides is important in order to be able to categorize the various strains that exist, but also to be able to develop future glycoconjugate vaccines. This is important as the WHO estimates that antibiotic resistance is expected to be more lethal than cancer by 2050, and therefore a vaccine is needed to slow down this development.

Nuclear Magnetic Resonance Spectroscopy (NMR) is a useful analytical tool to analyze these carbohydrates at the atomic level in order to determine their structures.

The first part (Paper I-III) of this thesis will summarize the structural determination of three Escherichia coli serogroups with hitherto unknown lipopolysaccharides.

The second part (Paper IV) will discuss the structure determination, using NMR spectroscopy, for various mono-, di-, and tri-saccharides that have recently been implemented in the structure-determination program, CASPER. The chapter will also present examples of predictions of complex carbohydrates that CASPER can perform.

The third part (Paper V) of the thesis will investigate conformational aspects of the polysaccharides from Shigella flexneri serotypes 7a and 7b using NMR spectroscopy and molecular dynamics simulations.

Stockholm 2022
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-202541

ISBN 978-91-7911-810-5
ISBN 978-91-7911-811-2

Department of Organic Chemistry

Stockholm University, 106 91 Stockholm

STRUCTURAL AND CONFORMATIONAL ANALYSIS OF BACTERIAL POLYSACCHARIDES USING NMR SPECTROSCOPY

Axel Furevi

Stockholm University

Structural and Conformational Analysis of Bacterial Polysaccharides using NMR Spectroscopy

Axel Furevi

ISBN print 978-91-7911-810-5
ISBN PDF 978-91-7911-811-2

Printed in Sweden by Universitetsservice US-AB, Stockholm 2022

Till min familj.

Abstract

Carbohydrates are one of the four classes of biomolecules found in nature. It is the most common one in comparison to the other three classes, lipids nucleic acids and proteins. However, this simple categorization does not reflect the reality since carbohydrates often are covalently linked to *e.g.*, proteins, so-called glycoproteins where, for example, *N*-glycans are used as markers of quality control during the process of protein folding. Another example is lipopolysaccharides, which cover the cell surfaces of Gram-negative bacteria and contain both a lipid moiety (Lipid A) and a carbohydrate chain. The outer part of the carbohydrate chain is a polysaccharide, also called O-antigen, as it may trigger host immune response. The polysaccharide has, like a polymer, a repeating unit consisting of 2-7 monosaccharides and can vary significantly between different bacteria. Determining the structure of these polysaccharides are important since they can be utilized for the development of glycoconjugate vaccines. WHO estimates that antibiotic resistance is expected to be more lethal than cancer by 2050, and therefore vaccines could help in slowing down this development.

Nuclear Magnetic Resonance Spectroscopy (NMR) is a useful analytical tool to analyze these carbohydrates at the atomic level in order to determine their structures.

The first part (Paper I-III) of this thesis will summarize the structural determination of three *Escherichia coli* serogroups with hitherto unknown lipopolysaccharides.

The second part (Paper IV) will discuss the structure determination, using NMR spectroscopy, for various mono-, di-, and tri-saccharides that have recently been implemented in the structure-determination program, CASPER. The chapter will also present examples of predictions of complex carbohydrates that CASPER can perform.

The third part (Paper V) of the thesis will investigate conformational aspects of the polysaccharides from *Shigella flexneri* serotypes 7a and 7b using NMR spectroscopy and molecular dynamics simulations.

Populärvetenskaplig Sammanfattning

Kolhydrater är en av tre klasser av biomolekyler som finns i naturen och är också den mest frekvent förekommande i jämförelse med de två andra klasserna, lipider och protein. Denna simpla kategorisering representerar dock inte verkligheten. Istället så finns kolhydrater oftast kovalent sammanbundna med t.ex. proteiner, så kallade glykoprotein där exempelvis *N*-glykaner används som markörer när proteiner viks ihop till sina slutgiltliga strukturer. Ett annat exempel är lipopolysackarider som täcker gram-negativa bakteriers cellyta och som både innehåller en lipid-del (Lipid A) och en lång kolhydratskedja. Den yttre delen av kolhydratskedjan är en polysackarid som även kallas för O-antigen då den interagerar med värdens immunförsvar. Polysackariden har, likt en polymer, en repeterande enhet bestående av 2-7 monosackarider och varierar mellan olika bakterier. Att bestämma strukturen för dessa polysackarider är viktigt för att utveckla framtida glykokonjugatvaccin. Detta är viktigt då WHO beräknar att antibiotikaresistensen kommer vara dödligare än cancer år 2050 och därför behövs vaccin för att bromsa denna utveckling.

Kärnmagnetisk resonansspektroskopi (NMR) är ett användbart analysinstrument som på atomnivå kan analysera dessa kolhydrater för att kunna bestämma dess strukturer.

Den första delen (Paper I-III) av denna avhandling kommer att sammanfatta strukturbestämningen av tre *Escherichia coli* serogrupper med hittills okända lipopolysackarider och jämföra med liknande redan kända strukturer.

Den andra delen (Paper IV) kommer att diskutera strukturbestämningen, med hjälp av NMR, för olika mono-, di-, och tri-sackarider som sedan har implementerats i strukturbestämningsprogrammet, CASPER. Kapitlet presenterar också exempel på prediktioner av komplexa kolhydrater som CASPER kan utföra.

Den tredje delen (Paper V) av avhandlingen handlar om konformationsstudier av *Shigella flexneri* serotyperna 7a och 7b med hjälp av NMR experiment och molekylär dynamik simuleringar.

List of publications

Paper I **Structural analysis of the O-antigen polysaccharide from *Escherichia coli* O188.** Axel Furevi, Jonas Ståhle, Claudio Muheim, Spyridon Gkotzis, Klas I. Udekwu, Daniel O Daley, Göran Widmalm.
Carbohydrate Research, 498, 108051, 2020.

Paper II **Structure Elucidation of *Escherichia coli* O125ac O-antigen.** Axel Furevi, Klas I. Udekwu, Göran Widmalm. *Manuscript.*

Paper III **Structure Elucidation of *Escherichia coli* O93 O-antigen.** Axel Furevi, Jonas Ståhle, Claudio Muheim, Spyridon Gkotzis, Daniel O. Daley, Klas I. Udekwu, Göran Widmalm. *Manuscript.*

Paper IV **Complete ^{1}H and ^{13}C NMR chemical shift assignments of mono- to tetrasaccharides as basis for NMR chemical shift predictions of oligo- and polysaccharides using the computer program CASPER**. Axel Furevi, Alessandro Ruda, Thibault Angles d'Ortoli, Hani Mobarak, Jonas Ståhle, Christoffer Hamark, Carolina Fontana, Olof Engström, Patricia Apostolica, Göran Widmalm.
Carbohydrate Research, 513, 108528, 2022.

Paper V **Structure and Dynamics of O-antigens from *Shigella flexneri* serotypes 7a and 7b: An NMR Spectroscopy and MD Simulation Study.** Axel Furevi[†], Zaheer Timol[†], Neil Ravenscroft, Michelle M. Küttel, Göran Widmalm.
Manuscript.
[†]: These authors contributed equally

Previous documents based on this work

This thesis is partially based on the content from the author's half-time report "Structural studies of *Escherichia coli* O-antigens" (8th of June 2021, Stockholm)

The introduction has been revised and updated based on feedback obtained from the Half-time control and to better suit the expanded content of this thesis.

Chapter 2 (Paper I) was presented in the Half-time report and only has minor updates and corrections.

Chapter 3 (Paper II) was also presented in the Half-time report but has been re-written into a thesis summary chapter instead of, in the prior case, only as a manuscript.

Chapter 4 (Paper III) initial bioinformatical studies have previously been presented in the thesis "Structure Elucidations of Bacterial Polysaccharides using NMR Spectroscopy and Bioinformatics" by Dr. Jonas Ståhle in 2017.

Chapter 5 and 6 (Papers IV-V) are presented for the first time in this thesis.

Contribution list

Paper I	Planned and performed hot phenol/water extraction and subsequent purification steps to obtain the lipopolysaccharide and polysaccharide material from *E. coli* serogroup O188. Performed the NMR experiments, NMR assignments and GC-analysis. Contributed to the writing of related parts of the manuscript.
Paper II	Planned and performed hot phenol/water extraction and subsequent purification steps to obtain the lipopolysaccharide and polysaccharide materials from *E. coli* serogroup O125ab and O125ac. Planned and performed NMR experiments and assignments and wrote parts of the manuscript and supporting information.
Paper III	Planned and performed preparation and purification of polysaccharide materials. Planned and performed NMR experiments and assignments and wrote parts of the manuscript and supporting information.
Paper IV	Performed NMR experiments, resonance assignments, iterative fitting of spectra, implemented and validated new datasets in the database for multiple compounds. Performed the case studies and contributed to the writing of related parts of the manuscript.
Paper V	Planned and performed sample preparation, NMR experiments and subsequent analysis. Participated in the writing of the manuscript.

List of Abbreviations

BS-CT-HMBC	Band Selective Constant Time HMBC
CASPER	Computer-Assisted SPectrum Evaluation of Regular polysaccharides
dd	doublet of doublet
E. coli	*Escherichia coli*
EAEC	EnteroAggragative *E. coli*
ECA	Enterobacterial Common Antigen
EHEC	EnteroHemorrhagic *E. coli*
EIEC	EnteroInvasive *E. coli*
EPEC	EnteroPathogenic *E. coli*
ETEC	EnteroToxigenic *E. coli*
Fuc*p*	6-deoxy-galactopyranose
GC	Gas Chromatography
Gal*p*	Galactopyranose
Gal*p*NAc	2-acetamido-2-deoxy-galactopyranose
Glc*p*	Glucopyranose
Glc*p*A	Glucuronic acid
Glc*p*NAc	2-acetamido-2-deoxy-glucopyranose
GT	GlycosylTransferase
H2BC	Heteronuclear 2-Bond Correlation
HMBC	Heteronuclear Multiple-Bond Correlation
HSQC	Heteronuclear Single-Quantum Correlation
ISPA	Isolated Spin-Pair Approximation
Kdo	3-deoxy-D-manno-2-octulosonic acid
LPS	Lipopolysaccharide
Man*p*	Mannopyranose
meHSQC	Multiplicity Edited HSQC
NOESY	Nuclear Overhauser Effect Spectroscopy
NUS	Non-Uniform Sampling
OM	Outer Membrane
PANIC	Peak Amplitude Normalization for Improved Cross-relaxation
ppb	parts-per-billion
PS	PolySaccharide
Sf	*Shigella flexneri*
SNFG	Standard Nomenclature For Glycans
STEC	Shiga-Toxin producing *E. coli*
TSP	3-TrimethylSilyl-(2,2,3,3-$^{2}H_4$)-Propanoate
TOCSY	Total Correlation SPectroscopY

Table of Contents

Abstract i

Populärvetenskaplig Sammanfattning ii

List of publications iii

Previous documents based on this work iv

Contribution list v

List of Abbreviations vi

Table of Contents vii

1 Introduction 1
- 1.1 Monosaccharides 2
- 1.2 Polysaccharides 4
- 1.3 Conformational aspects of sugars 4
- 1.4 Sugar analysis 5
- 1.5 Absolute configuration 5
- 1.6 *E*scherichia coli 6
- 1.7 Lipopolysaccharides 7
- 1.8 Structural diversities in *E. coli* O-antigens 9
- 1.9 Nuclear Magnetic Resonance spectroscopy 10
- 1.10 Nuclear Overhauser Effect 10
- 1.11 Structure elucidation of carbohydrates by NMR spectroscopy 11
- 1.12 CASPER 13
- 1.13 Conformational analysis 14
- 1.14 From cell cultivation to NMR samples 14
- 1.15 Enterobacterial common antigen (ECA) 15
- 1.16 Aim of this Thesis 17

2 *E. coli* O188 O-antigen (Paper I) 18
- 2.1 Introduction 18
- 2.2 Results and discussion 18
  - 2.2.1 Sample preparation and initial NMR-analysis 18
  - 2.2.2 Component Analysis 18
  - 2.2.3 NMR Studies 19
  - 2.2.4 Structural comparisons with *Shigella boydii* type 16 21
- 2.3 Conclusion and future outlook 21

3 Structural studies of *E. coli* O125ac and O125ab O-antigen (Paper II) 22
- 3.1 Introduction 22
- 3.2 Results and discussion 22
  - 3.2.1 Sugar Analysis 22
  - 3.2.2 NMR analysis 23
- 3.3 Conclusion and future outlook 26

4 Structural studies of *E. coli* O93 O-antigen (Paper III) 27
- 4.1 Introduction 27
- 4.2 Results and discussion 27

4.2.1 Initial observations 27
4.2.2 Bioinformatics 27
4.2.3 NMR Studies 28
4.2.4 Structure elucidation by CASPER 30
4.2.5 Assignment of the *O*-acetyl groups 31
4.2.6 Conformational insights 34
4.2.7 Structural comparisons 35
4.3 Conclusion and future outlook 36

5 Developments of the NMR-database CASPER (Paper IV) 37
5.1 Introduction 37
5.2 Results and discussion 37
5.2.1 NMR assignments 37
5.2.2 Extraction of new correction sets 38
5.2.3 Case study: predicting the structure of α/β-Gd1a 38
5.2.4 Small $^{3}J_{HH}$ observed in a 3,4-disubstituted β-D-Glc*p*NAc 41
5.3 Conclusion and future outlook 42

6 Conformational Analysis of *Shigella flexneri* O-antigens (Paper V) 43
6.1 Introduction 43
6.2 Result and discussion 45
6.2.1 Temperature studies 45
6.2.2 $^{3}J_{HH}$-couplings of the GlcNAc residue 46
6.2.3 ^{1}H,^{1}H-NOESY buildup curves *Sf* 7a 47
6.2.4 MD-simulations and comparisons with NMR-data 48
6.3 Conclusion and future aspects 51

7 Concluding remarks 52

8 Acknowledgment 53

9 References 55

1 Introduction

All life-forms on earth consist of four classes of biomolecules: carbohydrates, lipids, nucleic acids, and proteins. In terms of mass, carbohydrates are the most abundant, and cellulose is the most prevalent organic polymer on earth.[1] Not only are they used as a building blocks *e.g.,* in cellulose (glucose), DNA (deoxyribose) and RNA (ribose), but they are also essential for energy storage, metabolism, and much more. Most often are carbohydrates present in nature combined with the other classes of biomolecules and are involved in life-essential processes. In this regard, an example is represented by *N*-glycans which are covalently attached to specific amino acids to form what is called glycoproteins. These *N*-glycans act as biomolecular markers that quality control the process of protein folding.[2] Furthermore, carbohydrates are also prevalent on every type of cell surface, thus playing a crucial role in cellular recognition.[3]

The ability of bacteria and virus to infect and cause disease in their host, the virulence factor, can be altered based on which carbohydrate structure is expressed on their surface. Virus, for instance, cover their spike proteins with oligomannoses stolen from the host cells to disguise themselves from the immune system of the host.[4–6] Gram-negative bacteria have lipopolysaccharides covering the outer surface creating an almost impenetrable barrier protecting them from the host's immune system and harsh environments *e.g.,* stomach acid.[7] Recent studies also showed that some Gram-negative *E. coli* bacteria can quickly modulate the chain length of their lipopolysaccharide after attachment to the host cell to further improve virulence.[8]

In order to understand these complex carbohydrate structures found in nature at an atomic level, several techniques can be utilized *e.g.*, mass spectrometry (MS)[9], molecular dynamics (MD)[10] simulations and nuclear magnetic resonance spectroscopy (NMR).[11]

This thesis will utilize NMR spectroscopy as the main technique in order to decipher polysaccharide structures related to Gram-negative bacteria.

1.1 Monosaccharides

Monosaccharides are polyhydroxylated carbon chains with typical lengths of between three and nine carbon atoms with a varying degree of hydroxyl groups attached to them. They are categorized either as aldoses or ketoses depending on, if they have an aldehyde or a ketone group attached, respectively.

In general terms, an 'aldohexose' is a monosaccharide with six carbon atoms bearing an aldehyde group. The monosaccharide chain can then further be of either 'L' or 'D' configuration. This is determined by the stereochemistry of the chiral center furthest away from the aldehyde or keto function in the chain (Figure 1). It is 'L' configured if the chiral center is on the left-hand side in a Fischer projection and 'D', if it is on the right-hand side. To also account for the stereochemistry of the remaining hydroxylated carbon atoms, common names are used (*e.g.*, D-glucose and D-mannose) for simplification. Sugars lacking one or more of its hydroxyl groups, deoxy sugars, have trivial names, like L-rhamnose (6-deoxy-L-mannose), to further help distinguish the variety of monosaccharides. More complex monosaccharides with longer chain lengths and additional substituents also have common names, i.e., Kdo (3-deoxy-D-manno-oct-2-ulosonic acid) and substituents with a certain absolute configuration are commonly specified with the 'D/L-system', i.e., the monosaccharide often abbreviated as, LD-Hep, is in fact, L-g*lycero*-D-*manno*-heptose (Figure 2).

This unique plethora of functional groups makes it possible for monosaccharide chains with five or more carbons to also exist in cyclic forms by hemiacetal (aldoses) or hemiketal (ketoses) formation, leading to either a 5-membered ring (furanose) or a 6-membered ring (pyranose).

The hemiacetal or hemiketal carbon within the newly formed ring is referred to as the anomeric carbon and can form two epimers or, more specifically, *anomers*. To distinguish between these two stereochemical outcomes, one compares the highest-ranked substituent of the *anomeric* position with the substituent of the *last* stereogenic center within the ring. If these two substituents are oriented *cis* to each other, the anomeric configuration is α, and if they are *trans* the monosaccharide is 'β' configured (Figure 1).

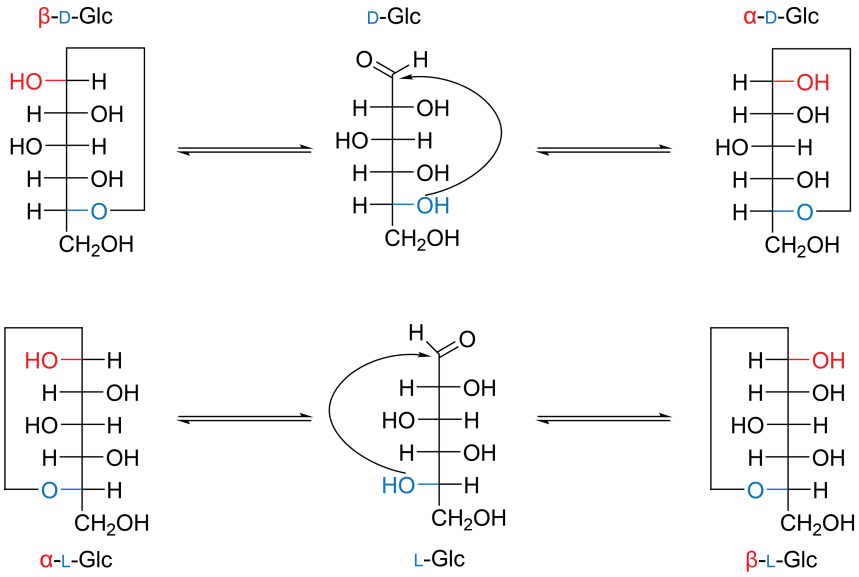


Figure 1. Fischer projections of D-glucose and L-glucose and their interconversion between linear and cyclic forms. The anomeric configuration (α/β) is highlighted in red, and the absolute configuration (D/L) in blue.

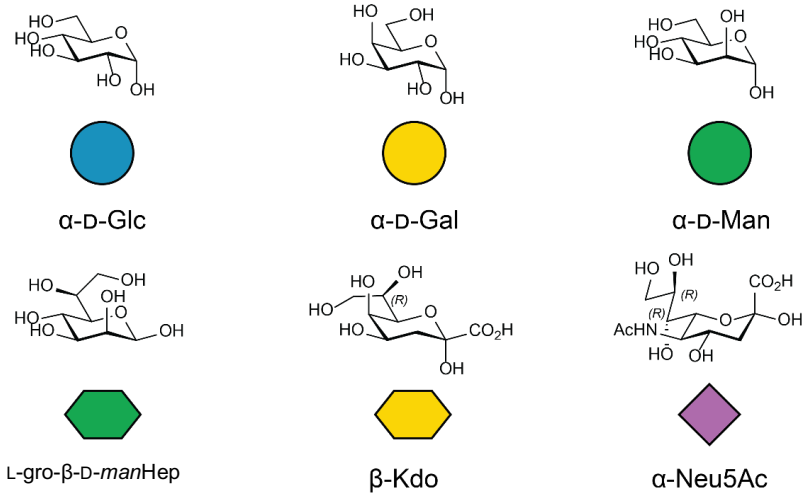


Figure 2. Examples of monosaccharides with various configurations and chain-lengths along with their associated SNFG representations.[12]

1.2 Polysaccharides

Polysaccharides form in nature with the help of glycosyltransferases (GT) by merging an activated sugar (donor) with the hydroxyl group of another sugar residue (acceptor) to first form di-, tri-, and oligomers in the most common biosynthetic pathways. These building blocks can then form longer chains, with a repeating motif, with the assistance of polymerases, to form a 'repeating unit'.

Simplified ways of representing these complex structures *e.g.*, symbol nomenclature for glycans (SNFG)[13] (Figure 2) displays each sugar with simplified shapes and other important information (linkage position, anomeric configuration, and degree of acetylation). Similar methods for 3D applications also have also been developed, *e.g.*, 3D-CFG[14] and 3D-SNFG.[15]

1.3 Conformational aspects of sugars

Sugars can adopt different conformational states. The different states are denoted with a letter describing the shape (*e.g.*, chair (*C*), boat (*B*), and skew (*S*)) along with superscript(s) and subscript(s) to explain where those atoms are located relative to a plane, *i.e.*, $^{1}C_{4}$ (C1 above the plane and C4 below the plane) (Figure 3). The reference plane for a $^{1}C_{4}$ is formed by remaining atoms; O5, C2, C3, and C5. For pyranoses, the typical states are the $^{4}C_{1}$ for D-configured pyranoses and $^{1}C_{4}$ chairs for L-configured sugars.

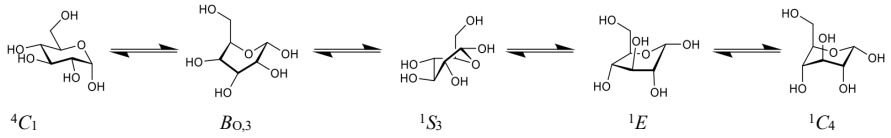


Figure 3. One of many possible conformation interconversion pathways for α-D-glucose pyranose between $^{4}C_{1}$ and $^{1}C_{4}$.[16]

For oligo- and polysaccharides, the two torsion angles in the glycosidic bond are represented as phi (ϕ) and psi (ψ). ϕ is the torsion angle consisting of H1'-C1'-O1-Cx and ψ is defined by positions C1'-O1'-Cx-Hx according to the NMR-definition. Due to steric and electronic effects, certain torsion angels are favoured (Figure 4).

The *exo*-anomeric effect involves the free electron pair of the glycosidic oxygen and stabilizes the antibonding orbital, σ*, of C1-O5 when ϕ is +60° for β-D-configured sugars and −60° for α-D-configured sugars. If the hydroxyl group of position C6 is substituted, the torsion angle ω is also used and is defined as O6-C6-C5-O5.

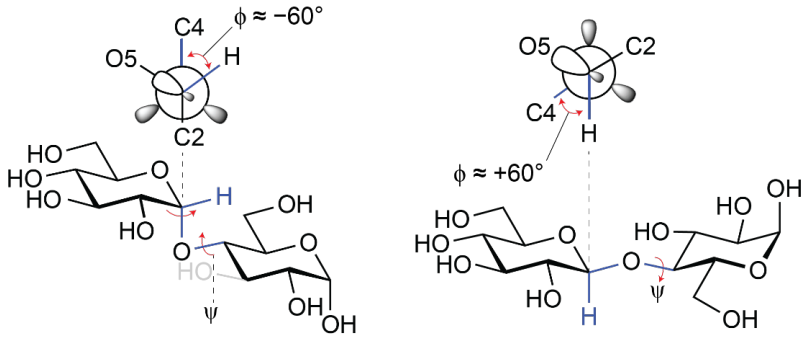


Figure 4. α-D-Glc-(1→4)-α-D-Glc and β-D-Glc-(1→4)-α-D-Glc disaccharides with their defined torsion angles, phi (ϕ) and psi (ψ) along with Newman projections related to torsion angle ϕ. The approximate values of torsion angle ϕ are given for both disaccharides when stabilized by the exo-anomeric effect.

1.4 Sugar analysis

In sugar analysis, a polysaccharide sample is first hydrolyzed under acidic conditions and is subsequently reduced under basic conditions with a reducing agent, sodium borohydride, to form the alditols and then per-*O*-acetylated using acetic anhydride and pyridine. This sample of per-*O*-acetylated alditols is then identified via gas-liquid chromatography (GLC) analysis by comparing it to reference samples.

1.5 Absolute configuration

The absolute configuration method is used to determine whether a monosaccharide is of L or D configuration. A polysaccharide material is first hydrolyzed and then glycosylated with an optically active alcohol compound (*e.g.* (+)-2-butanol). The formed mixture of butylfuranosides and butylpyranosides can then be analyzed by NMR or GLC.

The NMR approach utilizes the absolute configuration feature of CASPER[17] where the NMR chemicals shifts are compared to reference data. The GLC analysis[18] requires per-*O*-acetylated material which can then by the peak distribution patterns be compared to optically active monosaccharide references.

1.6 *Escherichia coli*

The Gram-negative bacteria *E. coli* can be both beneficial and detrimental to its host. The outer parts of Gram-negative bacteria consist of an inner membrane and an outer membrane.[19] A sugar and peptide network called the peptidoglycans is located in between the two membranes.[20,21] It consists of alternating *N*-acetyl glucosamine and *N*-acetyl muramic acid with peptides attached to position 3 of the lactyl moiety. This layer is thicker in Gram-positive bacteria since they lack the protection of the outer membrane. The outer membrane (OM) is an asymmetric layer where the outer leaflet is covered with lipopolysaccharides (LPS).[22] The layer in between is referred to as the periplasm.

Non-pathogenic strains of *E. coli* inhabit the gut of humans in a matter of minutes after birth[23,24] and pathogenic strains of *E. coli* may cause illnesses such as diarrheal disease, sepsis, and urinary tract infections.[25] Pathogenic strains of *E. coli* can be divided into different classes depending on the route of infection and infection mechanism: diffusely-adherent *E. coli* (DAEC),[26] enteroaggregative *E. coli* (EAEC),[27] enterohaemorrhagic *E. coli* (EHEC),[28] enteroinvasive *E. coli* (EIEC),[29] and enteropathogenic *E. coli* (EPEC).[30] For instance, several cases of infections involve the Shiga-toxin-producing class *E. coli* (STEC) O-antigen O157:H7,[31] which was responsible for numerous recent outbreaks.[32]

A related research topic is the study of intestinal microbiota[23] and how it affects the well-being of the host. A recent study by Lee *et al.* discovered that the specific strain, *E. coli* O16:H48, affected the maternal behavior in mice, which resulted in malnourished offspring.[33]

1.7 Lipopolysaccharides

The very first studies related to Lipopolysaccharides (LPS) date back to the middle of the 19th century. The Danish scientist Peter Ludvig Panum observed that injecting solutions of decomposed meat in dogs induced a lethal septic shock.[34] He named the solution 'putrid poison', and the substance was heat-resistant and insoluble in alcohol. He also observed tiny rod-shaped organisms referred to as *vibrios* in the solution, and he later suspected that the mysterious compound originated from the bacteria.

Richard Pfeiffer later coined the term 'endotoxin' for a substance that shared the putrid poison's physical and toxic properties derived from the Gram-negative bacteria *Vibrio cholerae*.[35] Nevertheless, it was not until the 1950s and 60s, that the LPS was discovered and analyzed more in-depth thanks to an optimized hot phenol/water extraction protocol,[36] chemical analysis, and studies of the biosynthetic pathways.[37]

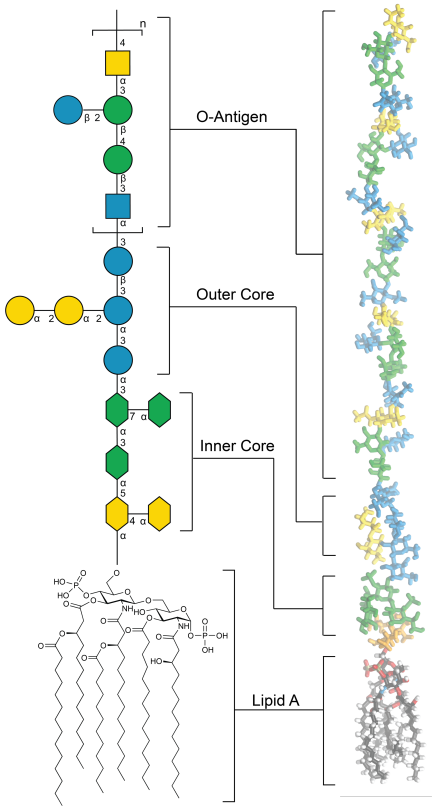


Figure 5. Schematic illustration of the LPS of *E. coli* O6 displayed in SNFG representation (left) and constructed in LPS Modeler[122] and visualized with VMD (right).[113]

The LPS consists of four parts (Figure 5): lipid A, inner core, outer core, and the O-antigen. The LPS constitutes the majority of the outer part of the outer membrane, and may cover up to 75% of the cell surface[7,38] providing Gram-negative bacteria with an efficient barrier against the host of the immune system.

Lipid A commonly has a D-Glc*p*N-(1→6)-D-Glc*p*N motif as a backbone that is substituted with phosphates, fatty acid chains, and additional sugars to various degrees. However, it is most commonly similar within species.[39]

For *E. coli*, the inner core is mainly constructed of heptose and Kdo monosaccharide residues. In contrast, the outer core consists of more common sugars (*e.g.,* D-Gal, D-Glc, and D-GlcNAc) with three sugars in its backbone and different additional monosaccharides present as side branches in different

configurations. Five different core regions have been observed in *E. coli* (R1, R2, R3, R4, and K12),[40] but are also present in other genera, *e.g., Salmonella* and *Shigella.*[41]

The O-antigen is the outermost part of the LPS consisting of polysaccharides with a specific repeating unit. It is the most common way for serotyping Gram-negative bacteria due to its high degree of structural variation.

The LPS is not synthesized on the surface of the outer membrane. Instead, complex biosynthetic pathways involving a plethora of proteins are used to construct and transport the large complex. The most common biosynthetic pathway for constructing the O-antigen is the Wzx/Wzy dependent pathway, where the construction of a single repeating unit takes place in the cytoplasm. The pathway starts with linking together an undecaprenyl lipid carrier and a phosphosugar. WecA catalyzes this process in *E. coli* by adding D-GlcNAc onto the undecaprenyl lipid. Serogroup-specific monosaccharides are then added depending on available GTs and NDP-sugar precursors. An epimerase, GlcNAc-P-P-Und epimerase (Gne), may also transform the initial D-GlcNAc into a D-GalNAc, if present in the serogroup.[42]

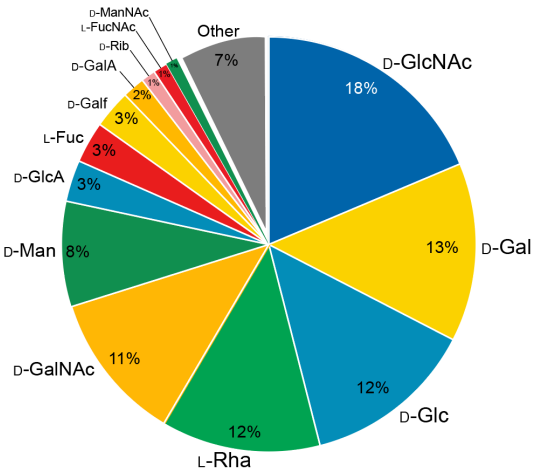
Upon construction of a polysaccharide repeating unit and its constituents, three key proteins, namely, Wzx, Wzy, and Wzz are involved in completing the O-antigen. Although the function and structure of these proteins are not fully understood, they are believed to have the following roles.[7,43,44]

- Wzx – flips an oligosaccharide being the precursor of the repeating unit into the periplasm.
- Wzy – connects the reducing end of the flipped oligosaccharide with the nascent polysaccharide.
- Wzz – regulates the total chain length of the polysaccharide.

The mature polysaccharide is then combined with lipid A and the core region by WaaL. Lastly, the entire LPS is transferred to the outer membrane by the LPS transport (Lpt) pathway. Lpt is a supercomplex of seven proteins that acts as a bridge for the LPS that is often referred to as the 'PEZ' model.[45,46] Recent studies even observed multiple lipopolysaccharides bound inside the Lpt complex in Cryo-EM[47] and X-ray studies.[48]

1.8 Structural diversities in *E. coli* O-antigens

E. coli O-antigens have high structural diversity. The repeating units vary in size with between two and seven sugars in their repeating unit and may contain up to three branch points. A five sugar residue repeating unit with a four sugar residue backbone is the most common topology.[49] The monosaccharides themselves also vary significantly, with a total of 49 different monosaccharides found so far. The most abundant ones are D-GlcNAc (18%), D-Gal (13%), and D-Glc (13%) (Fig. 6).



Topology	Abundance (%)
	3
	1
	<1
	7
	10
	3
	4
	2
	<1
	20
	28
	2
	4
	<1
	11
	3
	<1
	<1

Figure 6. Monosaccharide abundance distribution and colored similar to the SNFG-representation[50] (left) and the abundance of different repeating unit topologies (right) in *E. coli* O-antigen repeating units. (modified from Liu *et al.*).[49]

1.9 Nuclear Magnetic Resonance spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is nowadays an essential tool for organic chemists. The discovery awarded Bloch and Purcell the Nobel Prize in 1952[51,52] after the implementation of Fourier transform NMR and multi-dimensional NMR pulse sequences. It evolved into a powerful tool for elucidating complex biological macromolecules and awarded the Nobel Prize to Ernst in 1992[53] and Wüthrich in 2002.[54]

The technique utilizes the fact that some atomic nuclei, *e.g.,* ^{1}H and ^{13}C, have a spin-magnetic moment and when placed in a strong magnetic field, they can both absorb and re-emit energy in the range of radiofrequencies.

The read-out in an NMR spectrum is the chemical shift. It is the resonant frequency of a spin active nucleus relative to the frequency of a reference standard. The chemical shift reflects the local environment of the nucleus which is why it can be utilized as a powerful structure elucidation tool.

1.10 Nuclear Overhauser Effect

The Nuclear Overhauser effect (NOE) is generated when one spin, S, is saturated, which will affect a close-in-space spin, I. This measurable effect is due to cross-relaxation, σ, and can be used to obtain proton-proton distances when the isolated spin pair approximation (ISPA) is applicable:

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

where, r_{ij}, is the calculated distance between spins i and j derived from the measured cross-relaxation rates of the cross-peak between spins i and j, σ_{ij}, a reference cross-peak, σ_{ref}, and a reference distance, r_{ref}.

1D ^{1}H,^{1}H-NOESY and 2D ^{1}H,^{1}H-NOESY with short mixing times are suitable experiments for this approach. For more accurate measurement of proton-proton distances, several mixing times can be used to construct NOE build-up curves and calculated the cross-relaxation rates using the PANIC[55] or devised by Dixon *et al*.[56]

1.11 Structure elucidation of carbohydrates by NMR spectroscopy

Due to their already mentioned complex structure, any kind of carbohydrate often needs an extensive amount of NMR experiments to elucidate its structure. There are thus some general rules which simplify the process.[57]

The main bulk of ^{1}H and ^{13}C NMR chemical shifts resonances in carbohydrates, consisting of positions 2 to 5 in hexoses, spans from 3.0 to 4.5 ppm in ^{1}H NMR chemical shifts and from 65 to 85 ppm in ^{13}C NMR chemical shifts (Figure 7). For both proton and carbon chemical shifts the anomeric region resides downfield compared to positions 2-5 with δ_H and δ_C ranging from 4.0 to 6.0 ppm and 95-110 ppm, respectively.[57] The hydroxymethyl groups of hexoses, position 6, are often slightly upfield, at 60-65 ppm δ_C, whereas methyl groups of 6-deoxy sugars and acetyl groups occupy a region δ_C of 18-30 and δ_H of 1.2-2.2.

Positions 2 of *N*-acetylated aminosugars are shifted downfield in ^{13}C NMR spectra with δ_C of 55-60, and *O*-acetylated positions commonly have a distinct downfield change in ^{1}H chemical shifts with a $\Delta\delta_H$ of 0.5-1.

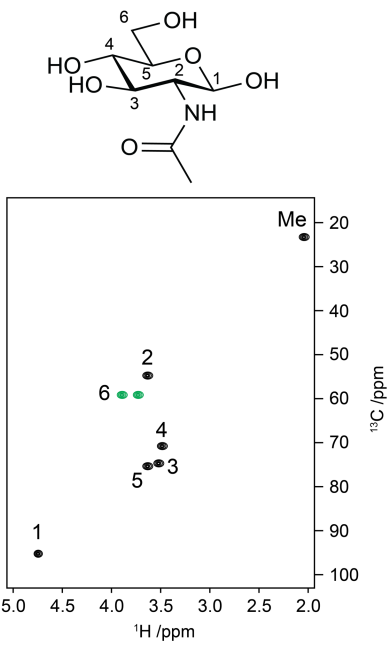


Figure 7. Schematic $^{1}H,^{13}C$-HSQC NMR spectrum of β-D-Glc*p*NAc.

As a 'rule of thumb', glycosylated positions are perturbed downfield by 5-10 ppm in ^{13}C NMR spectra, and neighbouring positions go upfield with 0-2 ppm. The corresponding chemical shifts in ^{1}H NMR spectra, deviate by ±0.25 ppm when compared to similar, unsubstituted, sugar residues.

To fully assign any type of sugar residue, $^{1}H,^{1}H$-TOCSY and $^{1}H,^{13}C$-HSQC-TOCSY are valuable NMR experiments since only cross-peaks within every sugar residue will be observed. An array of spectra with increasing mixing time can be acquired to identify the different spin systems for each sugar residue. These experiments utilize a spin-lock during which transfer the magnetization based on the size of the $^{3}J_{Hn,Hn+1}$ coupling takes place. Therefore, differently configured monosaccharides can be differentiated in the assignment based on the extent of observed cross-peaks from their anomeric proton.[58] The anomeric configuration of a sugar residue in pyranose form with an axial proton in position 2 (H2) (*e.g.* g*luco*- or *galacto*-configured sugar-

residues) can be determined by measuring the $^{3}J_{H1,H2}$ coupling constant.[57] For these sugar-residues, a large coupling constant (7-8 Hz) corresponds to a β-configured sugar, whereas an α-configured has a smaller coupling (3-4 Hz). In pyranoses with an equatorial H2 (*e.g.* D-Man) or if spectral overlap in the ^{1}H NMR is a problem – the $^{1}J_{C1,H1}$ coupling constant can instead be measured from, for instance, a coupled ^{1}H,^{13}C-HSQC or a ^{13}C-INEPT with the following rule for the $^{1}J_{C1,H1}$: α-anomer > 169 Hz > β-anomer. F_2-coupled ^{1}H,^{13}C-HSQC experiment should be analyzed with caution when strong couplings are present since the strong couplings can cause an error in the $^{1}J_{C1,H1}$ measurement of up to 4 Hz.[59]

A combination of ^{1}H,^{13}C-HMBC and ^{1}H,^{1}H-NOESY experiments can establish inter-residual connectivity to determine how the sugar residues are linked together. In ^{1}H,^{1}H-NOESY, these inter-residue correlations can be distinguished from intra-residue correlations by comparing the ^{1}H,^{1}H-NOESY cross-peaks to the ones in the ^{1}H,^{1}H-TOCSY (Figure 8). Similarly, inter-residual $^{3}J_{C,H}$ correlations across the glycosidic linkage can be extracted by comparing the cross-peaks observed in ^{1}H,^{13}C-HMBC with the ones observed in a ^{1}H,^{13}C-H2BC.[60]

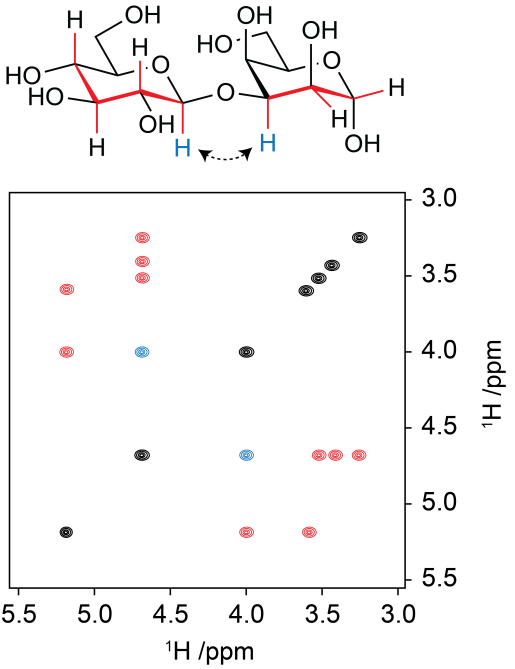


Figure 8. Overlay of schematic ^{1}H,^{1}H-NOESY and ^{1}H, ^{1}H-TOCSY NMR spectra of the disaccharide β-D-Glc*p*-(1→3)-α-D-Gal*p* with selected cross-peaks coloured in blue and red, respectively.

1.12 CASPER

Assigning the obtained experimental NMR chemical shifts of carbohydrate structures can be time-consuming and complicated despite having access to modern NMR spectrometers that can acquire multi-dimensional NMR experiments and even multiple experiments in parallel to save time.[61] Therefore any tool that can automate parts of this process will be useful for a carbohydrate chemists.

The computer program, CASPER (acronym for computer-assisted spectrum evaluation of regular polysaccharides), with its first rendition being developed in 1987[62] with subsequent updates ever since for instance to automate the assignment of complex human milk oligosaccharides[63] and to become more accessible with the implementation of a web-based version in 2006.[64]

Other NMR carbohydrate structure prediction program have since then been developed such as GlyNest,[65] GODESS[66] and GRASS.[67]

The structure prediction in CASPER is based on a database of collected mono-, di- and trisaccharide ^{1}H and ^{13}C NMR chemical shifts. In order to obtain more accurate values of the chemical shifts, a simulated ^{1}H NMR spectrum is adjusted by iteratively fitting it to the experimental ^{1}H-NMR spectrum until a satisfactory RMS-value has been reached using the computer software PERCH.[68] In addition, the $^{3}J_{HH}$ for each proton pair is also obtained since it is one of the important variables being adjusted by PERCH in order to optimize the fitting.

The obtained chemical shifts of the monosaccharides provide CASPER with a starting point for the prediction whereas the di- and trisaccharide datasets are used to construct what is called ‘correction sets’. These adjust the values of the chemical shifts for each monosaccharide of a constructed di-, tri-, oligo- or polysaccharide that is requested by the user to be predicted. The role of the disaccharide glycosylation sets and to predict the significant changes in ^{1}H and ^{13}C chemical shift that takes place at the linkage position and positions close-by.

The trisaccharide correction sets are needed to better predict the chemical shift of vicinal disubstituted residues, *e.g.*, a 3,4-substituted GlcNAc residue.

The database also contains datasets that account for substituents, *e.g.,* an *O*-acetyl groups or methyl substituents which will also affect the chemical shift of the monosaccharide.

1.13 Conformational analysis

As mentioned before, monosaccharides can exhibit fast conformational changes between multiple conformations. This, too, applies to longer oligo- and polysaccharides. These changes often occur faster compared to the NMR time scale. Therefore, time-averaged values, *e.g.*, scalar coupling constants and distance measurements through ^{1}H,^{1}H-NOESY, are obtained. CarbBuilder[69] is a recently developed software that can rapidly construct oligo- or polysaccharide structure models of interest. To simplify calculations, the software treats each monosaccharide as a rigid building block and contains multiple sets of ϕ, ψ dihedral angle values for several different types of disaccharide linkages which are iterated and adjusted until the entity disaccharide contains no steric clashes.

1.14 From cell cultivation to NMR samples

In the 1950s and 60s structure elucidation of polysaccharide structures mostly involved an array of different chemical degradations and analyses by GC. Since then, NMR has slowly taken over the structure elucidation process.

In order to fully assign these complex biomolecular structures, extraction methods to obtain purer materials are still essential for the structural elucidation of O-antigens. A flowchart showing the general workflow of a modern, NMR-based, structure elucidation process is shown in Figure 9.

The process usually involves the hot phenol/water extraction protocol optimized by Westphal and co-workers[36] and described by De Castro.[70] The extraction separates the lipophilic cell debris (phenol phase) from the water-soluble LPS.

Mild acidic delipidation is then performed to improve the solubility of the polysaccharide in D_2O.[71] In this process, the acid-sensitive α-Kdo-(2→6)-β-D-Glc*p*N bond between the inner core and lipid A is cleaved. Further purifications of the delipidated polysaccharide (PS) may include size exclusion chromatography and enzymatic treatments (*e.g.,* DNase and RNase) to remove unwanted impurities. *O*-deacetylation using alkaline conditions can also be performed to simplify structure elucidation if *O*-acetylation is observed during initial NMR analysis of the LPS.

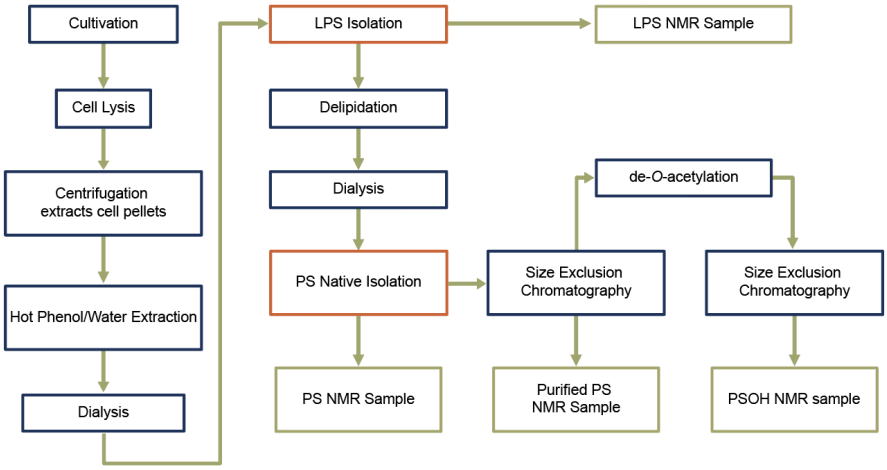


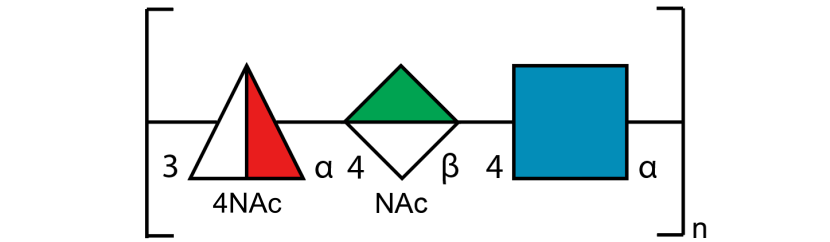
Figure 9. Flowchart of a typical lipopolysaccharide structure elucidation process.

1.15 Enterobacterial common antigen (ECA)

ECA is one of many other polysaccharides produced by Gram-negative bacteria (Figure 10) and was discovered in the 1960s.[72]

Similar to the O-antigen, the presence of ECA may affect the virulence factor of bacteria by, for instance, improving its resistance against bile salts thus, increasing the chances of an oral infection route.[73,74]

It is present in three forms, where two of them are linear polysaccharides just attached either to the cell membrane, ECA_{pg} or, attached to the core oligosaccharide, ECA_{LPS}, where the latter is attached to the core similar to the O-antigen. Lastly, ECA_{cyc} is a 4–6 repeating unit cyclic form of ECA.[75]



→)3-α-D-Fuc*p*4NAc-(1→4)-β-D-Man*p*NAcA-(1→4)-α-D-Glc*p*NAc-(1→

Figure 10. SNFG representation of the ECA repeating unit.

Gel-permeation chromatography is a useful technique to separate ECA from the PS since ECA usually has a lower molecular weight (M_w) than the PS (Figure 11a), and its triangular constellation in the anomeric- and the *N*-substituted ring proton regions is easily spotted in $^{1}H,^{13}C$-HSQC NMR spectra (Figure 11b).

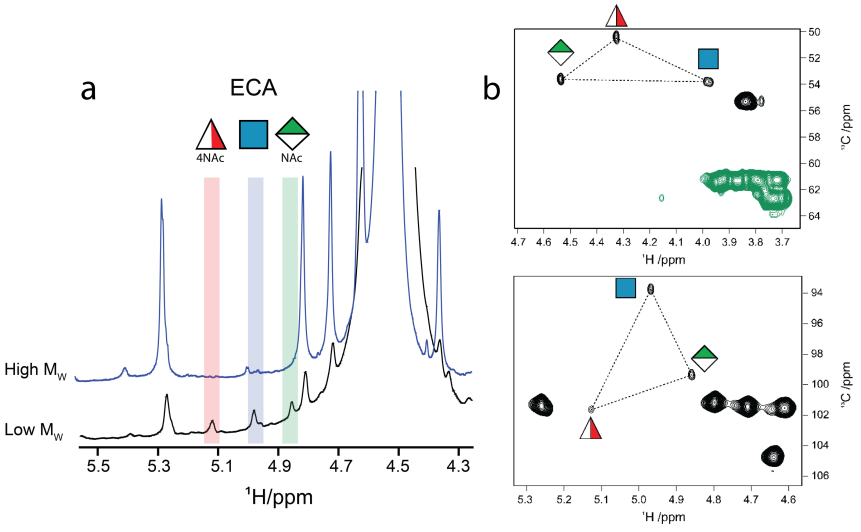


Figure 11. (a) ^{1}H NMR spectra of two different fractions from size exclusion chromatography run of the O188 PS. The highlighted proton signals correspond to each of the ECA monosaccharide residues shown with SNFG representation. (b) Selected regions from a $^{1}H,^{13}C$-HSQC NMR of the O188 LPS showing the presence of ECA signals easily identified due to their triangular constellations.

1.16 Aim of this Thesis

The aim of this thesis is to investigate structural and conformational aspects of different O-antigen structures isolated from the Gram-negative bacteria *Escherichia coli* and *Shigella flexneri.* The aim is also to further develop and, improve the performance of the carbohydrate structure prediction program, CASPER, by providing its database with additional datasets and correction sets.

In particular, the aim is to elucidate the structure of previously unknown *E. coli* serogroups O93, O125ac and O188 by utilizing CASPER in different ways in order to streamline the structure elucidation process and compare the elucidated structures with other, structurally similar serogroups (Chapter 2-4).

Another goal is to further improve and update the database of CASPER by fully assign and then implement ^{1}H and ^{13}C NMR chemical shift data from mono-, di- and tri-saccharides into the database (Chapter 5).

A final aim is to study conformational aspects of *Shigella flexneri* O-antigen serotypes 7a and 7b using NMR-experiments and MD-simulations and to compare the results with corresponding O-antigen structures of other *Shigella flexneri* serotypes (Chapter 6).

The main technique used in this thesis to achieve this goal is NMR spectroscopy with additional analytical methods, such as GC-MS, Bioinformatics, CASPER and MD-simulations used in conjunction with the NMR-studies.

2 *E. coli* O188 O-antigen (Paper I)

2.1 Introduction

The *E. coli* serogroup O188 is a newly added strain (*E. coli* O188:H10) with an unknown O-antigen repeating unit. Its gene sequence has not been elucidated but is still officially recognized as an *E. coli* O-antigen reference strain.

This chapter will summarize the structure elucidation process of the polysaccharide of *E. coli* O188 with a focus on the NMR analysis presented in Paper I.

2.2 Results and discussion

2.2.1 Sample preparation and initial NMR-analysis

After cultivation and subsequent centrifugation to isolate the cell pellets, the LPS of the *E. coli* O188 O-antigen was successfully obtained using the above-mentioned hot phenol/water extraction method, followed by dialysis (section 1.14).

Initial studies using $^{1}H,^{13}C$-HSQC, and ^{1}H-NMR of the LPS material did not reveal any *O*-acetyl groups usually present in the δ_H 2.1-2.2 region.

The LPS was then delipidated using mild acidic condition followed by dialysis to obtain delipidized PS material for further studies. In addition, gel-permeation chromatography was performed on a portion of the PS material in order to obtain an NMR sample with lower content of impurities, *i.e.*, ECA.

2.2.2 Component Analysis

A sugar analysis was conducted using alditol acetates obtained after hydrolysis, reduction, and per-*O*-acetylation of the polysaccharide (PS) material. The observed peaks in the GLC chromatogram had retention times corresponding to alditol acetates of Gal, Glc, Man, and GlcNAc (Figure S1 Paper I).

Furthermore, a sample of the PS material was 2-butyl glycosylated and subsequently per-*O*-acetylated and analyzed by GLC. In this case the observed

peaks in the GLC chromatogram corresponded to D-Gal, D-GlcA, D-GlcNAc, and D-Man.

2.2.3 NMR Studies

Five anomeric proton resonances in the region of 4.5-5.3 ppm were observed in the ^{1}H NMR spectrum and in the ^{1}H,^{13}C-HSQC spectrum (Figure 12) and are annotated **A**-**E** based on descending δ_H.

An array of ^{1}H,^{13}C-HSQC-TOCSY experiments with mixing times ranging from 30 to 120 ms were performed on the delipidated and dialyzed PS sample (Figure 10). Residue **A** has three correlations at 120 ms, indicating *galacto*-configuration. Residues **B** and **C** could quickly be confirmed as *manno*-configured due to their lack of cross-peaks from the respective anomeric signal. Both residues **D** and **E** appear to have *gluco*-configuration due to a total of 5 cross-peaks each from their respective anomers.

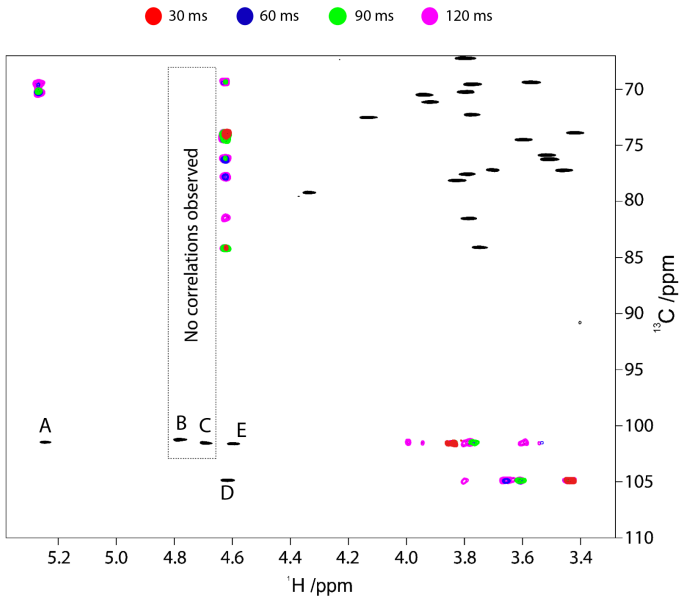


Figure 12. ^{1}H,^{13}C-HSQC spectrum overlayed with anomeric cross-peaks from ^{1}H,^{13}C-HSQC-TOCSY spectra with mixing times of 30 (red), 60 (blue), 90 (green) and 120 ms (purple).

A combination of ^{1}H,^{13}C-H2BC and ^{1}H,^{13}C-HMBC helped assigning position 2 of residues **B** and **C** to δ_H/δ_C 4.35/79.26 and δ_H/δ_C 3.93/71.11, respectively. The cross-peak is related to a large $^{2}J_{H1,C2}$ of ≈8 Hz[76,77] and is visible in the ^{1}H,^{13}C-HMBC. The remaining parts of both spin-systems could then be assigned with the ^{1}H,^{13}C-H2BC, and the ^{1}H,^{13}C-HSQC-TOCSY experiments (Table 1, Paper I).

The assignment of the two observed carbonyl groups at δ_C 174.18 and 175.40 in the ^{13}C NMR was possible with a $^{1}H,^{13}C$-BS-CT-HMBC experiment (Figure 13) showing the correlation from C6 to H5 for residue **D** and CO to H2 for residue **E**.

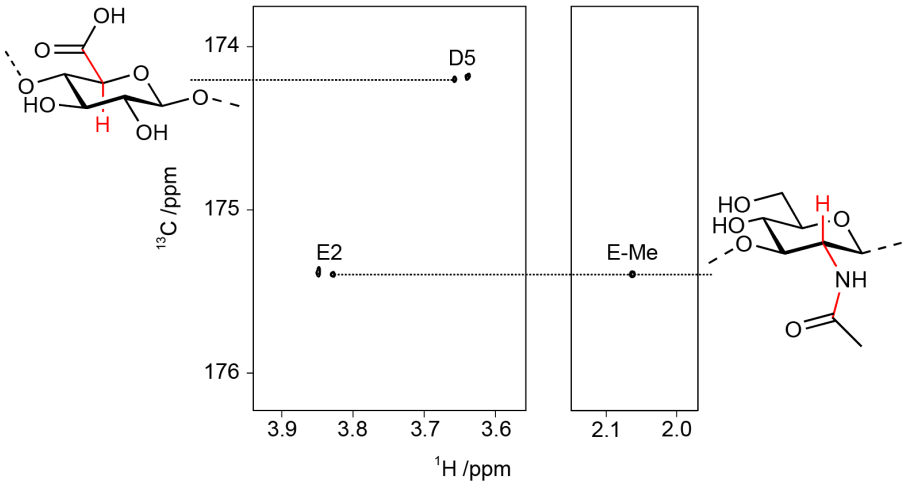


Figure 13. Selected regions of a $^{1}H,^{13}C$ BS-CT-HMBC spectrum from the O188 PS with F_1 centred in the carbonyl region; δ_C 175.40 shows a cross-peaks to the H2 of residue **E** (β-D-Glc*p*NAc), and δ_C 174.18 and shows a cross-peak to H5 of residue **D** (β-D-Glc*p*A).

The fact that position 5 of residue **D** had a significant change in δ_C/δ_H when performed at pD 4 and pD 7 confirms the presence of a close-by carboxyl group. The change in pD also slightly affected the δ_H of **D1**, which had a slight downfield displacement at pD 4. This chemical shift displacement was sufficient to determine the $^{3}J_{H1,H2}$ for both **D1** (7.9 Hz) and **E1** (8.5 Hz) (Figure 3b, Paper I) and made their cross-peaks along F_1 distinguishable which simplified the assignment.

The ^{13}C NMR chemical shifts of the polysaccharide were compared to their respective monosaccharide chemical shifts to determine their linkage positions.[78] The δ_C of positions, **B2**, **B3**, **C4**, **D4**, and **E3** (Table 1, Paper I) were significantly shifted downfield in comparison to their respective monosaccharide. Correlations *to* these linkage positions were confirmed with $^{1}H,^{1}H$-NOESY and $^{1}H,^{13}C$-HMBC *from* the anomeric positions (Table 2, Paper I).

The $^{1}H,^{13}C$-HSQC-TOCSY with a 20 ms mixing was particularly useful for assigning position 6 of residues **B**, **C**, and **E** (Figure 5, Paper I). This is an excellent alternative to $^{1}H,^{13}C$-H2BC when an improved resolution in F_1 is needed since the $^{1}H,^{13}C$-H2BC is a constant-time experiment thus limiting the increment range in the indirect dimension.

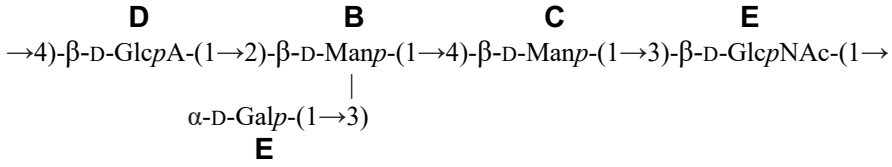
2.2.4 Structural comparisons with *Shigella boydii* type 16

The gene sequence of *E. coli* serogroup O188 has not yet been revealed and can thus not be fully compared to the gene sequencies of other *E. coli* serogroups. Although initial studies show that it agglutinates with *S. boydii* type 16 antiserum which confirms the similarities in structure and thus also the gene sequence of *S. boydii* type 16. *Liu et al.* observed the presence of four glycosyltransferases and genes involved in the biosynthesis of GDP-D-Mannose (*manB*, and *manC*) which further validates the assignment of the O188 O-antigen repeating unit.[79]

2.3 Conclusion and future outlook

The O188 O-antigen repeating unit was determined by using NMR spectroscopy as the primary tool. The sample with the lower pD, simplified the assignment process since the anomers of residue **D** and **E** were separated in δ_H. The high resolution of $^{1}H,^{13}C$-HSQC-TOCSY with a 20 ms mixing was essential for assigning the position 6 of residues **B**, **C** and **E**.

The *E. coli* O188 O-antigen repeating unit could be determined as:



and it is thus a member of the most common topology in *E. coli* O-antigens with a four-sugar backbone and one branched sugar residue.[49] A structural comparison with other known O-antigen repeating units showed that the O188 O-antigen repeating was shared with *Shigella boydii* type 16.[79] In addition, the *S. boydii* type 16 O antigen repeating unit also contain an O-acetyl substituent in position 6 of the branched β-D-Man*p* (Residue B).

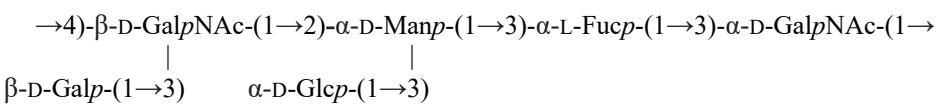
Further studying how the additional *O*-acetyl group in *S. boydii* type 16 affects the conformation of the polysaccharide and the survivability of the bacteria in the host would be of great interest. A comparison of the gene clusters of *E. coli* O188 and *S. boydii* type 16 would also deepen the knowledge of the gene-structure relationship in bacteria.

3 Structural studies of *E. coli* O125ac and O125ab O-antigen (Paper II)

3.1 Introduction

Serogroup O125 of *E. coli* belongs to the group of enteropathogenic *E. coli* (EPEC) O-serogroups and was initially discovered in 1952 by Taylor and Charter due to an outbreak of acute diarrhea in a children's daycare center in London.[80] The serogroup of *E. coli* O125 can further be subdivided in two groups, O125ab and O125ac. Over the years, several different serotypes of serogroup O125 have been involved additional outbreaks.[81]

The structure of *E. coli* O125ab O-antigen was previously determined by Kjellberg *et al*. in 1996[82] using NMR analysis, methylation analysis and a Smith degradation product. The elucidated O-antigen had an unusual highly branched hexasaccharide repeating unit with adjacent disubstituted branch-point residues:



This chapter will focus on the unknown O-antigen structure of *E. coli* subgroup O125ac and make structural comparisons with the O125ab O-antigen structure.

3.2 Results and discussion

Strains of O125ac (*E. coli* O125ac:H6) and O125ab (*E. coli* O125ab:H19) were cultivated and extracted using the same protocol as described (Section 1.14, Paper 1).

3.2.1 Sugar Analysis

A conducted sugar analysis revealed fucose, galactose, galactosamine, glucose, and mannose for both O125ab and O125ac. Interestingly, the relative integral of the peak related to the glucose residue observed with GLC-MS was significantly lower in O125ac compared to in O125ab which may indicate a lower degree (or absence) of the glucose residue in the repeating unit (Figure 14).

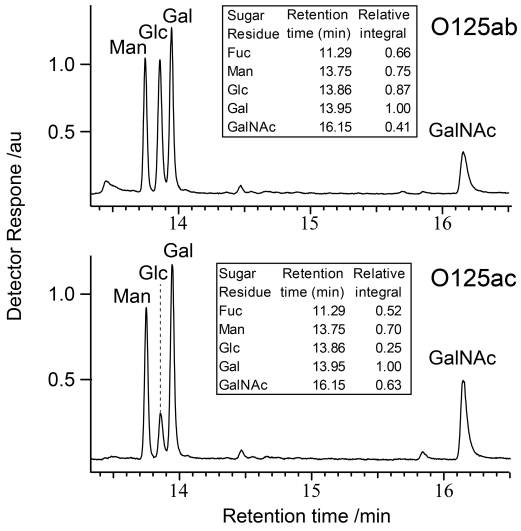


Figure 14. Selected regions of GLC-MS chromatograms of O125ab and O125ac with relative integrals for the different sugar-residues displayed.

3.2.2 NMR analysis

The elucidation for the O125ac O-antigen polysaccharide then proceeded with NMR analysis. Five resonances were observed in the anomeric region of the ^{1}H,^{13}C-meHSQC NMR spectrum compared to the six resonances observed for the O125ab O-antigen polysaccharide. The resonances were annotated based on descending proton NMR chemical shifts, **A-E** (Figure 15).

The ^{1}H,^{13}C-meHSQC NMR spectrum also revealed two signals resonating at δ_H/δ_C 2.07/23.01 and 2.05/23.36, and one signal at δ_H/δ_C 1.22/16.05 which are characteristic for *N*-acetylated and 6-deoxy sugar-residues, respectively.

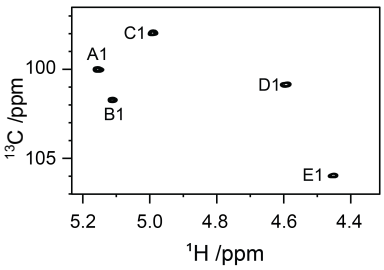


Figure 15. Selected region of the anomeric region of a ^{1}H,^{13}C-meHSQC NMR spectrum of the O125ac O-antigen polysaccharide.

A ^{1}H,^{1}H-TOCSY NMR spectra acquired with mixing times of 30-200 ms helped to determine the configuration of each sugar-residue and are summarized in Table 1. Residue **A** was determined to be of *manno*-configuration since the transfer ended at H2 which is caused by the axially-oriented proton resulting in a $^{3}J_{H1,H2} < 2$ Hz. Interestingly, the transfer reached H6b from the ^{1}H,^{1}H-TOCSY with a 200 ms mixing time which is expected since the remaining $^{3}J_{HH}$ are large.

In a similar way, the four other sugar-residues were all determined to be *galacto*-configured since the magnetization transfer ended at H4 of the respective sugar-residue.

Table 1. Furthest transfer of magnetization from the anomeric proton observed in ^{1}H,^{1}H-TOCSY NMR at different mixing times and concluded ring-configuration.

Correlation from anomeric proton in spin-system	To furthest correlation in ^{1}H,^{1}H-TOCSY at mixing time (ms)					Configuration
	30	**60**	**90**	**120**	**200**	
A	H2	H2	H2	H2	H6b	*manno*
B	H2	H4	H4	H4	H4	*galacto*
C	H2	H3	H4	H4	H4	*galacto*
D	H3	H3	H4	H4	H4	*galacto*
E	H3	H4	H4	H4	H4	*galacto*

An F_2-coupled ^{1}H,^{13}C-coupled-HSQC then determined the anomeric configuration with measured $^{1}J_{CH}$ of 172, 172, 174, 165 and 162 Hz for residues **A**, **B**, **C**, **D** and **E,** respectively, and where residues **A-C** have the α-configuration and **D-E** have the β-configuration.

2D NMR experiments (^{1}H,^{13}C-HMBC and ^{1}H,^{13}C-H2BC) were also acquired and their peak-picked resonances were used as input for CASPER in order to predict the structure of the *E. coli* O125ac O-antigen polysaccharide along with Biological rules (WecA), component analysis (GLC-MS) and magnitude of $^{3}J_{H1,H2}$ (Scheme 1, Paper II).

The highest-ranked structure resembled the one previously reported of O125ab but lacked the branched glucose residue that was linked to the mannose (Figure 16). The second- and third-ranked structures have 'normalized relative rankings' of 1.20 and 1.25 which are significantly larger.

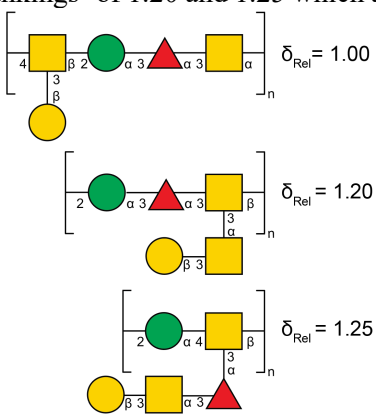


Figure 16. Top 3 ranked structures predicted by CASPER for the *E. coli* O125ac O-antigen.

Only 5 out of the 31 proton resonances had a δ_H chemical shift deviation of more than 0.1 ppm for proton NMR chemical shifts compared to the chemical shifts predicted by CASPER with four of them originating from the GalNAc

residue **C** (Table 2). This demonstrated that the prediction of the 3,4 di-substitued GalNAc (residue **D**) is well predicted by the correction sets in CASPER but also highlights possibilities of improvements regarding the prediction of GalNAc residue **C**.

Table 2. Comparison of experimental and calculated δ_H chemical shifts for residues with a $\Delta\delta_H > 0.1$.

Position	δ_H Expt	δ_H Calc	$\Delta\delta_H$
C2	4.44	4.21	0.23
C4	4.14	4.35	0.21
C3	4.09	4.28	0.19
C1	4.99	5.15	0.16
B3	4.02	3.91	0.11

CASPER could also successfully predict the location of the glucose residue linkage in the structure of O125ab (Figure 4, paper II) when including data from ^{1}H,^{13}C-HSQC and ^{1}H,^{13}C-HMBC.

The elucidation process of *E. coli* O125ac then proceeded with manual assignment of the structure to confirm what CASPER suggested and the full resonance assignment is summarized in Table 3.

Table 3. ^{1}H and ^{13}C NMR chemical shifts of the O-antigen from E. coli O125ac.

	Sugar residue	1	2	3	4	5	6	Me	CO
A	→2)-α-D-Man*p*-(1→	5.16 100.08	4.21 77.71	3.95 70.37	3.58 68.39	3.74 74.33	3.66, 3.90 62.58		
B	→3)-α-D-Fuc*p*-(1→	5.11 101.79	3.89 68.46	4.02 77.98	3.97 72.42	4.18 67.88	1.22 16.05		
C	→3)-α-D-Gal*p*NAc-(1→	4.99 98.09	4.44 49.85	4.09 76.38	4.14 69.27	4.50 70.90	~3.80 60.82	2.07 23.01	175.37
D	→3,4)-β-D-Gal*p*NAc-(1→	4.59 100.91	4.20 52.63	3.98 78.94	4.29 74.43	3.75 76.31	3.72, 3.74 61.29	2.05 23.36	175.75
E	β-D-Gal*p*-(1→	4.45 106.00	3.54 71.55	3.63 73.48	3.93 69.64	3.66 75.90	3.77, 3.80 61.97		

When comparing the assigned resonances of the two polysaccharides, a significant downfield displacement was observed for carbon chemical shift in **A3** in the polysaccharide of O125ab, which further confirms that the structure is similar to the previously studied O125.[82]

The suggested sequence was determined with a combination of ^{1}H,^{13}C-HMBC and ^{1}H,^{1}H-NOESY NMR spectra and was confirmed to have a backbone consisting of [-**D**-**A**-**B**-**C**-]$_n$ with the side-chain **E** linked to **D** (Figure 15).

Apart from the glycosidic correlations, a strong inter-residue correlation, from **A5** to **B4,** was observed in the ^{1}H,^{1}H-NOESY for both O125ac and O125ab (Figure 7a, Paper II). This correlation fitted well with an *exo*-anomeric conformation of the involved glycosidic linkage (**A1**-**B3**).

^{1}H,^{13}C-meHSQC and ^{1}H,^{13}C-HSQC-TOCSY NMR experiments made it possible to decipher the assignment of all hydroxymethyl resonances.

Interestingly, a larger proton chemical shift difference between protons of **D6a** and **D6b** was observed resonating at δ_H 3.72 and 3.74 in O125ac versus δ_H 3.72 and 3.78 respectively in O125ab. This may be explained by a change in the rotameric distribution of ω (O5-C5-C6-O6) torsion angle in residue **D** (branched →3,4)-β-D-GalpNAc-(1→) most likely due to the added sterical interactions caused by the α-D-Glc*p*-(1→ (residue **G**).

3.3 Conclusion and future outlook

The O-antigen structure of *E. coli* O125ac was elucidated using a combination of NMR spectroscopy and GLC-MS analysis (Figure 17). Furthermore, CASPER was a powerful tool in the elucidation process. In addition, the hydroxymethyl group spectral region for the O-antigen structure of *E. coli* O125ab was fully elucidated using high resolution $^{1}H,^{13}C$-meHSQC and $^{1}H,^{13}C$-HSQC-TOCSY NMR spectra.

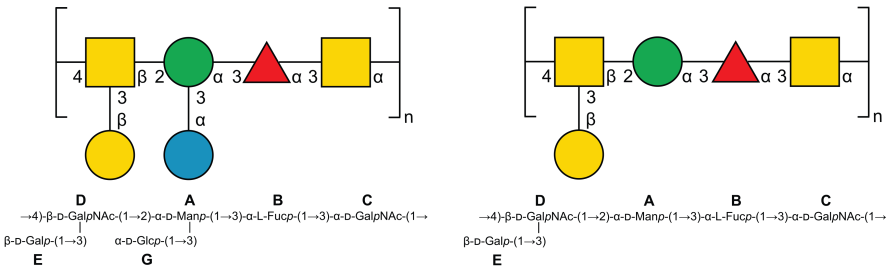


Figure 17. The polysaccharide structures of *E. coli* O125ab (left) and O125ac (right) O-antigens.

4 Structural studies of *E. coli* O93 O-antigen (Paper III)

4.1 Introduction

The *E. coli* O93 serogroup was recently re-introduced and available from SSI Diagnostica as a reference strain[83] after initially being excluded due cross-reacting with serogroup O8.[84,85] Thereafter, its gene sequence was investigated in studies by Iguchi *et al*.[86]

In this chapter the structural elucidation of the *E. coli* O93 O-antigen polysaccharide will be investigate and follows a slightly different approach compared to two above-mentioned studies by utilizing both bioinformatics in combination with NMR studies.

4.2 Results and discussion

4.2.1 Initial observations

The LPS material was obtained using the extraction protocol described in the *E. coli* O188 O-antigen structural study (Chapter 2 and Paper I). Initial studies of the LPS by ^{1}H and ^{1}H,^{13}C-HSQC NMR revealed multiple resonances in both the anomeric region and in the *O*-acetyl/*N*-acetyl regions, but it was too heterogeneous for further in-depth analysis. Therefore, the study proceeded with a mild acidic delipidation to obtain a better, lipid-free PS, while also retaining any potential *O*-acetyl groups.

A sugar analysis conducted on the PS material of the *E. coli* O93 O-antigen structure revealed peaks with retention times corresponding to the reference standards of mannose, glucose and glucosamine alditol acetates.

4.2.2 Bioinformatics

E. coli O93 O-antigen was sequenced in 2015 by Iguchi *et al.* and the gene cluster nearby the housekeeping genes *galF* and *gnd* are displayed below (Figure 18).[49] Two glycosyltransferases (GT), *wfdV* and *wfdW,* were identified along with GDP-mannose pathway genes, *manB* and *manC* and one gene, *ugd,* known to function as a UDP-glucose 6-dehydrogenase which forms D-GlcA*p*-UDP. The genes responsible for coding *Wzx* (flippase)*, Wzy* (polymerase) and *Wzz* (chain-length modulator) were also observed and therefore the *E. coli*

O93 O-antigen is assumed to follow the *Wzx*/*Wzy* dependant pathway when constructing its lipopolysaccharides. No genes encoding for *gne* of *gnu* were identified meaning that the repeating unit is initialized by a D-GlcNAc.[87,88] The first GT, *wfdV*, was identified through a protein BLAST search to belong to the GT family 2 which is of inverting mechanism. The sequence was also added into ECODAB[89] and a BLASTp search revealed a close similarity to the GT, *WeiG*, in O161 which is responsible for a β-D-Glc*p*A-(1→3)-D-GlcNAc linkage. The second identified GT, *wfdW,* belongs to GT family 1 which is inverting.

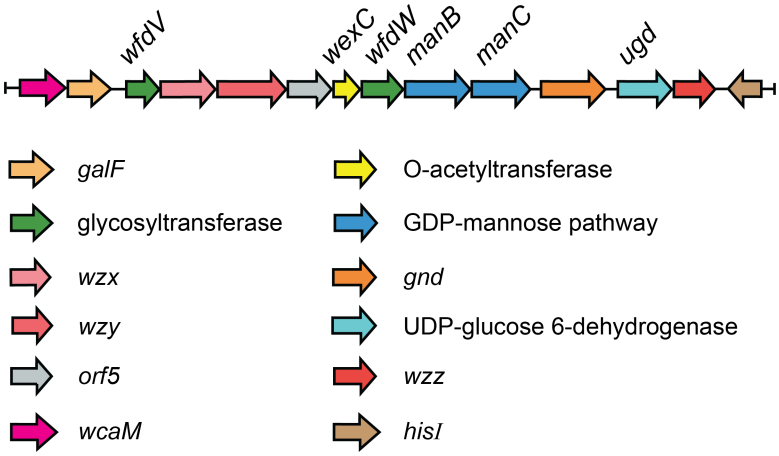


Figure 18. O-antigen gene cluster of *E. coli* O93 sequenced by Iguchi *et al*.[86] and annotated according to Liu *et al*.[49]

4.2.3 NMR Studies

After a performed gel-permeation chromatography to lower the content of ECA, a 10 mg PS sample was prepared for further NMR experiments. 2 mg of the PS-material was in addition *O*-deacetylated to obtain the PS-OH material of <1 mg for further NMR studies.

^{1}H NMR of the PS shows the presence of *O*-acetyl groups in the region of δ_H 2.15-2.25 and also *N*-acetyl group resonating at 2.00-2.05 ppm which indicates the presence of D-GlcNAc residues.

A ^{1}H,^{13}C-HSQC NMR spectrum of the PS-OH material shows four resonances in the anomeric region which are denoted as **A**-**D** based on descending proton chemical shifts. A total of 27 resonances were observed with, four anomeric resonances, one nitrogen-bearing ^{13}C resonance at δ_H/δ_C 4.06/53.70, six resonances in the hydroxymethyl group region at 60-64 ppm for ^{13}C, 16 resonances found in the sugar-ring region at δ_C 65-85 and one methyl group resonance characteristic of *N*-acetyl groups at δ_H/δ_C 2.04/22.87.

The anomeric region in the ^{1}H-NMR spectrum revealed that H-1 of residue **B** and **C** are of *manno*-configuration due to their small $^{3}J_{HH}$ estimated to be less than 2 Hz based on their narrow full width at half-height. The peak of H-1 of residue **A** at δ_H 5.25 instead had a wider peak which indicates a medium-sized $^{3}J_{H1,H2}$ of between 2 and 7 Hz, characteristic for α-*gluco*/*galacto*-configured monosaccharides. The $^{3}J_{HH}$ of the anomeric proton in residue **D** was measured to 7.8 Hz which is typical for β-*gluco*/*galacto*-configured sugars.

A F_2-coupled ^{1}H,^{13}C-HSQC spectrum was also acquired for the PS-OH (Figure 19) and the measured $^{1}J_{C1,H1}$ are 179, 159, 160 and 163 Hz respectively for positions **A1**, **B1**, **C1** and **D1**, and thus residue **A** has an α-anomeric configuration whereas **B, C** and **D** have β-configured anomeric positions.

In addition, an array of ^{1}H,^{1}H-TOCSY NMR experiments was recorded with mixing times of 20-120 ms (Figure 19). H1 of residue **A** shows a transfer of magnetization until H4. H1 in both residues **B** and **C** only shows correlations to H2 which indicates that they are of *manno*-configuration. A correlation from H1 of residue **D** to H5 of residue **D** was observed, indicating *gluco*-configuration.

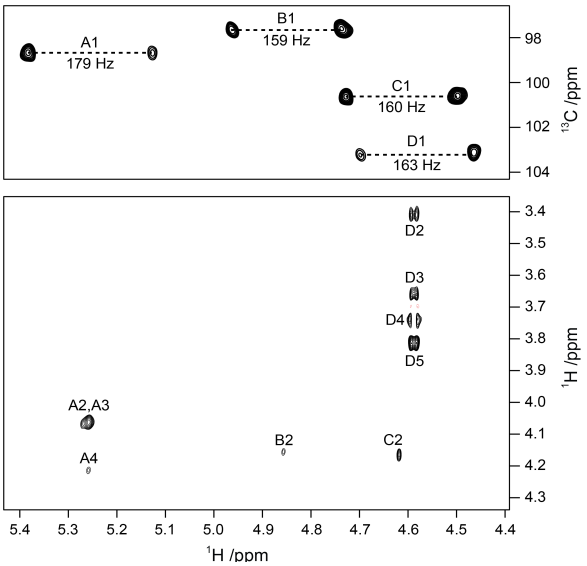


Figure 19. Selected regions of a F_2-coupled ^{1}H,^{13}C-HSQC (top) and ^{1}H, ^{1}H-TOCSY with a 120 ms mixing time (bottom) NMR spectra of the O93 O-antigen PS-OH.

4.2.4 Structure elucidation by CASPER

Data obtained from the above-mentioned analysis of the F_2-coupled ^{1}H,^{13}C-HSQC (measured $^{1}J_{CH}$), ^{1}H,^{13}C-HSQC and ^{1}H,^{1}H-TOCSY (peak-picked resonances), and ^{1}H NMR (measured $^{3}J_{HH}$) experiments along with insights from the bioinformatics (contains a β-D-Glc*p*A and a D-GlcNAc linked together) and sugar analysis (D-Man and D-GlcNAc present) were used as inputs for the 'Determine Glycan Structure' section in CASPER.

The highest-ranked structure consists of a linear tetrasaccharide repeating unit (Figure 20) which belongs to one of the most common topologies for *E. coli* O-antigens with an abundancy of 20 % (Section 1.8, Figure 6). It contains the anticipated β-D-Glc*p*A-(1→3)-D-GlcNAc linkage that was suggested by the bioinformatical data and also contains a β-D-Man-(1→3)-β-D-Man sequence which has not been observed in any previous *E. coli* O-antigen structure before.

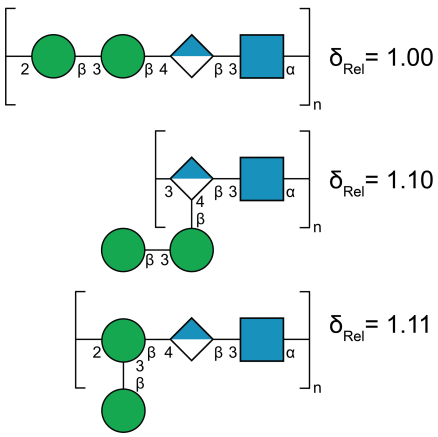


Figure 20. The three highest-ranked structures of the *E. coli* O93 O-antigen PS-OH suggested by CASPER.

The assignment suggested by CASPER of the experimental chemical shifts was then validated with a manual assignment and is summarized in Table 4. The sequence was confirmed by analysis of a recorded ^{1}H,^{1}H-NOESY NMR experiment (Figure 3, Paper III) which shows correlations between H1 of residue **A** to H2 of residue **B**, from H1 of residue **B** to H3 of residue **C**, from H1 in residue **C** to H4 of residue **D** and from H1 of residue **D** to H3 of residue **A** resulting in a sequence [-**B**-**C**-**D**-**A**-]$_n$.

Table 4. ^{1}H and ^{13}C NMR chemical shifts of the *O*-deacetylated polysaccharide from the *E. coli* O93 O-antigen structure.

	Sugar residue	^{1}H and ^{13}C chemical shifts							
		1	2	3	4	5	6	2Ac	2CO
A	→3)-α-D-GlcNAc-(1→	5.25 (M)[b]	4.06	4.06	3.61	4.21	~3.84	2.04	
		98.78 [179][c]	53.70	81.15	69.31	72.04	61.24	22.87	n.d[a]
B	→2)-β-D-Man-(1→	4.85 (S)	4.16	3.80	3.72	3.42	3.79, 3.94		
		97.55 [159]	77.33	74.37	68.04	77.54	61.78		
C	→3)-β-D-Man-(1→	4.62 (S)	4.17	3.85	3.62	3.39	3.75, 3.92		
		100.66 [161]	68.50	80.11	65.80	77.21	61.82		
D	→4))-β-D-GlcA-(1→	4.59 (L)	3.40	3.65	3.74	3.81			
		103.27 [162]	73.41	74.78	81.26	76.45	n.d.[a]		

[a] n.d. = not determined.
[b] $^{3}J_{H1,H2}$ value used as input for 'Determine Glycan Structure' in CASPER: S < 2 Hz, M = 2 – 7 Hz and L > 7 Hz.
[c] $^{1}J_{C1,H1}$ values given in Hertz in square brackets and obtained from a coupled ^{1}H,^{13}C-HSQC NMR spectrum.

4.2.5 Assignment of the *O*-acetyl groups

The study then proceeded with assignment of the observed *O*-acetyl groups in the PS material. This was deciphered with the 2D NMR experiments, ^{1}H,^{1}H-TOCSY, ^{1}H,^{13}C-HSQC and ^{1}H, ^{13}C-BS-CT-HMBC (Figure 21).

In the ^{1}H,^{13}C-HSQC NMR spectrum, two resonances at δ_H/δ_C 4.37/63.88 and δ_H/δ_C 4.46/63.88 were observed that are characteristic of a 6-*O*-acetylated hydroxymethyl group. Furthermore, two significantly downfield displaced resonances at δ_H/δ_C 5.53/69.97 and δ_H/δ_C 5.55/69.87 corresponding to *O*-substituted sugar ring protons were found.

These four resonances were subsequently assigned to adjacent protons in the spin-system of residue **C,** using ^{1}H,^{1}H-TOCSY NMR spectra with shorter mixing times (30-90 ms), with δ_H/δ_C 4.37/63.88 correlating to H5 of residue **C** at δ_H 3.63 and resonances at δ_H/δ_C 5.53/69.97 5.55/69.87 showing correlations to δ_H H1 of residue **C** at 4.81 and 4.82, respectively.

Interestingly, in a ^{1}H,^{1}H-TOCSY NMR spectrum with a longer mixing time of 120 ms correlations *from* δ_H 5.55 (H2) *to* δ_H 4.37/4.46 (H6) were also observed (Figure 21 top) which suggests the presence of a 2,6-*O*-acetylated minor population of residue **C** in the PS in addition to a 2-*O*-acetylated major population. The mono-*O*-acetylated population was denoted with a single prime (**C'**) and the di-*O*-acetylated population with double prime (**C''**).

A ^{1}H, ^{13}C-BS-CT-HMBC then further shows correlations *from* their respective carbonyl carbons from the *O*-acetyl groups (Figure 21 bottom) *to* the protons of H2 and H6 of residue **C** with correlations *from* δ_C 174.67 *to* δ_H 4.37/4.46 for **C6''** and *from* δ_C 174.04/173.95 *to* δ_H 5.53/5.55, respectively for **C2'** and **C2''**.

The degree of *O*-acetylation in the PS was subsequently determined, by integration of the anomeric resonance in a ^{1}H,^{13}C-HSQC, to be ~½ 2,6-*O*-acetylated, ~¼ 2-*O*-acetylated while ~¼ is without any *O*-acetyl group.

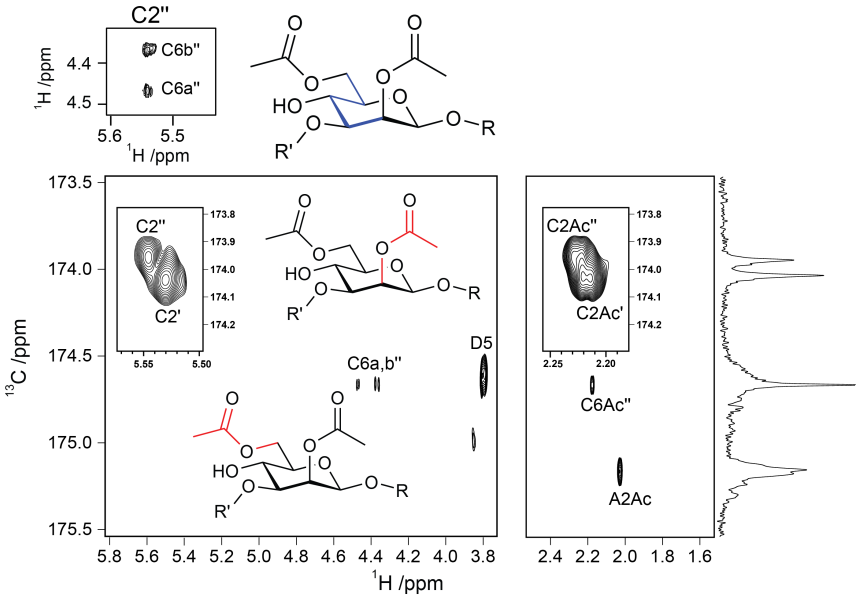


Figure 21. Selected regions of ^{1}H, ^{1}H-TOCSY (top) and ^{1}H, ^{13}C-BS-CT-HMBC (bottom) spectra summarizing the *O*-acetyl assignment of residue **C'/C''** with inserted chemical drawings highlighting observed correlations in blue or red.

When validating the assignment of resonances from the *E. coli* O93 PS with the 'Calculate Chemical Shift' feature in CASPER, the majority of the resonances are in good agreement (Figure 22). One significant outlier is observed in δ_H of H2 for the *O*-acetylated populations (**B'**/**B''**) (Figure 6 bottom, Paper III). This observation is unexpected, since the only change from the PS-OH material is the 2,6-*O*-acetylation of the adjacent sugar-residue (**C'**/**C''**). In addition, large chemical shift displacements caused by *O*-acetylations are usually *downfield* in δ_H on adjacent positions of the same sugar-residue (Section 1.11) whereas, in this case, it is an *upfield* displacement in δ_H of 0.33 when comparing the experimental chemical shifts of H2 for residue **B** vs **B'** (Table 1 Paper III). We therefore speculate that this chemical shift change originates from conformational aspects involving the 2-*O*-acetyl group which CASPER cannot predict.

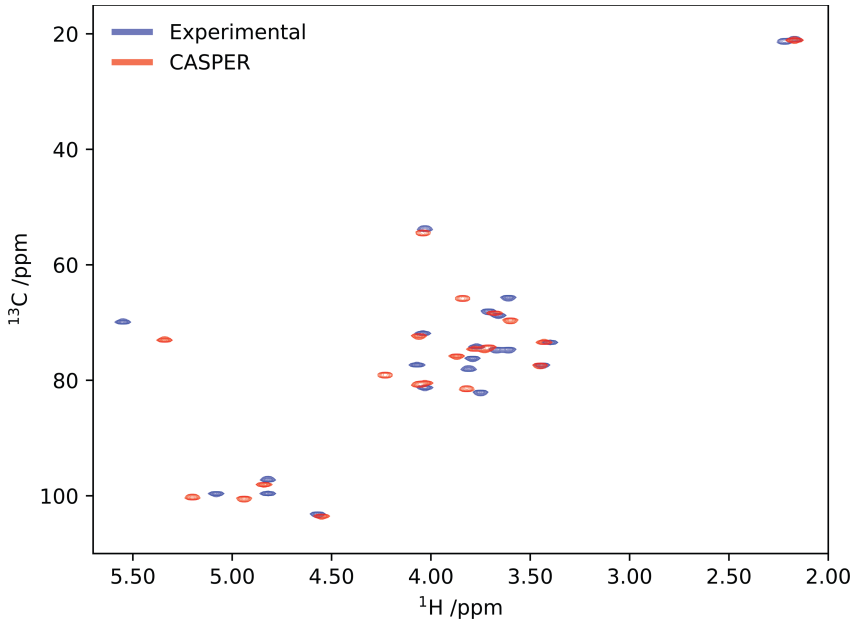


Figure 22. Overlay of experimental (blue) and CASPER-predicted (red) chemical shifts from the PS-OH of *E. coli* O93 O-antigen plotted using with NMRglue.[90]

4.2.6 Conformational insights

To further investigate this, we recorded 1D ^{1}H,^{1}H-NOESY with 300 ms and 600 ms mixing times with selective irradiation set on the H-Me of **C2Ac** in order to map the conformational space. The 1D ^{1}H,^{1}H-NOESY revealed correlations *from* H-Me of **C2Ac** *to* protons of **A1**, **A5**, **A6** and also **B2**. A constructed 3D model with torsion angles of $\varphi, \psi \approx -35°, -45°$ for the α-D-Glc*p*NAc-(1→2)-β-D-Man*p* linkage and $\varphi, \psi \approx 25°, 60°$ for the β-D-Man*p*-(1→3)-β-D-Man*p*2Ac fitted well with the observed NOEs (Figure 23). This aligned the shielding anisotropy of the close-by carbonyl group of **C2Ac** in close proximity to the proton of **B2** which may explain the significant upfield displacement observed.

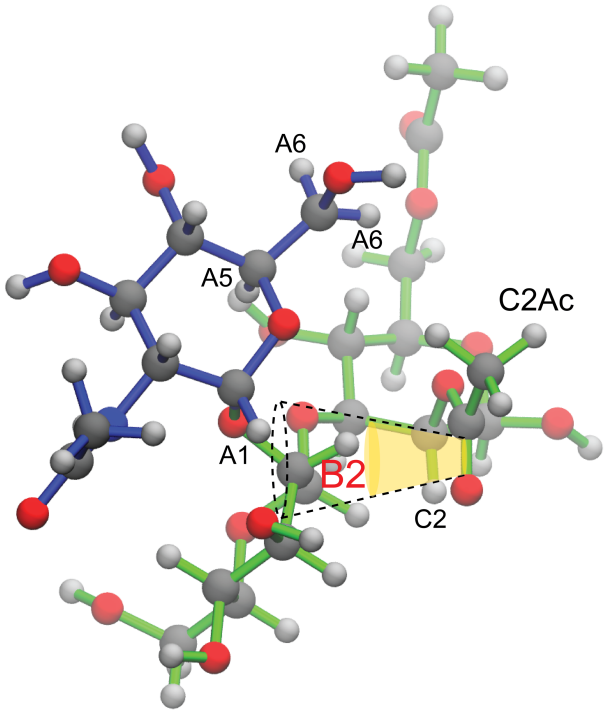


Figure 23. 3D model of trisaccharide α-D-Glc*p*NAc-(1→2)-β-D-Man*p*-(1→3)-β-D-Man*p*2Ac from the *E. coli* O93 O-antigen structure. The protons that show correlations from a 1D ^{1}H,^{1}H-NOESY 300 ms irradiating the **C2Ac** signal. Part of the shielding anisotropy of the carbonyl group is highlighted in yellow.

A capsular polysaccharide from *E. coli* O20:K84:H25 has the structure:→4)-α-D-Gal*p*A-(1→3)-α-D-Man*p*-(1→4)-β-D-Glc*p*-(1→3)-α-D-Glc*p*NAc-(1→2)-β-D-Man*p*-(1→3)-β-D-Man*p*NAc-(1→, with a similar trisaccharide fragment (underlined) also having its H2 of the central β-D-Man*p* resonating upfield at 3.91 ppm, most likely caused by the close-by *N*-acetyl group of the adjacent β-D-Man*p*NAc.[91]

4.2.7 Structural comparisons

No other *E.* coli O-antigen structure in the *E.coli* structure database, ECODAB[92] were found to have the β-D-Man*p*-(1→3)-β-D-Man*p* structural element. A similar structural search was conducted in the Carbohydrate Structure Database (CSDB) which contains polysaccharide structures originating from a wider range of biological material (incl. prokaryotes, plants and fungi).[93] There, a structural match was found in the O-antigen structure of *Cellulophaga baltica*[94] even having a similar degree of *O*-acetylation (70%) in position 2 of the 3-linked mannose residue (**C2'** in this study) but with no *O*-acetylation reported in position 6.

A comparison of ^{1}H and ^{13}C NMR chemical shifts of the two assigned O-deacetylated polysaccharides were in good agreement with each other with only one resonance having a δ_C deviation of more than 2 ppm which was the linkage position (**C3** in this study). This is most likely due to the NMR spectra of the polysaccharides being recorded at different temperatures (70°C vs 30°C for O93 PS/*C. baltica* PS). The essentially identical O-antigen structures in the bacteria may be explained by a possible horizontal gene transfer event involving the related gene cluster encoding for the O-antigen synthesis from one species to the other.[29,95]

4.3 Conclusion and future outlook

The *E. coli* O93 O-antigen structure was elucidated (Figure 24) with combined efforts from bioinformatics, sugar analysis and NMR studies. CASPER predicted the correct structure of the *O*-deacetylated polysaccharide (PS-OH) and assisted in validating the assignment of the *O*-acetylated polysaccharide. The combination of an unusual β-D-Man*p*-(1→3)-β-D-Man*p*-linkage and an adjacent *O*-acetyl substitution resulted in an unexpected upfield chemical shift displacement for H2 in residue **B** which was further investigated using 1D ^{1}H,^{1}H-NOESY NMR.

The elucidated PS was similar to the O-antigen structure of *Cellulophaga baltica* and therefore an additional study comparing their gene clusters would be of interest in the future.

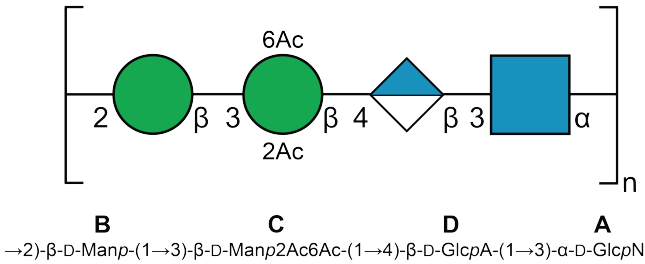


Figure 24. The elucidated structure of *E. coli* O93 O-antigen depicted with a SNFG representation.

5 Developments of the NMR-database CASPER (Paper IV)

5.1 Introduction

CASPER is a useful tool, as mentioned in Section 1.12, that has been frequently used by researchers in the field of carbohydrate chemistry to assist in their structure elucidation studies. It has been used as a fast way to obtain ^{1}H and ^{13}C chemical shifts of monosaccharides to compare with the user's own experimental data[96,97] or to validate the user's assignment of exopolysaccharides by simulating the suggested structure's ^{1}H and ^{13}C chemical shifts.[98,99]

Continuous updates of CASPER with additional mono-, di- and trisaccharide residues will further improve the performance of CASPER in predicting the chemical shifts of more complex oligo- and polysaccharides.

This chapter will summarize the most recent update of CASPER and demonstrate its usefulness.

5.2 Results and discussion

5.2.1 NMR assignments

The study contains ^{1}H and ^{13}C chemical shift data from 58 different sugars consisting of 30 monosaccharides, 17 disaccharides, 10 trisaccharides and one tetrasaccharide. One of the assigned monosaccharides is *N*-acetylmuramic acid (MurNAc), which is most known for being present in the peptidoglycan layer of Gram-positive bacteria. Similar to the assignment of most sugars in this study, MurNAc was assigned using a combination of the typical 2D NMR experiments ($^{1}H,^{13}C$-HSQC, $^{1}H,^{13}C$-HMBC, $^{1}H,^{13}C$-H2BC, $^{1}H,^{13}C$-BS-CT-HMBC, and $^{1}H,^{1}H$-TOCSY). Exact chemical shifts of the ring proton region at δ_H 3.65-3.95 were obtained by iterative fitting with PERCH (Figure 25). PERCH also calculated the anomeric population in MurNAc to 66% in α- and 34% in β-configuration.

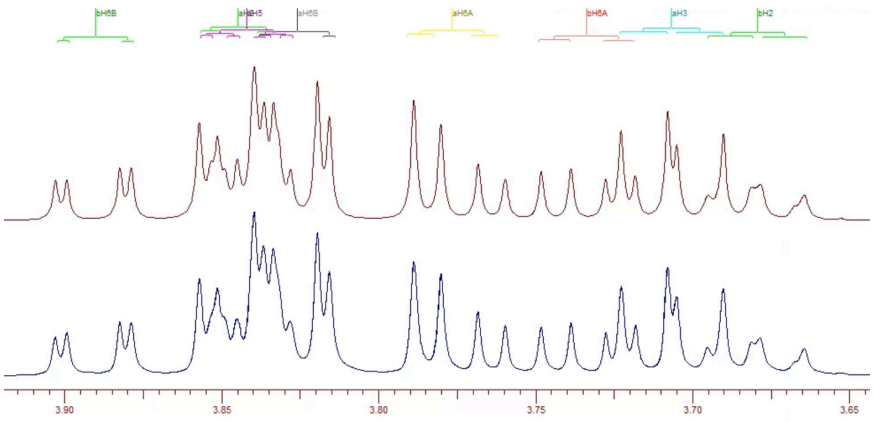


Figure 25. Selected region of a recorded ^{1}H-NMR *N*-acetylmuramic acid (MurNAc) dissolved in D_2O (blue) and a PERCH-simulated spectrum.[68]

5.2.2 Extraction of new correction sets

New correction sets were manually constructed to further improve the chemical shift prediction of common sialic acid-containing oligosaccharides. One important correction set was derived from the ganglioside oligosaccharide, GM2 (Table 6, Paper IV), with the following structure: β-D-Gal*p*NAc-(1→4)[α-Neu5Ac(1→3)]-β-D-Gal*p*(1→4)-D-Glc*p*.

The correction set for the β-D-Gal*p*NAc-(1→4)[α-Neu5Ac(1→3)]-β-D-Gal*p* trisaccharide was extracted according to the procedure thoroughly described by Ståhle & Widmalm.[100] The trisaccharide correction set suggests a significant change in δ_C of −2.3 in position 3 of the αNeu5Ac residue resulting in a upfield adjustment.

Table 5. ^{1}H and ^{13}C chemical shift corrections displayed in ppm for the αNeu5Ac residue from the GM2 trisaccharide correction set.

Sugar residue		1	2	3	4	5
αNeu5Ac-(1→	^{1}H			0.11 −0.08	0.08	−0.03
	^{13}C	0.1	1.5	−2.3	0.3	−0.1

5.2.3 Case study: predicting the structure of α/β-Gd1a

The newly implemented correction set was tested by trying to predict the structure of α/β-Gd1a.

Acquired 2D NMR experiments (^{1}H,^{13}C-HSQC, ^{1}H,^{13}C-H2BC and ^{1}H,^{13}C-HMBC) were peak-picked and the resulting list of correlations were added as input in the 'Determine Glycan Structure' feature in CASPER (Figure 26)

along with the sugar-residues and all linkage positions set as wildcards and the box ‘Free anomer in reducing end’ activated.

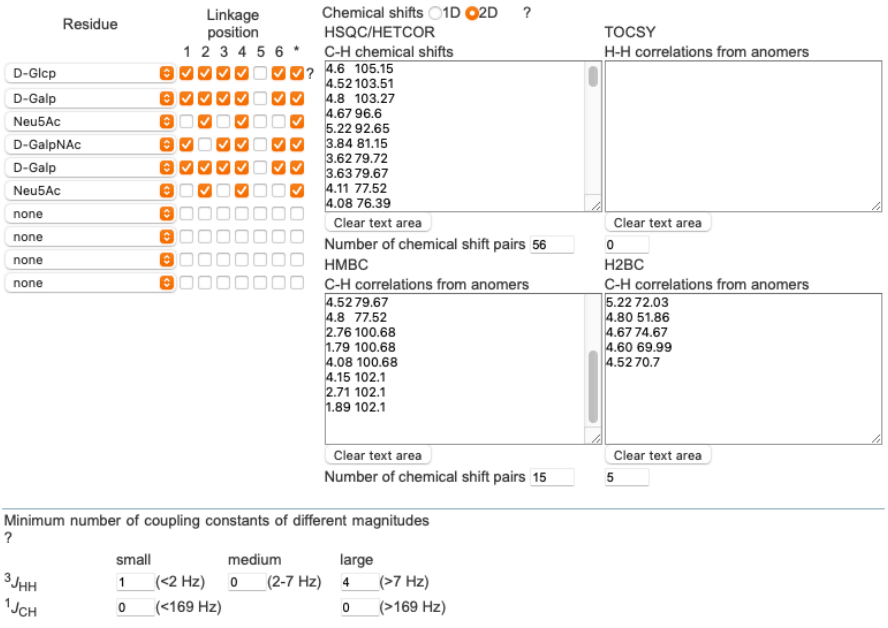


Figure 26. Screenshot of the ‘Determine glycan structure’ webpage in CASPER with input-data added for the α/β-GD1a oligosaccharide from 2D NMR experiments (^{1}H,^{13}C-HSQC, ^{1}H,^{13}C-HMBC and ^{1}H,^{13}C-H2BC) and with the ‘Free anomer in reducing end’ activated.

The prediction results from CASPER propose the correct structures ranked as #1 and #2 while #3 and #4 have chemical shifts that deviate considerably more with larger ‘normalized relative rankings’ at 1.43 and 1.44 respectively (Figure 27).

Proposed structures

1. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-Gal-(1→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→4)-β-D-Glc , δ_{Rel}=1.00
2. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-Gal-(1→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→4)-α-D-Glc , δ_{Rel}=1.17
3. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→3)-β-D-Gal-(1→4)-β-D-Glc , δ_{Rel}=1.43
4. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→3)-β-D-Gal-(1→4)-α-D-Glc , δ_{Rel}=1.44
5. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-Gal-(1→4)]β-D-GalNAc-(1→4)-β-D-Gal-(1→4)-β-D-Glc , δ_{Rel}=1.43
6. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-Gal-(1→4)]β-D-GalNAc-(1→4)-β-D-Gal-(1→4)-α-D-Glc , δ_{Rel}=1.48
7. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-GalNAc-(1→4)-β-D-Gal-(1→4)]β-D-Gal-(1→4)-β-D-Glc , δ_{Rel}=1.44
8. aNeu5Ac(2→3)[aNeu5Ac(2→3)-β-D-GalNAc-(1→4)-β-D-Gal-(1→4)]β-D-Gal-(1→4)-α-D-Glc , δ_{Rel}=1.49
9. aNeu5Ac(2→3)[aNeu5Ac(2→3)[β-D-Gal-(1→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→4)]β-D-Glc , δ_{Rel}=1.46
10. aNeu5Ac(2→3)[aNeu5Ac(2→3)[β-D-Gal-(1→3)-β-D-GalNAc-(1→4)]β-D-Gal-(1→4)]α-D-Glc , δ_{Rel}=1.49

Figure 27. Proposed structures suggested by CASPER from the GD1a structure prediction with β-GD1a and α-GD1a ranked as #1 and #2, respectively.

A comparison of predicted and experimental ^{1}H and ^{13}C chemical shift data for β-GD1a was also done and was in good agreement (Figure 28). This shows that CASPER is capable of predicting complex oligosaccharides with Neu5Ac residues.

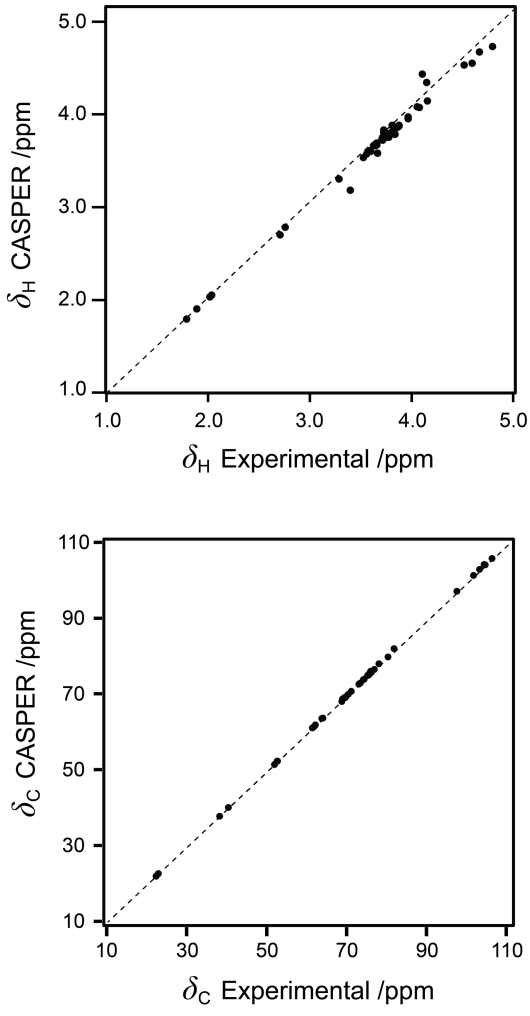
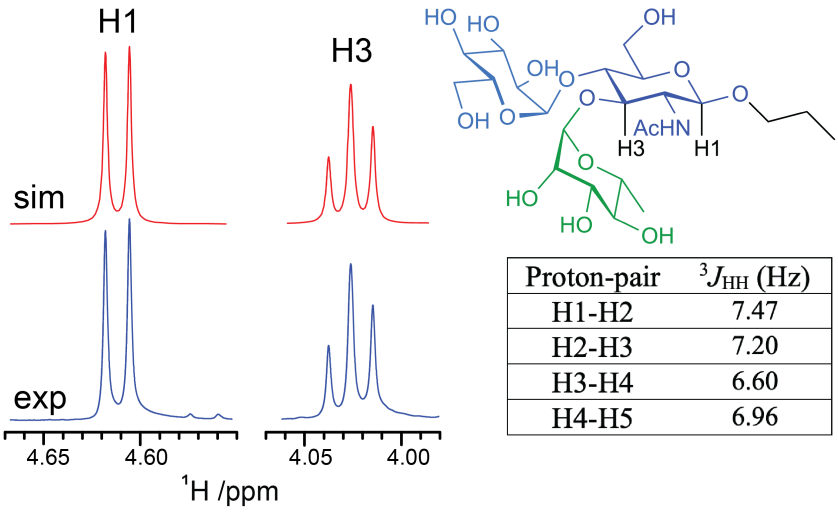


Figure 28. Comparison of ^{1}H and ^{13}C NMR chemical shifts from experimental data and predicted by CASPER of β-GD1a.

5.2.4 Small $^{3}J_{HH}$ observed in a 3,4-disubstituted β-D-Glc*p*NAc

Another observation in this study was the $^{3}J_{HH}$ of the proton pairs in the central β-D-Glc*p*NAc residue of the trisaccharide α-L-Rha*p*-(1→3)-[α-D-Glc*p*-(1→4)]-β-D-Glc*p*NAc-OPr (Table 5 Paper IV) were noticeably smaller than expected *e.g.*, $^{3}J_{H3,H4}$ of 6.60 Hz (Figure 29) compared to a monosaccharide counterpart *e.g.*, β-D-Glc*p*NAc-O*p*NP, that had a $^{3}J_{H3,H4}$ of 9.0 Hz. This indicates that the proton pairs of the branched β-D-Glc*p*NAc may adopt conformations where they are deviating from the antiperiplanar arrangement, for instance due to ring puckering.



Proton-pair	$^{3}J_{HH}$ (Hz)
H1-H2	7.47
H2-H3	7.20
H3-H4	6.60
H4-H5	6.96

Figure 29. Selected regions of a PERCH-simulated (top-left) and an experimental (bottom-left) ^{1}H NMR spectrum of the vicinal branched trisaccharide: α-L-Rhap-(1→3)[α-D-Glcp-(1→4)]-β-D-GlcNAcOPr (top-right) and related $^{3}J_{HH}$ of the GlcNAc residue (bottom-right).

5.3 Conclusion and future outlook

This chapter summarizes parts of the work included in the new update of the computer program CASPER where 58 compounds have had their ^{1}H and ^{13}C NMR resonances assigned and added to the database. A correction set related to the ganglioside oligosaccharide, GM2, was presented. Also, an example showcased the 'Determine Glycan Structure' feature in which CASPER successfully predicted the structure of the α/β-GD1a oligomer.

In the future, the toolbox of CASPER could be further expanded in order to successfully handle 'hyper-branched' structures containing multiple vicinal disubstituted sugar residues found in nature as well as linkages to more non-sugar residues used in for instance automated solid-phase carbohydrate synthesis.[101,102] Another feature that could further improve CASPER prediction would be to implement $^{1}H,^{1}H$-NOESY NMR data as input for CASPER for determining the sugar ring configuration of unknown monosaccharides using key intra-residue correlations and, more importantly, for sequence determination from observed inter-residue correlations.

6 Conformational Analysis of *Shigella flexneri* O-antigens (Paper V)

6.1 Introduction

Shigella flexneri is a species within the genus *Shigella* and is known to be closely related to *E. coli* [103] and has previously been referred to as "*E. coli* in disguise".[104] However, there are certain characteristic differences *e.g.* that strains of *E. coli* have flagella, thus being motile, while strains of *Shigella* lack this feature.[105]

Out of the 1.6 million deaths associated with diarrheal infections, 15% of them were caused by strains from *Shigella* with low- and middle-income countries being most affected.[106]

As in other Gram-negative bacteria, the antibiotic resistance is rapidly increasing in *Shigella* with strains of *Shigella sonnei* (highly prevalent in South America) that have obtained resistance to many antibiotics.[107] Therefore, the development of vaccines targeting these strains of bacteria could help improve both the quality of life among the affected communities while also decreasing the rate of antibiotic resistance and decreasing the risk for potential epidemics/outbreaks.

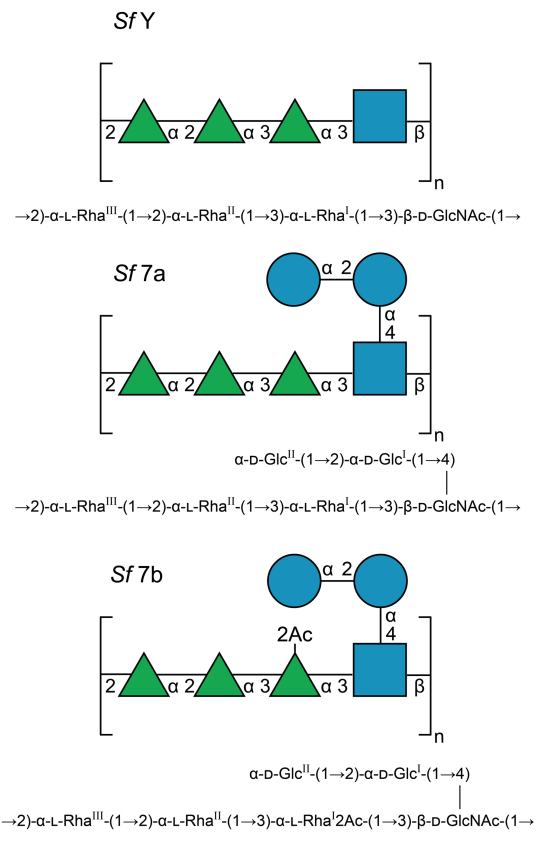


Figure 30. The O-antigen structures of *Shigella flexneri* Y, 7a and 7b depicted with SNFG representation.

The O-antigen structure of *Shigella flexneri* can be categorized into serotypes named X, Y and 1−7 with various subtypes. A majority of the serotypes have a common tetrasaccharide backbone: →2)-α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I}-(1→3)-β-D-Glc*p*NAc-(1→ which is also the O-antigen structure of *Shigella flexneri* Y (*Sf* Y) with its structure elucidated in 1978 (Figure 30 top).[108] The rhamnose residues are

commonly numbered starting from the one linking to the GlcNAc residue according to the biological repeating unit,[109] and are in this study denoted using roman letters. For clarification, this study also annotates the two glucose residues as Glc^{I} and Glc^{II}.

This study involves *Shigella flexneri* 7a[110] (also known as 1c) and 7b[111] (*Sf* 7a and *Sf* 7b) which, in addition, have a branched disaccharide of two α-D-Glc*p* (*Sf* 7a, Figure 30 middle), and further, an *O*-acetyl substitution in position 2 of the α-L-Rha*p*^{I}-residues (*Sf* 7b, Figure 30 bottom).

More in-depth conformational analysis of the O-antigen structures in *Shigella flexneri* is needed to get a better understanding of which structures to be included in a potential multi-valent glycoconjugate vaccine[106] since cross-protection has been observed between the certain serotypes.[112,113]

The cross-protection is caused by the serotypes sharing key structural elements, O-antigen epitopes, that induce a similar immune response. The epitopes are summarized in a review by Knirel *et al.*[114] with *Sf* 7a having the O-antigen epitope, IC, formed by the α-D-Glc*p*^{II}-(1→2)-α-D-Glc*p*^{I}-(1→4)-β-D-Glc*p*NAc structural elements, and *Sf* 7b also having the O-antigen epitope 6 constituted by the 2-*O*-acetylation of α-L-Rha*p*^{I}.

Recent conformational studies on the O-antigen structure of *Shigella flexneri* Y*,* 2, 3, 5 and 6[69,115–117] have been investigated by MD-simulations and a study combining both MD-simulations and NMR data has been performed for *Sf* serotypes: X, Y and 1−5.[118]

In this chapter, the goal is to investigate the conformational preferences of *Sf* 7a and 7b by using a combination of NMR spectroscopy and MD-simulations and to compare the results with *Sf* Y.

6.2 Result and discussion

6.2.1 Temperature studies

When adjusting the temperature to prevent spectral overlap between the anomeric protons and the HDO peak, we discovered that the ^{1}H chemical shifts of some anomeric resonances were changing considerably.

A temperature study was conducted since variations in chemical shift may indicate signs of conformational flexibility. ^{1}H NMR spectra between 5-70 °C were acquired for both *Sf* 7a and *Sf* Y to investigate any differences. For *Sf* 7a, a significant chemical shift change was observed for the anomeric proton of Glc^{I} (Figure 31) with −2.3 ppb/°C thus displaced upfield when the temperature increases. The adjacent anomeric proton of the GlcNAc residue instead moves in the opposite direction, having a +1.3 ppb/°C displacement downfield. H1 of Rha^{III} has an upfield change of +1.1 ppb/°C whereas the remaining anomeric protons has smaller temperature dependencies (Rha^{I} −0.3 ppb/°C, Rha^{II} <+0.2 ppb/°C and Glc^{II} −0.9 ppb/°C).

In contrast, the anomeric protons of *Sf* Y showed only minor changes in chemical shift over temperature with H1 of Rha^{I} at +0.7 ppb/°C having the largest temperature dependence while H1 of GlcNAc, Rha^{II} and Rha^{III} had changes of +0.3, +0.2 and −0.4 ppb/°C, respectively.

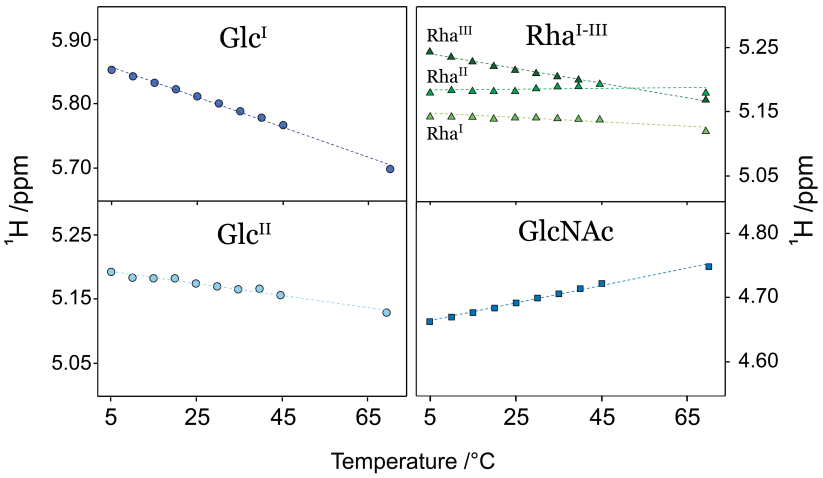


Figure 31. Change in ^{1}H NMR chemical shift as a function of temperature (ppb/°C) for the anomeric protons of the S. flexneri 7a and Y O-antigens.

6.2.2 $^{3}J_{HH}$-couplings of the GlcNAc residue

Sf 7a has the same vicinal disubstituted trisaccharide structure within the repeating unit as the trisaccharide compound previously presented (Section 5.2.4). We therefore investigated whether *Sf* 7a also shows the smaller than usual $^{3}J_{HH}$.

1D ^{1}H,^{1}H-TOCSY NMR spectra were recorded in order to extract each of the $^{3}J_{HH}$ couplings in the spin-system of the GlcNAc residue in *Sf* 7a. This was deemed possible since its anomeric proton was free from spectral overlap and could therefore be selectively irradiated. Similar to the temperature study, this was performed for both *Sf* 7a and *Sf* Y. The $^{3}J_{H1,H2}$ for both anomeric protons were measured in their respective ^{1}H NMR spectra while the $^{3}J_{H2,H3}$ was obtained by subtracting the $^{3}J_{H1,H2}$ from the total width of the *dd* of H2. The same procedure was performed for the remaining $^{3}J_{HH}$ (Table 6 and Figure 32).

Table 6. The $^{3}J_{HH}$ coupling constants (Hz) for selected protons in the GlcNAc residue of the O-antigens from *S. flexneri* 7a and Y.

Proton pair	*Sf* 7a	*Sf* Y
	$^{3}J_{HH}$	
H1-H2	8.1	8.5
H2-H3	9.0	10.0
H3-H4	6.2	9.0
H4-H5	9.0	9.6

Interestingly, the $^{3}J_{H3,H4}$ coupling of *Sf* 7a is unusually small (6.2 Hz), following the previously described observation from compound **53** in Paper IV which has a corresponding $^{3}J_{H3,H4}$ coupling of 6.6 Hz while the one in *Sf* Y is larger (9.0 Hz), similar to a monosaccharide GlcNAc residue. The considerably lower $^{3}J_{H3,H4}$ coupling in *Sf* 7a indicates that H3 and H4 are less frequently occupying an anti-periplanar orientation (*e.g.*, when in a common $^{4}C_{1}$ ring pucker) than in *Sf* Y.

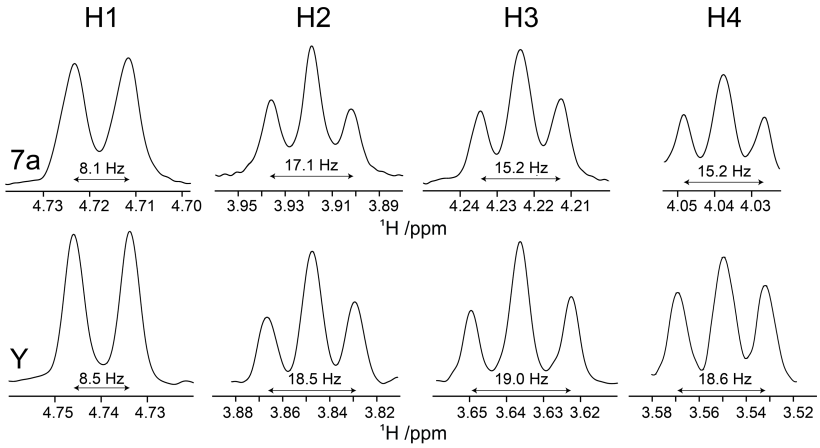


Figure 32. Selected regions from ^{1}H NMR and 1D ^{1}H,^{1}H-TOCSY experiments displaying the $^{3}J_{HH}$ and the total width of involved $^{3}J_{HH}$ in the *dd* of H2, H3 and H4 in *Shigella flexneri* 7a and Y.

6.2.3 ^{1}H,^{1}H-NOESY buildup curves *Sf* 7a

The study proceeded by recording ^{1}H,^{1}H-NOESY NMR spectra with zero-quantum coherence suppression filter at 500 MHz and with the different mixing times 30, 40, 50, 60, 70, 80, 90 and 110 ms (Figure 33).

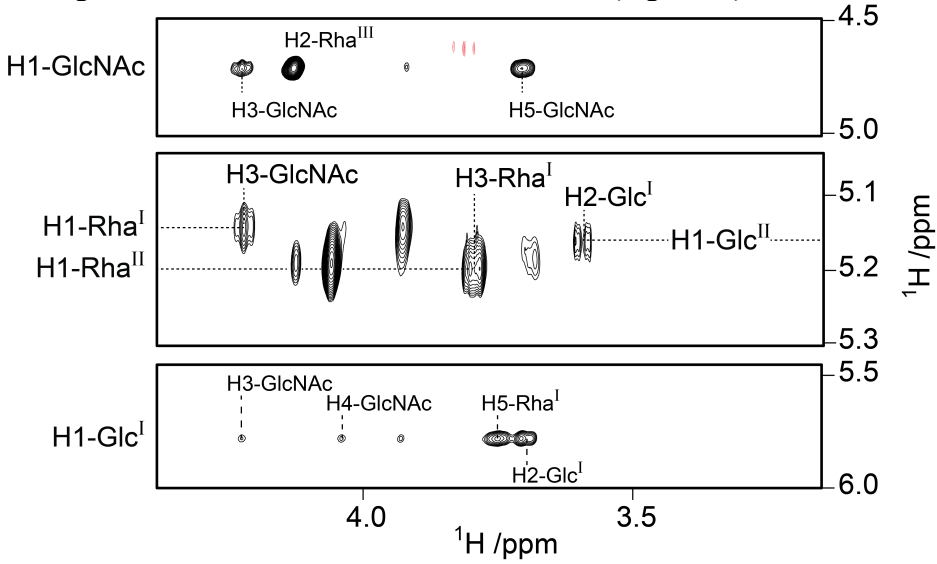


Figure 33. Panels along the F_2 dimension of the anomeric region of *Shigella flexneri* 7a from a ^{1}H, ^{1}H-NOESY at 500 MHz recorded with a mixing time of 110 ms.

Build-up curves were then generated according to the PANIC approach in order to obtain the cross-relaxation rates (σ). Cross-relaxation rates for five of the six glycosidic linkages were possible to obtain for *Sf* 7a (Figure 34). In addition, other inter-residue correlations were found *e.g.*, from H1-Glc^{I} to H3-GlcNAc and a methyl-methine cross-peak from Me-Rha^{I} to H1-Glc^{I}.

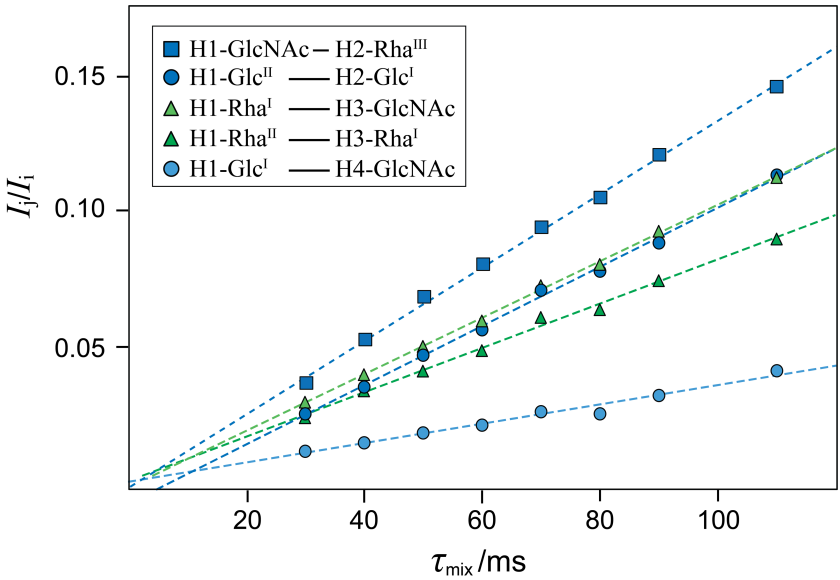


Figure 34. Selected ^{1}H,^{1}H-NOESY buildup curves from the O-antigen polysaccharide of *Shigella flexneri* 7a with I_j/I_i vs. τ_{mix} plotted, following the PANIC approach. The cross-relaxation rates, ($-\sigma_{ij} = I_j/I_i$) are determined from the slope of the fitted data. Markers are depicted using SNFG-representation.

6.2.4 MD-simulations and comparisons with NMR-data

MD-simulation of 1 μs on a six repeating unit O-antigen structure in solution was performed in a ~90 Å^3 water box using a modified CHARMM36 additive force field[119] for both *Sf* 7a and 7b with the 4 innermost repeating units only considered for the data collection of distance, torsion angles and pucker states. Data collected in a previous study of *Sf* Y will also be used as a comparison.[117]

The initial analysis of the two polysaccharides show that the GlcNAc residue occupies, in addition to the expected $^{4}C_1$, the ring-pucker conformations $^{1}C_4$, $^{3}S_1$, $B_{1,4}$ and $^{5}S_1$ (Figure S1, Paper V) whereas in *Sf* Y this is not observed to a considerable degree.

The $^{1}C_4$ conformation was ruled out as unlikely due to the magnitude of the $^{3}J_{HH}$ couplings since a β-D-Glc*p*NAc in a $^{1}C_4$ conformation would have had multiple protons in equatorial-equatorial orientation thus leading to smaller $^{3}J_{HH}$ couplings (1-3 Hz). The remaining observed pucker conformations, $^{3}S_1$,

$B_{1,4}$ and $^{5}S_{1}$, were deemed likely since they would still maintain anti-periplanar-like orientations for the majority of the proton-pairs.

The distances were then calculated following the ISPA using the equation, $r_{ij} = r_{ref} \cdot (\sigma_{ref}/\sigma_{ij})^{1/6}$ in which the r_{ref} was calculated with $r_{MD} = \langle r^{-6} \rangle^{-1/6}$. Cross-relaxation-rates used as references (σ_{ref}) were H1-Glc^{II}–H2-Glc^{II} for methine-methine correlations and H5-Rha^{II}–Me-Rha^{II} for methyl-methine cross-peaks. The distances from the MD-simulation, r_{MD}, were initially calculated from the frames where the GlcNAc residues adopts a $^{4}C_{1}$ conformation.

Overall, the distances correlated well when comparing the NOE-derived distances with the ones from the MD-simulation (Table 7) except for the distance H1-Glc^{I}–H4-GlcNAc which was considerably longer (r_{MD}=2.94 Å vs r_{expt}=2.68Å). We hypothesize that this is due to only including the $^{4}C_{1}$ conformations and we, therefore, tested whether an inclusion of the frames related to the $^{3}S_{1}$, $B_{1,4}$ and $^{5}S_{1}$ conformations would fit better with the experimental data both suggested from ^{1}H,^{1}H-NOESY and the above-mentioned derived $^{3}J_{HH}$ couplings. By calculating the root-mean-square deviation of all measured inter-residue and trans-glycosidic distances at different ratios of $^{4}C_{1}$:($^{3}S_{1}$, $B_{1,4}$, and $^{5}S_{1}$) we observed a minimum at a 82.5:17.5 ratio of $^{4}C_{1}$ vs. $^{3}S_{1}$, $B_{1,4}$, and $^{5}S_{1}$ conformations mostly due to a change of the H1-Glc^{I}–H4-GlcNAc distance changing from 2.94 Å to 2.70 Å. The remaining distances only observed minor changes (r_{MDadj} column, Table 7).

Table 7. Proton-proton cross-relaxation rates (σ, s^{-1}) from the O-antigen polysaccharide of *Shigella flexneri* 7a derived from ^{1}H,^{1}H-NOESY NMR buildup curves analyzed by the PANIC approach experimentally derived distances (r_{expt} in Å) and effective proton-proton distances (r_{MD} and r_{MDadj} in Å). The effective distances r_{MD} refer to only the $^{4}C_{1}$ conformation of the GlcNAc residue while r_{MDadj} distances were calculated based on an 82.5:17.5 ratio of $^{4}C_{1}$ vs. $^{3}S_{1}$, $B_{1,4}$, and $^{5}S_{1}$ distance conformations.

Proton pair		σ_{expt}	r_{expt} [a]	r_{MD} [b]	r_{MDadj} [b]
Trans-glycosidic					
H1-Rha^{II}	H3-Rha^{I}	–0.83	2.33	2.32	2.32
H1-Rha^{I}	H3-GlcNAc	–1.06	2.24	2.22	2.24
H2-Rha^{III}	H1-GlcNAc	–1.33	2.16	2.29	2.29
H1-Glc^{I}	H4-GlcNAc	–0.36	2.68	2.94	2.70
H1-Glc^{II}	H2-Glc^{I}	–1.11	2.22	2.43	2.43
Inter-residue					
H1-Glc^{I}	H5-Rha^{I}	–0.90	2.30	2.30	2.36
H1-Glc^{I}	H3-GlcNAc	–0.43	2.61	2.65	2.66
H1-Glc^{I}	H1-Glc^{II}	–0.57	2.48	2.22	2.23
Methyl-methine					
Me-Rha^{I}	H1-Glc^{II}	–0.26	3.22	3.15	3.16
Me-Rha^{I}	H1-Glc^{I}	–0.23	3.28	3.24	3.24
Reference					
H1-Glc^{II}	H2-Glc^{II}	–0.59	2.47	2.47	2.47
H5-Rha^{II}	Me-Rha^{II}	–0.31	2.60	2.60	2.60

[a] Calculated with $r_{ij} = r_{ref} \cdot (\sigma_{ref}/\sigma_{ij})^{1/6}$ with reference proton distances derived from 1 µs, of MD simulation of the six-repeating unit polysaccharide in solution.

[b] Effective distance derived from MD simulations was calculated according to $r_{MD} = \langle r^{-6} \rangle^{-1/6}$.

The NMR-study for *Sf* 7b was less extensive due to significant spectral overlap since the degree of *O*-acetylation is non-stoichiometric and does not displace many of the other resonances significantly from the *O*-deacetylated population also present in the PS sample. The autopeak of H1-Glc^{I} in *Sf* 7b was slightly displaced downfield and its autopeak, along with the observed cross-peak, could be utilized to obtain the cross-relaxation rates for H1-Glc^{I}–H3 GlcNAc/H4-GlcNAc/H5-Rha^{I} (Table 3, Paper V). The calculated distances agreed with the corresponding ones in *Sf* 7a.

The glycosidic torsion angles (ϕ, ψ) of the polysaccharides of *Sf* Y, 7a and 7b are visualized with heatmaps (Figure 6 Paper V). When comparing *Sf* 7a and 7b, no large differences in ϕ, ψ torsion angles were observed in any of the glycosidic linkages. More expected, the conformational space is slightly different between *Sf* Y and *Sf* 7a/b for the H1-Rha^{I}–H3-GlcNAc glycosidic linkage, where the ϕ, ψ torsion angles in *Sf* 7a/b adopted a more restricted conformational space, especially in ψ, only ranging from ≈ 0° to +30° while the corresponding ψ torsion angle of *Sf* Y ranged between ≈ ±60°. The H1-Glc^{I}–H4-GlcNAc glycosidic linkage mainly adopts torsion angles of ϕ, ψ ≈ –60°, –60° which agrees with the exo-anomeric effect of an α-D-Glc*p* residue.

In addition, snapshots of the 4 innermost repeating units of *Sf* 7a and 7b were taken to visualize the epitopes present in the O-antigen structures (Figure 35).

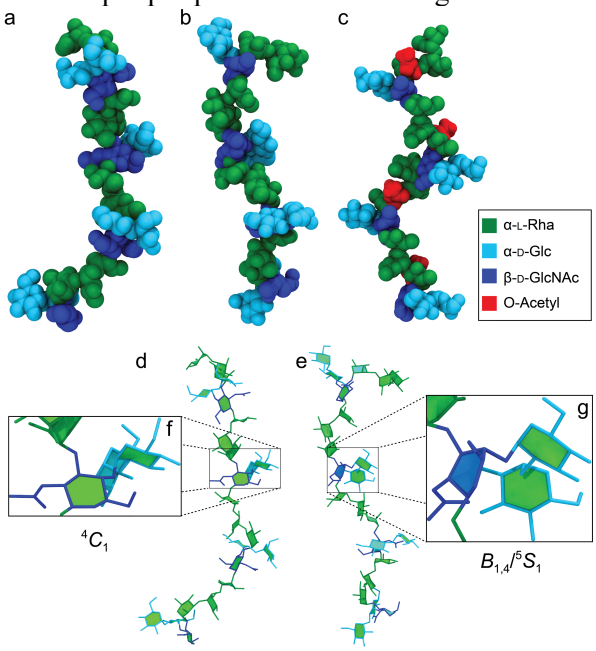


Figure 35. Representative snapshots from the MD-simulation of the 4 innermost repeating units of *Shigella flexneri* O-antigen structures visualized with VMD[120] using either vdW representation or a licorice/paperchain representation for *Shigella flexneri* 7a (a-b, d-g) and 7b (c).

6.3 Conclusion and future aspects

A combination of NMR-studies and MD-simulations were used to investigate various conformational aspects of the *Sf* 7a and 7b O-antigen polysaccharides.

NOE-derived distances related to five out of six glycosidic linkages were successfully obtained from an array of ^{1}H,^{1}H-NOESY using the PANIC approach and the approach devised by Dixon *et al.*[56] was utilized to obtain H1-Glc^{I}-related distances for *Sf* 7b. Interestingly, the NOE-derived distances for H1-Glc^{I}–H4-GlcNAc and H1-Glc^{I}–H3-GlcNAc were equidistant in both *Sf* 7a and 7b.

The combined efforts of $^{3}J_{HH}$ coupling measurements and MD-simulations suggest that the 3,4-vicinal disubstituted β-D-Glc*p*NAc-residue, to a certain degree, adopts a $^{3}S_{1}$, $B_{1,4}$, or $^{5}S_{1}$ ring pucker conformation in addition to the common $^{4}C_{1}$ conformation.

Further investigating the conspicuous α-L-Rha*p*-(1→3)-[α-D-Glc*p*-(1→4)]-β-D-Glc*p*NAc trisaccharide would be of interest in the future since the MD-simulations suggests the uncommon $^{1}C_{4}$ conformation for the GlcNAc residue while the obtained NMR data shows no sign of that conformation. Further development of the CHARMM additive force field to better handle these highly branched sugars is needed.

7 Concluding remarks

During the course of this thesis, three previously unknown *E. coli* O-antigen structures have been investigated. Their elucidated repeating units all belong to the most common topology in *E. coli* with a 4 sugar-residue backbone.

CASPER was an essential tool in these three studies and was utilized in different ways:

- As a chemical shift database, by rapidly obtaining the monosaccharide NMR chemical shifts to thereafter compare with corresponding ones of the *E. coli* O188 O-antigen to confirm its linkage positions.
- As an automated structure elucidation tool, by using the 'Determine Glycan Feature' – adding inputs from multiple 1D and 2D NMR experiments, sugar analysis and biological rules to obtain repeating unit structures of *E. coli* O125ac and O93 (PSOH).
- To validate the manual assignment of the *O*-acetylated polysaccharide in *E. coli* O93 using the 'Chemical Shift Calculation' section which in turn highlighted a conspicuous signal (proton of **B2'/ B2''**) that was shifted upfield in the experimental data due to a shielding anisotropy.

Further updates of CASPER will extend its usage towards automated solid phase-based synthetic carbohydrates by adding sugar-residues with common linkers that are used in that process.

Due to the rapid increase in antibiotic resistance WHO predicts that by 2050 infections from these bacterial strains will surpass the number of deaths caused by cancer.[121] Therefore, vaccines targeting Gram-negative bacteria will be of increasing demand and importance to avoid the risk of future pandemics and a mutated strain belonging to one of the above-elucidated serogroups could therefore be a future vaccine candidate with an isolated or synthesized polysaccharide conjugated with bacterial toxin as one possible solution.

For similar reasons the conformational studies of the *Shigella flexneri* 7a and 7b O-antigen structure with NMR and MD simulations and comparisons to the *Shigella flexneri* Y counterpart provide a deeper insight into how the 3,4 vicinal disubstituted β-D-Glc*p*NAc is affected by the branched glucose disaccharide and highlight key epitopes of the polysaccharide structure.

8 Acknowledgment

I would like to thank the following: / jag skulle vilja tacka:

Min handledare *Göran Widmalm*, som antog mig som doktorand. Jag har lärt mig så ofantligt mycket under denna tid. Tack för intressanta och givande diskussioner/anekdoter genom åren, både vetenskapligt och om allt möjligt annat.

My co-supervisor *Prof. Belén Matute* and *Prof. Kalman Szabo* for proof-reading my thesis and providing me with valuable comments.

Collaborators in all of the projects, in particular, *Prof. Daniel Daley, Dr. Klas Udekwu, Prof. Michelle Kuttel, Dr. Claudio Muheim* and *Zaheer Timol.*

All past and present members of the Göran Widmalm group, *Dr. Alessandro Ruda, Kevin Dorst, Dr. Federico Riu, Dr. Geeta Karki, Dr. Mateusz Pallach, Zaheer Timol, Dr. Elena Stepanova, Dr. Jonas Ståhle, Dr. Hani Mobarak* and *Prof. Carolina Fontana*, and the students *David Bitar, Li Wang, Louise Marizy* and *Nevén Tiyé*.

The Lennart Kenne foundation for a travel grant to EUROCARB in Leiden, 2019.

TA-staff members *Kristina* and *Matthew*, for all the help with the NMR instruments and, *Carin, Jonas*, *Martin*, and *Ola* for helping me out with the GC-instruments, lab-related problems, and computer server-related issues.

To all past and present colleagues at the department for the nice working atmosphere, *Aitor Bermejo López, Alba Carretero Cerdán, Dr. Andrey Shatskiy, Beatriz Meana, Bram Peters, Dr. Cristiana Margarita, Dr. Davide Rigo, Daria Lebedeva, Dr. Davide Di Francesco, Dr. Elena Stepanova, Dr. Elena Subbotina, Dr. Gabriella Kervefors, Dr. Sybrand Jonker, Dr. Denise Meyer, Emanuelle Silvi, Ester Di Tommaso, Flavia Ferrara, Jesper Schwarz, Dr. Kuntawit Witthayolankowit, Dr. Laura Castoldi, Luca Massaro, Marie Deliaval, Dr. Marvin Lübcke, Dr. Matteo Costantini, Dr. Michael Meyer, Dr. Rajdip Chowdhury, Dr. Haibo Wu, Dr. Jianping Yang, Pedro Tortajada Palmero, Dr. Stefano Parisotto, Dr. Stephanie Kohlkepp, Dr. Supaporn Sawadjoon, Suthawan Muangmeesri, Dr Thanya Rukkijakan, Dr. Sybrand Jonker, Tautvydas Kireilis, Victor Garciá Vázguez, Dr. Wangchuk Rabten* and many more, as well as *Brando Adranno*, *Francesco Nosenzo* from outside this department.

Dr. Alexander, Emil, Dr. Hannes, Johan, Liv, Max, Niklas, Oliver, Sebastian, Simon och *Teo* för support och trevligt sällskap, både på distans och nere i Skåne under somrarna.

Magdalena för allt!

min familj, *Catrin, Fredrik* och *Svante* för er oändliga kärlek och stöd genom livets alla svängar.

9 References

1. Brigham, C. Biopolymers: Biodegradable Alternatives to Traditional Plastics. in *Green Chemistry: An Inclusive Approach* (2017).
2. Xu, C. & Ng, D. T. W. Glycosylation-directed quality control of protein folding. *Nat. Rev. Mol. Cell Biol.* **16**, 742–752 (2015).
3. Varki, A. Biological roles of glycans. *Glycobiology* (2017).
4. Bonomelli, C., Doores, K. J., Dunlop, D. C., Thaney, V., Dwek, R. A., Burton, D. R., Crispin, M. & Scanlan, C. N. The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One* **6**, 1–7 (2011).
5. Binley, J. M., Ban, Y.-E. A., Crooks, E. T., Eggink, D., Osawa, K., Schief, W. R. & Sanders, R. W. Role of Complex Carbohydrates in Human Immunodeficiency Virus Type 1 Infection and Resistance to Antibody Neutralization. *J. Virol.* **84**, 5637–5655 (2010).
6. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T. & Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).
7. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).
8. Liu, B., Qian, C., Wu, P., Li, X., Liu, Y., Mu, H., Huang, M., Zhang, Y., Jia, T., Wang, Y., Wang, L., Zhang, X., Huang, D., Yang, B., Feng, L. & Wang, L. Attachment of Enterohemorrhagic Escherichia coli to Host Cells Reduces O Antigen Chain Length at the Infection Site That Promotes Infection. *MBio* **12**, (2021).
9. Gray, C. J., Migas, L. G., Barran, P. E., Pagel, K., Seeberger, P. H., Eyers, C. E., Boons, G. J., Pohl, N. L. B., Compagnon, I., Widmalm, G. & Flitsch, S. L. Advancing Solutions to the Carbohydrate Sequencing Challenge. *Journal of the American Chemical Society* 14463–14479 (2019).
10. Fadda, E. & Woods, R. J. Molecular simulations of carbohydrates and protein–carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* **15**, 596–609 (2010).
11. Widmalm, G. A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr. Res.* **378**, 123–132 (2013).
12. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F. G., Lütteke, T. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* (2015).
13. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T.,

Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F. G., Lütteke, T. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324 (2015).
14. Pendrill, R., Jonsson, K. H. M. & Widmalm, G. Glycan synthesis, structure, and dynamics: A selection. *Pure Appl. Chem.* **85**, 1759–1770 (2013).
15. Sehnal, D. & Grant, O. C. Rapidly Display Glycan Symbols in 3D Structures: 3D-SNFG in LiteMol. *J. Proteome Res.* (2019).
16. Mayes, H. B., Broadbelt, L. J. & Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* **136**, 1008–1022 (2014).
17. Lundborg, M., Fontana, C. & Widmalm, G. Automatic structure determination of regular polysaccharides based solely on NMR spectroscopy. *Biomacromolecules* **12**, 3851–3855 (2011).
18. Leontein, K. & Lönngren, J. Determination of the absolute configuration of sugars by gas-liquid chromatography of their acetylated 2-octyl glycosides. *Methods Carbohydr. Chem.* **9**, 87–89 (1993).
19. Nikaido, H. Diffusion of inhibitors across the cell wall. *New Compr. Biochem.* (1994).
20. Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications. *Bacteriol. Rev.* (1973).
21. Rogers, H. J. Peptidoglycans (mucopeptides): structure, function, and variations. *Ann. N. Y. Acad. Sci.* (1974).
22. Osborn, M. J., Gander, J. E. & Parisi, E. Mechanism of Assembly of the Outer Membrane of Salmonella typhimurium. *J. Biol. Chem.* (1972).
23. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* **148**, 1258–1270 (2012).
24. Charbonneau, M. R., Blanton, L. V., Digiulio, D. B., Relman, D. A., Lebrilla, C. B., Mills, D. A. & Gordon, J. I. A microbial perspective of human developmental biology. *Nature* **535**, 48–55 (2016).
25. Kaper, J. B., Nataro, J. P. & Mobley, H. L. T. Pathogenic Escherichia coli. *Nat. Rev. Microbiol.* **2**, 123–140 (2004).
26. Servin, A. L. Pathogenesis of Afa/Dr diffusely adhering Escherichia coli. *Clin. Microbiol. Rev.* **18**, 264–292 (2005).
27. Jensen, B. H., Olsen, K. E. P., Struve, C., Krogfelt, K. A. & Petersen, A. M. Epidemiology and clinical manifestations of enteroaggregative escherichia coli. *Clin. Microbiol. Rev.* **27**, 614–630 (2014).
28. Nguyen, Y. & Sperandio, V. Enterohemorrhagic E. coli (EHEC) pathogenesis. *Front. Cell. Infect. Microbiol.* **2**, (2012).
29. Pasqua, M., Michelacci, V., Di Martino, M. L., Tozzoli, R., Grossi, M.,

Colonna, B., Morabito, S. & Prosseda, G. The intriguing evolutionary journey of enteroinvasive E. coli (EIEC) toward pathogenicity. *Front. Microbiol.* **8**, (2017).
30. Deborah Chen, H. & Frankel, G. Enteropathogenic Escherichia coli: Unravelling pathogenesis. *FEMS Microbiol. Rev.* **29**, 83–98 (2005).
31. Wang, L. & Reeves, P. R. Organization of escherichia coli O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.* **66**, 3545–3551 (1998).
32. Heiman, K. E., Mody, R. K., Johnson, S. D., Griffin, P. M. & Hannah Gould, L. Escherichia coli O157 Outbreaks in the United States, 2003–2012. *Emerg. Infect. Dis.* **8**, 1293–1301 (2015).
33. Lee, Y. M., Wallace, M., Gengatharan, J. M., Furst, A. J., Bode, L., Metallo, C. M. & Ayres, J. S. Microbiota control of maternal behavior regulates early postnatal growth of offspring. *Sci. Adv.* **7**, (2021).
34. Kolmos, H. J. Panum's studies on 'putrid poison' 1856. An early description of endotoxin. *Dan. Med. Bull.* **53**, 450–452 (2006).
35. Knirel, Y. A. & Valvano, M. A. *Bacterial lipopolisaccharides. Structure, Chemical synthesis, biogenesis and interaction with Host cells. Springerverlag* (2011).
36. Westphal, O. & Jann, K. Extraction with Phenol-Water and Further Applications of the Procedure. *Methods Carbohydr. Chem.* **5**, 83–91 (1965).
37. Nikaido, H. Biosynthesis of Cell Wall Lipopolysaccharide in Gram-Negative Enteric Bacteria. in *Advances in Enzymology and Related Areas of Molecular Biology* (1968).
38. König, H., Claus, H. & Varma, A. *Prokaryotic cell wall compounds: Structure and biochemistry*. *Prokaryotic Cell Wall Compounds: Structure and Biochemistry* (2010).
39. Silipo, A. & Molinaro, A. Lipid A Structure. in *Bacterial Lipopolysaccharides* (2011).
40. Amor, K., Heinrichs, D. E., Frirdich, E., Ziebell, K., Johnson, R. P. & Whitfield, C. Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. *Infect. Immun.* **68**, 1116–1124 (2000).
41. Silipo, A. & Molinaro, A. The diversity of the core oligosaccharide in lipopolysaccharides. *Subcell. Biochem.* (2010).
42. Ståhle, J. & Widmalm, G. Lipopolysaccharides of gram-negative bacteria: Biosynthesis and structural aspects. *Trends Glycosci. Glycotechnol.* **31**, E159–E171 (2019).
43. Caroff, M. & Karibian, D. Structure of bacterial lipopolysaccharides. *Carbohydrate Research* (2003).
44. Ståhle, J. & Widmalm, G. Lipopolysaccharides of Gram-negative bacteria: Biosynthesis and structural aspects. *Trends Glycosci. Glycotechnol.* **31**, E159–E171 (2019).
45. Sherman, D. J., Xie, R., Taylor, R. J., George, A. H., Okuda, S., Foster,

P. J., Needleman, D. J. & Kahne, D. Lipopolysaccharide is transported to the cell surface by a membrane-to- membrane protein bridge. *Science,* 798–801 (2018).
46. Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N. & Kahne, D. Lipopolysaccharide transport and assembly at the outer membrane: The PEZ model. *Nat. Rev. Microbiol.* **14**, 337–345 (2016).
47. Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB 2 FGC complex. *Nature* **567**, 486–490 (2019).
48. Owens, T. W., Taylor, R. J., Pahil, K. S., Bertani, B. R., Ruiz, N., Kruse, A. C. & Kahne, D. Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550–553 (2019).
49. Liu, B., Furevi, A., Perepelov, A. V, Guo, X., Cao, H., Wang, Q., Reeves, P. R., Knirel, Y. A., Wang, L. & Widmalm, G. Structure and genetics of Escherichia coli O antigens. *FEMS Microbiol. Rev.* **44**, 655–683 (2020).
50. Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A. & Woods, R. J. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29**, 620–624 (2019).
51. Bloch, F. Bloch, Felix: The Principle of Nuclear Induction Nobel Lecture, December 11, 1952 © Le Prix Nobel, 1952 . in *Encyclopedia of Magnetic Resonance* (2007).
52. Purcell, E. M. Purcell, Edward M.: Research in Nuclear Magnetism Nobel Lecture December, 11, 1952 © Les Prix Nobel, 1952 . in *Encyclopedia of Magnetic Resonance* (2007).
53. Ernst, R. R. Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture). *Angewandte Chemie International Edition in English* (1992).
54. Wüthrich, K. NMR studies of structure and function of biological macromolecules (Nobel Lecture). in *Angewandte Chemie - International Edition* (2003).
55. Macura, S., Farmer, B. T. & Brown, L. R. An improved method for the determination of cross-relaxation rates from NOE data. *J. Magn. Reson.* **70**, 493–499 (1986).
56. Dixon, A. M., Widmalm, G. & Bull, T. E. Modified GOESY in the Analysis of Disaccharide Conformation. *J. Magn. Reson.* **147**, 266–272 (2000).
57. Widmalm, G. General NMR Spectroscopy of Carbohydrates and Conformational Analysis in Solution. in *Comprehensive Glycoscience: From Chemistry to Systems Biology* (2007).
58. Gheysen, K., Mihai, C., Conrath, K. & Martins, J. C. Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chem. - A Eur. J.* **14**, 8869–8878 (2008).
59. Yu, B., Van Ingen, H., Vivekanandan, S., Rademacher, C., Norris, S.

E. & Freedberg, D. I. More accurate 1JCH coupling measurement in the presence of 3JHH strong coupling in natural abundance. *J. Magn. Reson.* **215**, 10–22 (2011).
60. Nyberg, N. T., Duus, J. & Sørensen, O. W. Heteronuclear two-bond correlation: Suppressing heteronuclear three-bond or higher NMR correlations while enhancing two-bond correlations even for vanishing 2JCH. *J. Am. Chem. Soc.* **127**, 6154–6155 (2005).
61. Kupče, Ē., Yong, J. R. J., Widmalm, G. & Claridge, T. D. W. Parallel NMR Supersequences: Ten Spectra in a Single Measurement. *JACS Au* **1**, 1892–1897 (2021).
62. Jansson, P.-E., Kenne, L. & Widmalm, G. CASPER-a computerised approach to structure determination of polysaccharides using information from n.m.r. spectroscopy and simple chemical analyses. *Carbohydr. Res.* **168**, 67–77 (1987).
63. Hermansson, K., Jansson, P. E., Kenne, L., Widmalm, G. & Lindh, F. A 1H and 13C NMR study of oligosaccharides from human milk. Application of the computer program CASPER. *Carbohydr. Res.* **235**, 69–81 (1992).
64. Jansson, P.-E., Stenutz, R. & Widmalm, G. Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program casper. *Carbohydr. Res.* **341**, 1003–1010 (2006).
65. Loß, A., Stenutz, R., Schwarzer, E. & von der Lieth, C. W. GlyNest and CASPER: Two independent approaches to estimate 1H and 13C NMR shifts of glycans available through a common web-interface. *Nucleic Acids Res.* **34**, 733–737 (2006).
66. Kapaev, R. R. & Toukach, P. V. Simulation of 2D NMR Spectra of Carbohydrates Using GODESS Software. *J. Chem. Inf. Model.* **56**, 1100–1104 (2016).
67. Kapaev, R. R. & Toukach, P. V. GRASS: Semi-Automated NMR-based structure elucidation of saccharides. *Bioinformatics* **34**, 957–963 (2018).
68. Laatikainen, R., Niemitz, M., Weber, U., Sundeun, J., Hassinen, T. & Vepsäläinen, J. General strategies for total-lineshape-type spectral analysis of NMR spectra using integral-transform iterator. *J. Magn. Reson. - Ser. A* **120**, 1–10 (1996).
69. Kuttel, M. M., Ståhle, J. & Widmalm, G. CarbBuilder: Software for building molecular models of complex oligo- and polysaccharide structures. *J. Comput. Chem.* **37**, 2098–2105 (2016).
70. De Castro, C., Parrilli, M., Holst, O., Molinaro, A., Castro, C. De, Parrilli, M., Holst, O. & Molinaro, A. Microbe-associated molecular patterns in innate immunity. Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. *Methods Enzymol.* **480**, 89–115 (2010).
71. Knirel, Y. A., Kocharova, N. A., Hynes, S. O., Widmalm, G.,

Andersen, L. P., Jansson, P. E. & Moran, A. P. Structural studies on lipopolysaccharides of serologically non-typable strains of helicobacter pylori, af1 and 007, expressing lewis antigenic determinants. *Eur. J. Biochem.* **233**, 123–131 (1999).
72. Kunin, C. M. Separation, Characterization, and Biological Significance of a Common. *J. Exp. Med.* **118**, 565–586 (1963).
73. Valtonen, M. V., Larinkari, U. M., Plosila, M., Valtonen, V. V. & Mäkelä, P. H. Effect of enterobacterial common antigen on mouse virulence of Salmonella typhimurium. *Infect. Immun.* **13**, 1601–1605 (1976).
74. Ramos-Morales, F., Prieto, A. I., Beuzón, C. R., Holden, D. W. & Casadesús, J. Role for Salmonella enterica Enterobacterial Common Antigen in Bile Resistance and Virulence. *J. Bacteriol.* **185**, 5328–5332 (2003).
75. Erbel, P. J. A., Seidel, R., Macintosh, S. E., Gentile, L. N., Amor, J. C., Kahn, R. A., Prestegard, J. H., McIntosh, L. P. & Gardner, K. H. Cyclic enterobacterial common antigen: Potential contaminant of bacterially expressed protein preparations. *J. Biomol. NMR* **29**, 199–204 (2004).
76. Oikawa, M., Adachi, S. & Kusumoto, S. 2JC,Hindex: A nondestructive NMR method for differentiation of aldohexopyranosyl residues. *Org. Lett.* **7**, 661–664 (2005).
77. Larsson, E. A., Urbina, F., Yang, Z., Weintraub, A. & Widmalm, G. Structural and immunochemical relationship between the O-antigenic polysaccharides from the enteroaggregative Escherichia coli strain 396/C-1 and Escherichia coli O126. *Carbohydr. Res.* **339**, 1491–1496 (2004).
78. Jansson, P.-E., Kenne, L. & Widmalm, G. Computer-assisted structural analysis of regular polysaccharides. *Pure Appl. Chem.* **61**, 1181–1192 (1989).
79. Liu, B., Senchenkova, S. N., Feng, L., Perepelov, A. V, Xu, T., Shevelev, S. D., Zhu, Y., Shashkov, A. S., Zou, M., Knirel, Y. A. & Wang, L. Structural and molecular characterization of *Shigella boydii* type 16 O antigen. *Gene* **380**, 46–53 (2006).
80. Freitas Do Valle, G. R., Tardelli Gomes, T. A., Irino, K. & Trabulsi, L. R. The traditional enteropathogenic Escherichia coli (EPEC) serogroup O125 comprises serotypes which are mainly associated with the category of enteroaggregative E. coli. *FEMS Microbiol. Lett.* **152**, 95–100 (1997).
81. Elias, W. P., Barros, S. F., Moreira, C. G. & Serogroups, E. O. Enteroaggregative Escherichia coli Strains among Classical Enteropathogenic Escherichia coli O Serogroups Letters to the Editor Enteroaggregative Escherichia coli Strains among Classical Enteropathogenic. *J. Clin. Microbiol.* **40**, 3540–3541 (2002).
82. Kjellberg, A., Urbina, F., Weintraub, A. & Widmalm, G. Structural

Analysis of the O-Antigenic Polysaccharide from the Enteropathogenic Escherichia coli O125. *Eur. J. Biochem.* **239**, 532–538 (1996).
83. SSI Diagnostica. SSI Diagnostica list of E. coli strains. https://ssidiagnostica.com/international/solutions/bacterial-strains/e-coli-strains/ (2022).
84. Jann, K., Schmidt, G., Wallenfels, B. & Oulbert, E. F.-M. Isolation and Characterization of Escherichia coli bacteriophage Ω8 specific for E. coli strains belonging to sero-group O8. *Microbiology* **67**, 289–297 (1971).
85. Wallenfels, B. & Jann, K. The action of bacteriophage Ω8 on two strains of Escherichia coli O8. *J. Gen. Microbiol.* **81**, 131–144 (1974).
86. Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., Hayashi, T. & Thomson, N. R. A complete view of the genetic diversity of the Escherichia coli O-antigen biosynthesis gene cluster. *DNA Res.* **22**, 101–107 (2015).
87. Rush, J. S., Alaimo, C., Robbiani, R., Wacker, M. & Waechter, C. J. A novel epimerase that converts GlcNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in Escherichia coli O1570. *J. Biol. Chem.* **285**, 1671–1680 (2010).
88. Wang, L., Huskic, S., Cisterne, A., Rothemund, D. & Reeves, P. R. The O-antigen gene cluster of Escherichia coli O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. *J. Bacteriol.* **184**, 2620–2625 (2002).
89. Fontana, C., Lundborg, M., Weintraub, A. & Widmalm, G. Rapid structural elucidation of polysaccharides employing predicted functions of glycosyltransferases and NMR data: Application to the O-antigen of Escherichia coli O59. *Glycobiology* **24**, 450–457 (2014).
90. Helmus, J. J. & Jaroniec, C. P. Nmrglue: An open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* **55**, 355–367 (2013).
91. Whittaker, D. V, Parolis, L. A. S. & Parolis, H. Structural elucidation of the capsular polysaccharide produced by Escherichia coli O20 : K84 : H26. *Carbohydr. Res.* **262**, 323–334 (1994).
92. Rojas-Macias, M. A., Ståhle, J., Lütteke, T. & Widmalm, G. Development of the ECODAB into a relational database for Escherichia coli O-antigens and other bacterial polysaccharides. *Glycobiology* **25**, 341–347 (2015).
93. Toukach, P. V & Egorova, K. S. Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic Acids Res.* **44**, D1229–D1236 (2015).
94. Tomshich, S. V., Komandrova, N. A., Widmalm, G., Nedashkovskaya, O. I., Shashkov, A. S. & Perepelov, A. V. Structure of acidic O-specific polysaccharide from the marine bacterium Cellulophaga baltica. *Russ. J. Bioorganic Chem.* **33**, 83–87 (2007).

95. Arnold, B. J., Huang, I.-T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **20**, 206–218 (2022).
96. Hamedi, A., Pasdaran, A. & Pasdaran, A. A trisaccharide phenylethanoid glycoside from Scrophularia flava Grau with potential anti-type 2 diabetic properties by inhibiting α-glucosidase enzyme and decreasing oxidative stress. *Bioorg. Chem.* **99**, 103776 (2020).
97. Botto, E., Reina, L., Moyna, G., Menéndez, P. & Rodríguez, P. Insights into the hydrolysis of Eucalyptus dunnii bark by xylanolytic extracts of Pseudozyma sp. *Biomass Convers. Biorefinery* (2020).
98. Kumar, A., Mukhia, S. & Kumar, R. Production, characterisation, and application of exopolysaccharide extracted from a glacier bacterium Mucilaginibacter sp. ERMR7:07. *Process Biochem.* **113**, 27–36 (2022).
99. Wang, J., Salem, D. R. & Sani, R. K. Two new exopolysaccharides from a thermophilic bacterium Geobacillus sp. WSUCF1: Characterization and bioactivities. *N. Biotechnol.* **61**, 29–39 (2021).
100. Ståhle, J. & Widmalm, G. CHAPTER 15. NMR Chemical Shift Predictions and Structural Elucidation of Oligo- and Polysaccharides by the Computer Program CASPER. in *NMR in Glycoscience and Glycotechnology* 335–352 (The Royal Society of Chemistry, 2017).
101. Seeberger, P. H. & Haase, W. C. Solid-phase oligosaccharide synthesis and combinatorial carbohydrate libraries. *Chem. Rev.* **100**, 4349–4393 (2000).
102. Seeberger, P. H. Glycan arrays and other tools produced by automated glycan assembly. *Perspect. Sci.* **11**, 11–17 (2017).
103. Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. & Wang, L. Structure and genetics of Shigella O antigens. *FEMS Microbiol. Rev.* **32**, 627–653 (2008).
104. Lan, R. & Reeves, P. R. Escherichia coli in disguise: Molecular origins of Shigella. *Microbes Infect.* **4**, 1125–1132 (2002).
105. Zuo, G., Xu, Z. & Hao, B. Shigella Strains Are Not Clones of Escherichia coli but Sister Species in the Genus Escherichia. *Genomics, Proteomics Bioinforma.* **11**, 61–65 (2013).
106. Citiulo, F., Necchi, F., Mancini, F., Rossi, O., Aruta, M. G., Gasperini, G., Alfini, R., Rondini, S., Micoli, F., Rappuoli, R., Saul, A. & Martin, L. B. Rationalizing the design of a broad coverage shigella vaccine based on evaluation of immunological cross-reactivity among s. Flexneri serotypes. *PLoS Negl. Trop. Dis.* **15**, 1–20 (2021).
107. Sack, D., Lyke, C., McLaughlin, C. & Suwanvanichkij, V. Antimicrobial Resistance in Shigellosis, Cholera and Campylobacteriosis. *World Health Organization* https://apps.who.int/iris/handle/10665/66875 (2001).
108. Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E. & Romanowska, E. Structural studies of Shigella flexneri O-antigens.

Eur. J. Biochem. **91**, 279–284 (1978).
109. Carlin, N. I. A., Lindberg, A. A., Bock, K. & Bundle, D. R. Nature of the biological repeating unit. *Eur. J. Biochem.* **139**, 189–194 (1984).
110. Wehler, T. & Carlin, N. I. A. Structural and immunochemical studies of the lipopolysaccharide from a new provisional serotype of Shigella flexneri. *Eur. J. Biochem.* **176**, 471–476 (1988).
111. Chassagne, P., Fontana, C., Guerreiro, C., Gauthier, C., Phalipon, A., Widmalm, G. & Mulard, L. A. Structural Studies of the O-Acetyl-Containing O-Antigen from a Shigella flexneri Serotype 6 Strain and Synthesis of Oligosaccharide Fragments Thereof. *European J. Org. Chem.* 4085–4106 (2013).
112. Livio, S., Strockbine, N. A., Panchalingam, S., Tennant, S. M., Barry, E. M., Marohn, M. E., Antonio, M., Hossain, A., Mandomando, I., Ochieng, J. B., Oundo, J. O., Qureshi, S., Ramamurthy, T., Tamboura, B., Adegbola, R. A., Hossain, M. J., Saha, D., Sen, S., Faruque, A. S. G. *et al.* Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin. Infect. Dis.* **59**, 933–941 (2014).
113. Noriega, F. R., Liao, F. M., Maneval, D. R., Ren, S., Formal, S. B. & Levine, M. M. Strategy for cross-protection among Shigella flexneri serotypes. *Infect. Immun.* **67**, 782–788 (1999).
114. Knirel, Y. A., Sun, Q., Senchenkova, S. N., Perepelov, A. V., Shashkov, A. S. & Xu, J. O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms. *Biochem.* **80**, 901–914 (2015).
115. Kang, Y., Barbirz, S., Lipowsky, R. & Santer, M. Conformational diversity of O-antigen polysaccharides of the gram-negative bacterium Shigella flexneri serotype y. *J. Phys. Chem. B* **118**, 2523–2534 (2014).
116. Richardson, N. I., Ravenscroft, N., Arato, V., Oldrini, D., Micoli, F. & Kuttel, M. M. Conformational and Immunogenicity Studies of the Shigella flexneri Serogroup 6 O-Antigen: The Effect of O-Acetylation. *Vaccines* **9**, 432 (2021).
117. Hlozek, J., Ravenscroft, N. & Kuttel, M. M. Effects of Glucosylation and O-Acetylation on the Conformation of Shigella flexneri Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B* **124**, 2806–2814 (2020).
118. Theillet, F. X., Simenel, C., Guerreiro, C., Phalipon, A., Mulard, L. A. & Delepierre, M. Effects of backbone substitutions on the conformational behavior of Shigella flexneri O-antigens: Implications for vaccine strategy. *Glycobiology* **21**, 109–121 (2011).
119. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W. & Mackerell Jr, A. D. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564 (2008).
120. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

121. United Nations, W. H. O. No time to wait: Securing the future from drug-resistant infections. *Artforum Int.* **54**, 1–25 (2019).
122. Patel, D. S., Qi, Y. & Im, W. Modeling and simulation of bacterial outer membranes and interactions with membrane proteins. *Curr. Opin. Struct. Biol.* **43**, 131–140 (2017).