Downloaded from orbit.dtu.dk on: Apr 26, 2024

DTU Library

Structural elucidation of polysaccharides and investigations of enzymatic syntesis of oligosaccharides using NMR spectroscopy

Kjeldsen, Christian

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Kjeldsen, C. (2017). *Structural elucidation of polysaccharides and investigations of enzymatic syntesis of oligosaccharides using NMR spectroscopy*. Technical University of Denmark.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Structural elucidation of polysaccharides and investigations of enzymatic synthesis of oligosaccharides using NMR spectroscopy

PhD Thesis

Christian Kjeldsen

Department of Chemistry

Technical University of Denmark

November 2017

Preface

This thesis is the product of three years of research as a PhD student at the Department of Chemistry, Technical University of Denmark. The work was carried out under the supervision of Professor Jens Øllgaard Duus and Senior Researcher Sebastian Meier with financing from the Novo Nordisk Foundation (Biotechnology-based synthesis and production research, program grant no 5371).

First and foremost, I would like to thank Jens for giving me the opportunity to work with NMR spectroscopy and for guidance and support during the last three years, it has been a real inspiration working with you. I would also like to thank Sebastian for guidance, discussions and proofreading of the thesis.

Next, I would like to thank Professor Thomas Peters for allowing me to visit his lab for three months and learn a different field of, and also Friedemann Flügge and the rest of the group at the University of Lübeck for a pleasant stay. Additionally, I would like to thank the Frants Allings legat, Oticon Fonden and Otto Mønsted Fonden for financial support during my external stay.

I owe a big thanks to everyone from the old building 201 and building 211 as well as Cecilie, Nanna, Sofie and Stine who all contributed during their bachelor studies, and additional gratitude to Charlotte, Casper, Kasper and Anne for always taking the time to help me out when the spectrometers and I had disagreeances. I acknowledge the Carlsberg Foundation and the Villum Foundation for the support to the NMR Center · DTU.

Thank you to Jan, Magnus, Sean, Mathilde, Pernille and the entire Hypermag group for a pleasant work environment and fruitful discussions. I also acknowledge the Danish National Research Foundation for funding of the Hypermag center.

I want to thank all of our collaborators from DTU Bioengineering, SSI Diagnostica, Public Health England and The University of Oslo. Especially David Teze for fruitful discussions.

Also, I want to thank all of my friends and family for support and understanding throughout the last three years, and especially Ronja and Daniel for many good discussions, both in and outside the lab.

Last but not least, I would like to thank my amazing wife Tina for always supporting me throughout the project. Without you this would not have been possible.

Abstract

Carbohydrates is a source of complex biomolecules with numerous roles in biology, and elucidation of the structures is imperative for understanding their functions. NMR spectroscopy is an important tool for structure analysis of organic molecules, and for carbohydrates it is often the best option for determining the structures. This thesis is made up of several projects focusing on NMR spectroscopy for carbohydrate structural elucidation as well as transglycosylation studies, and the following sections are small abstracts for each project. All of the projects were part of collaborations with researchers from adjacent fields, and most of the results would not have been obtainable without the use of high-field NMR spectroscopy.

The capsular polysaccharide of *Streptococcus pneumoniae* is responsible for much of its virulence, and as such understanding the structures of the polysaccharides from different serotypes can further the understanding of the pathogen. The capsular polysaccharide structure of a novel serotype in serogroup 7 was elucidated, and the novel serotype turned out to be a mixture of two other serotypes capsular polysaccharides and named 7D in the Danish nomenclature. As one of these two polysaccharides was from serotype 7C, the structure of which was previously undetermined, this was also elucidated.

Inonotus obliquus, more commonly known as the chaga mushroom, has been used in Eastern European folk medicine to treat a variety of symptoms, and contains several compounds of medicinal interest such as polyphenols, triterpenoids and polysaccharides, the last of which was subjected to several types of structural analysis, including NMR spectroscopy. The fungal polysaccharides were extracted and purified, and they turned out to be heterogeneous, but different structural trends were determined.

β-Lactoglobulin is a major protein present in dairy products which is known to bind different types of molecules, but little is known of its ability to bind carbohydrates. For this purpose several polysaccharides were fragmented into oligosaccharide mixtures to be used for binding studies, and in order to understand the implications of those results the composition and structural analysis of the mixtures is required.

Human milk oligosaccharides are important for the immune system of infants, but unfortunately it is barely present in the milk of any domesticated mammal, so a large scale synthesis of human milk oligosaccharides can help improve infant formula. However, these are fairly complex, and when synthesising them via transglycosylation several products can formed, making the structural

determination rather troublesome, even with NMR spectroscopy. The transglycosylation products of three different β-acetylglucosaminidases using lactose as acceptor were identified.

Dissolution dynamic nuclear polarisation can be used to increase the sensitivity of single scan solution ^{13}C-NMR with up to four orders of magnitude, and using a doubly isotopically labelled β-galactopyranoside the *lacZ* β-galactosidase was investigated. Using this method, previously undescribed short-lived transglycosylation products of the enzyme were observed, and using kinetic model the hydrolysis and transglycosylation rates were determined.

Resumé

Kulhydrater er en kompleks klasse af biomolekyler med vidt forskellige roller i biologiske systemer, og udredning af deres strukturer er nødvendigt for at forstå deres funktioner. NMR spektroskopi er et vigtigt redskab til udredning af organiske molekylers strukturer, og for kulhydrater er det ofte den bedste metode. Denne afhandling er opbygget af adskillelige projekter med fokus på NMR spektroskopi til kulhydrat strukturopklaring, og hver af de følgende afsnit er små resuméer af hvert projekt. Alle projekterne er udført i samarbejde med forskere fra tilstødende felter, og de fleste af resultaterne havde ikke været muligt at opnå uden højfelts NMR spektroskopi.

Kapsel polysakkariderne fra *Streptococcus pneumoniae* er ansvarlige for meget af deres virulens, og derfor kan strukturel kendskab til de forskellige serotypers polysakkarider øge forståelsen af patogenet. Kapsel polysakkarid strukturen fra en nyligt opdaget serotype tilhørende serogruppe 7 blev opklaret, og den nye serotype viste sig at være en blanding af to andre serotypers kapsel polysakkarider, og den blev døbt 7D i det danske nomenklatur system. Da en af disse to polysakkarider stammede fra serotype 7C, der tidligere havde en ukendt struktur, blev denne også udredet.

Inonotus obliquus, bedre kendt som chaga svampen, har været brugt i østeuropæisk folkemedicin til at behandle en bred vifte af symptomer, og indeholder adskillelige forbindelser med medicinsk interesse, og en af disse er polysakkarider. Efter ekstraktion fra svampen blev disse undersøgt med forskellige typer strukturanalyse, deriblandt NMR spektroskopi. Polysakkariderne viste sig at være heterogene, men på trods af dette blev diverse strukturelle tenderser beskrevet.

β-Lactoglobulin er valleprotein der findes i mælk fra køer og det er kendt for at binde mange typer molekyler, men kun lidt er kendt angående dets evner til at binde kulhydrater. For at undersøge disse evner blev tre forskellige polysakkarider fragmenteret til oligosakkaridblandinger, og for at forstå resultaterne fra disse bindingsstudier er det nødvendigt at kende sammensætningen og strukturerne af oligosakkariderne i disse blandinger.

Oligosakkarider fra modermælk er vigtige for spædbørns immunsystem, men eftersom disse oligosakkarider er unikke for mennesker og stort set ikke findes i mælk fra noget husdyr er der stor interesse i at udvikle metoder til at syntetisere disse oligosakkarider i stor skala. Ved enzymatisk syntese af disse oligosakkarider opnår man komplekse blandinger af molekyler, og NMR spektroskopi er nødvendigt for at identificere de forskellige transglycosyleringsprodukter. I dette projekt blev transglycosyleringsprodukterne fra tre forskellige β-acetylglucosaminidaser analyseret.

Opløsnings dynamisk nuklear polarisering er en metode til at øge følsomheden af enkelt-skan ^{13}C-NMR med op til fire størrelsesordner, og ved hjælp af et dobbelt isotopmærket β-galaktopyranosid blev *lacZ* β-galaktosidase undersøgt. Ved hjælp af denne metode blev nye kortlivede transglycosyleringsprodukter observeret, og ved hjælp af en kinetisk model blev hydrolyse og transglycosylerings hastighederne bestemt.

List of abbreviations

1D	one-dimensional
2D	two-dimensional
AATGalp	2-acetamido-4-amino-2,4,6-trideoxy-galactopyranoside
Ac	acetate
ax	axial
Cho	choline
CLIP	clean inphase
COSY	correlation spectroscopy
cps	Capsular polysaccharide
CWPS	cell wall polysaccharide
dDNP	dissolution dynamic nuclear polarisation
DNP	dynamic nuclear polarisation
DQF	double quantum filtered
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
eq	equatorial
f	furanoside (for example Galf)
FID	free induction decay
Gal	galactose
GalA	galacturonic acid
GalNAc	*N*-acetyl galactosamine
GC	gas chromatography
Glc	glucose
GlcA	glucuronic acid
GlcNAc	*N*-acetyl glucosamine
GTA	α-1,3-galactosyltransferase
GTB	α-1,3-*N*-acetylgalactosaminyltransferase
GulA	guluronic acid
H2BC	heteronuclear two-bond correlation
HMBC	heteronuclear multiple-bond correlation
HMO	human milk oligosaccharide
HMQC	heteronuclear multiple-quantum correlation
HPLC	high-performance liquid chromatography
LNT2	lacto-*N*-triose II

Man	mannose
ManA	mannuronic acid
Me	methyl
MRI	magnetic resonance imaging
MS	mass spectrometry
NAG-oxazoline	2-methyl-(1,2-dideoxy-α-D-glucopyrano) [2,1-d]-oxazoline
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
onp	ortho-nitrophenol
p	pyranoside (for example Galp)
PCV	protein-conjugate vaccine
pnp	para-nitrophenol
ppm	parts per million
PPV	pneumococcal polysaccharide vaccine
RF	radio-frequency
Rha	rhamnose
Rib	ribose
ROESY	rotating frame Overhauser effect spectroscopy
SNR	signal to noise ratio
TLC	thin layer chromatography
TOCSY	total correlation spectroscopy
UDP	uridine diphosphate
Xyl	xylose

Publications

Paper 1

Kjeldsen, C.; Slott, S; Elverdal, P. E.; Sheppard, C. L.; Kapatai, G.; Fry, N. K.; Skovsted, I. C.; Duus, J. Ø. Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D, *manuscript in preparation*

Paper 2

Wold, C. W.; **Kjeldsen, C.**; Corthay, A.; Rise, F.; Christensen, B. E.; Duus, J. Ø.; Inngjerdingen, K. T. Structural characterization of bioactive heteropolysaccharides from the medicinal fungus *Inonotus obliquus* (Chaga), *manuscript submitted to Carbohydrate Polymers*

Paper 3

Stender, E. G. P.; Birch, J.; **Kjeldsen, C.**; Nielsen, L. D.; Duus, J. Ø.; Kragelund, B. B.; Svensson, B. pH dependent binding patches of alginate trisaccharides on β-lactoglubolin investigated by HSQC-NMR, *manuscript in preparation*

Paper 4

Kjeldsen, C.; Ardenkjær-Larsen, J. H.; Duus, J. Ø. Discovery of novel intermediates of *lacZ* β-galactosidase catalyzed hydrolysis using dDNP NMR, *manuscript in preparation*

Paper 5

Visnapuu, T.; Teze, D.; **Kjeldsen, C.**; Duus, J. Ø.; Lie, A.; André-Miral, C.; Pedersen, L. H.; Stougaard, T.; Svensson, B. GH20 β-*N*-acetyl hexosaminidase from psychrotrophic marine bacterium *Paraglaciecola hydrolytica*, *manuscript in preparation*

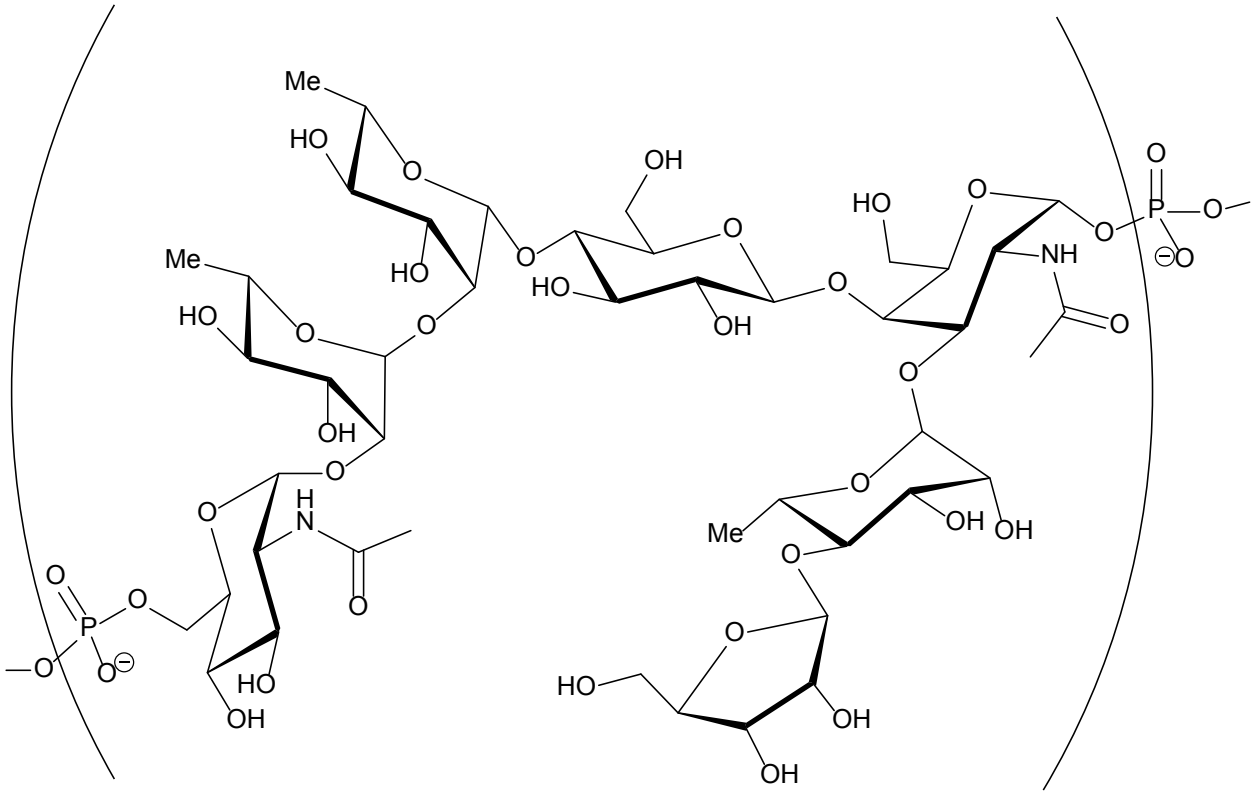
All of these papers are included in the thesis.

Table of Contents

1 Introduction

1.1 Carbohydrates

Of the four major biological macromolecule classes, carbohydrates are by far the most complex, as they can be highly branched and derivatised, whereas other biological polymers such as nucleic acids and proteins are mostly linear[1]. An example of a polysaccharide repeating unit structure is shown in scheme 1, and each of the hydroxyl groups could be a point for further branching or derivatisation, highlighting the potential for complexity of carbohydrate structures.



Scheme 1 – Repeating unit structure of the capsular polysaccharide from Streptococcus pneumoniae serotype 7C. The full structure elucidation can be found in chapter 2.1.

This high complexity makes carbohydrates optimal for specific recognition, which also happens to be one of the major functions of oligo- and polysaccharide. For example in *E. coli*, the *lac* operon will produce β-galactosidase in the presence of lactose, enabling the metabolism of lactose in the bacteria, and while the enzyme will hydrolyse most β-galactopyranosides, structurally similar molecules like α-galactopyranosides or β-glucopyranosides, where only one stereocenter is inverted, are not substrates[2].

This type of specificity can be seen in numerous protein-carbohydrate interactions, and another example could be the ABO(H) blood group A and B antigen enzymatic synthesis, in which two

glycosyltransferases that only differ in four out of 354 amino acid residues can differentiate between uridine diphosphate (UDP) *N*-acetylated galactosamine and galactose as the donor[3], while using the same acceptor.

Other uses of highly complex and diverse polysaccharide structures can be found in bacterial pathogens capsular polysaccharides (cps), which is used to mask it from the host immune system, and consequently many antibacterial vaccines are based on the cps[4,5]. An example of a bacterial cps is shown in scheme 1, and if one of these seven monosaccharides was substituted for another, or if one of them had any additional substitution or derivatisation on any of the free hydroxyl groups, it would most likely have a different immune response.

Carbohydrates can also assist with protection from diseases in other ways, human milk, for example, contains oligosaccharides that have a number of different health benefits for infants, including protection against pathogens. The oligosaccharides can prevent pathogen attachment to mucosal surfaces by acting as antiadhesives and lowers the risk infections. The problem with these oligosaccharides is that they are barely present in milk from any domesticated mammal, leaving infants without access to human milk at a higher infection risk[6]. Improving the current availability of those oligosaccharides as prebiotics could potentially increase the health of infants relying on formula[7,8].

Another potential source of bioactive and potentially health improving carbohydrates is from fungi[9]. Generally, fungi are known to produce unique structures, and this might also be the case for carbohydrates obtained from fungi[10]. The fungi *Inonotus obliquus*, more commonly known as the chaga mushroom, have been used as a medicinal mushroom in Russian and Eastern European folk medicine[11], and it has been observed to have immunomodulatory effects on mice[12], but elucidation of the structural properties of the polysaccharides present in the mushroom have been lacking.

Therefore, while carbohydrates are perhaps most commonly known for their role as an energy source for biological life, their high potential for complexity allows them to have a broad range of utilisations in nature. In order to better understand the roles of carbohydrates, knowledge of their structures is important to understand how and why they are used in biology.

1.2 NMR spectroscopy for carbohydrate elucidation

Nuclear magnetic resonance (NMR) spectroscopy is often the most reliable tool for carbohydrate structure elucidation[13], as other structure elucidation methods such as mass spectrometry and X-ray crystallography have more limited uses for carbohydrates compared to other biological polymers. While analytical methods such as gas chromatography (GC), liquid chromatography (LC) and mass spectrometry (MS) can provide structural information, they lack information regarding glycosidic linkages and α/β-configurations, but generally these methods are much more sensitive than NMR, requiring less analyte[13–15].

For example, in paper 2 GC and GC-MS analysis was used alongside NMR spectroscopy to extract structural knowledge of structures of the chaga heteropolysaccharides, with GC-MS analysis giving information regarding substitution patterns and concentrations of the different monosaccharides present in the polysaccharide, and NMR spectroscopy yielding information regarding α/β-configuration and glycosidic linkages.

1.2.1 Basics of NMR spectroscopy

Very briefly described, NMR spectroscopy relies on the alignment, or polarisation, of nuclear spins when subjected to a magnetic field and the subsequent manipulation of said spins. The outcome of these spin manipulations are then measured using a radio-frequency (RF) module, which is also used to induce the aforementioned manipulations, present in the NMR probe. The simplest of NMR methods would be a standard proton NMR experiment, in which the polarised spins are subjected to a hard pulse, usually of 90° or 30°, and the following relaxation to equilibrium produces RF signals, known as free induction decay (FID), which can then be Fourier transformed to a frequency spectrum.

Due to low sensitivity of NMR spectroscopy, signal averaging by summation of scans is usually required, but this also adds noise. However, the random nature of the noise makes it increase at a slower rate than the signal intensity, yielding the relation that the signal-to-noise ratio (SNR) increases by the square root of the number of scans. So to achieve double the SNR, four times the number of scans, and consequently four times the acquisition time, is required. This results in NMR both requiring large amounts of analyte, compared to most other analytical methods, and long acquisition times, limiting its uses for time sensitive applications such as reaction monitoring or measuring of unstable compounds.

1.2.2 Carbohydrates and NMR

As carbohydrates are primarily made up of hydrogen, carbon and oxygen, the structural elucidation will rely on the information gained from these nuclei. NMR spectroscopy is relying on spin ½ nuclei, and for hydrogen the most abundant isotope ^{1}H (>99.9%) is a spin ½ nuclei. For carbon, the second most abundant isotope, ^{13}C (1.1%), is also a spin ½ nuclei. Oxygen, however, is less useful for NMR spectroscopy, and thus less commonly used, as the only NMR active isotope, ^{17}O (0.04%), is a spin 5/2 and quadrupolar. Therefore, while it is indeed possible to do oxygen NMR, it is highly impractical and is seldom used. Other nuclei, such as nitrogen and phosphorous, which can be present in carbohydrates, are rarely used for assignments.

Bearing this in mind, the structural elucidation of carbohydrates generally rely on the NMR properties of protons and carbons. In the following sections, the NMR experiments used throughout the different projects will be discussed, and as most of the carbohydrates were pyranosides, the discussion of each experiment is in regard to glycopyranoside elucidation, but much of it also applies to furanosides. Lastly, as carbohydrate NMR is often performed in aqueous solution, the use of solvent chemical shifts[16,17] for calibration is rather unreliable, and an added standard is often needed.

1.2.2.1 1D ^{1}H- and ^{13}C-NMR

The 1D ^{1}H and ^{13}C experiments are the simplest, and in the case if ^{1}H also the most sensitive NMR experiment. The information obtainable from a ^{1}H experiment includes coupling constants, integrals and information regarding the chemical environment of each ^{1}H signal. However, in some cases the usefulness of 1D ^{1}H NMR can be quite limited due to the fairly narrow spectral frequency, approximately 15 ppm, which causes signals from similar protons to have a tendency to overlap. This is especially true for carbohydrates, in which the bulk protons, H2-H6, all usually lie between 3 and 4 ppm. However, for the anomeric proton signals there is usually far less overlap, and the information gained from the anomeric region, usually 4.5-5.5 ppm, alone is usually invaluable for the full assignment.

Carbon NMR is much less prone to overlap as the spectral frequency is much broader, approximately 200 ppm, and in commonly used ^{13}C-NMR pulse sequences the ^{1}H-^{13}C couplings will be decoupled, and carbon will appear as singlets, further reducing the probability of spectral overlap. The drawback of ^{13}C NMR is the low sensitivy, caused by the low natural abundance and gyromagnetic ratio, which results in long acquisition times being required compared to ^{1}H.

However, due to the complexity of carbohydrates, 1D NMR is rarely sufficient to perform structure elucidations, and spectra with more information are required.

1.2.2.2 COSY

Correlation spectroscopy (COSY) was the first 2D experiment to be developed[18,19], and as the name suggests it correlates signals between homonuclear spins, usually ^{1}H. This method is immensely useful for assigning the protons of each spin system, which often represents an entire monosaccharide. The double quantum filtered (DQF)[20] version of COSY improves the usefulness of COSY, as it allows for the coupling constants to be measured as well and suppresses signals from singlets, although at the cost of sensitivity, and thus requires longer acquisition times. For carbohydrates, the $^{3}J_{^{1}H,^{1}H}$ coupling constants are often the best way to differentiate between the different monosaccharide configurations, making DQF-COSY invaluable for carbohydrate structural elucidation.

However, due to the narrow signal frequency of ^{1}H, COSY spectra, and indeed any ^{1}H homonuclear 2D experiment, is prone to signal overlap.

1.2.2.3 TOCSY

Total correlation spectroscopy (TOCSY)[21], or homonuclear Hartmann-Hahn spectroscopy (HOHAHA), is similar to COSY, except that it shows correlations through several couplings, which can assist with discerning couplings in spectra with overlap. This works better for systems with strong couplings, which in pyranosides means large H_{ax}-H_{ax} couplings.

1.2.2.4 NOESY and ROESY

Nuclear Overhauser effect spectroscopy (NOESY)[22,23] uses the effects of cross relaxation between nuclei to establish correlations between nuclei of up to 5 Å appart. For carbohydrates this has some important uses: Correlations between axial protons on the same face of a pyranoside, e.g. between H1/H3/H5 in β-galactopyranosides, in which finding a $^{3}J_{^{1}H,^{1}H}$ correlation between H3/H4/H5 can be troublesome. Furthermore, correlations between monosaccharides in a di-, oligo- or polysaccharide can be used to determine glycosidic linkages, and NOE signals is often one of the most important tools for polysaccharide structural elucidation.

One of the drawbacks of NOESY is that some molecules will tumble too quickly to develop the dipolar correlations required and thus give little to no signal. This tumbling is dependent mainly on the size of the molecule and the magnetic field.

Rotating frame Overhauser effect spectroscopy (ROESY)[24] is quite similar to NOESY, but does not suffer from the problem of NOESY with tumbling of small molecules, and for macromolecules they give very similar results.

1.2.2.5 HSQC

Heteronuclear single-quantum correlation spectroscopy (HSQC)[25] correlates two heteronuclear spin through their one-bond correlation. This method is extremely useful for carbohydrates, as it can be used to take advantage of the broad frequency range of carbon and the high sensitivity of proton NMR. Furthermore, similar to the distortionless enhancement by polarisation transfer (DEPT)[26] experiment, HSQC can be edited to differentiate between CH/CH_3 and CH_2 groups[27]. Generally speaking, biggest drawback of HSQC data is the lack of information regarding quaternary carbon, however, as carbohydrates rarely have many, if any, quaternary carbon, this is less problematic compared to most other organic molecules. Lastly, for determining the $^1J_{^1H,^{13}C}$ coupling constants, a clean inphase (CLIP) HSQC experiment can be used[28].

HSQC can also be coupled with other experiments such as NOESY or TOCSY, the latter of which is known as HSQC-TOCSY[29] and can be of great use when assigning carbohydrate structures, as regular ^1H homonuclear 2D experiments are prone to extensive overlap, especially in the bulk region.

A similar experiment to HSQC is the heteronuclear multiple-quantum correlation spectroscopy (HMQC)[30] experiment, but generally HSQC is more suited for larger molecules, and for smaller molecules the methods produce very similar results.

1.2.2.6 HMBC

The heteronuclear multiple-bond correlation spectroscopy (HMBC)[31] experiment is quite similar to the HMQC experiment, except, as the name suggests, it is tailored towards the correlations between nuclei several bonds away, usually 2-4. For carbohydrates, this is useful for similar reasons as NOESY, except that it is less sensitive and less prone to spectral overlap. Furthermore, HMBC is one of the only experiments that can detect quaternary carbon, along 1D ^{13}C-NMR, and while this might not be important for most carbohydrates, as they generally have few quaternary positions, it does become relevant, for example when analysing acetylated or uronic acid containing carbohydrates where the carbonyls can be difficult to detect.

One of the drawbacks of HMBC is that it can be difficult to discern 2-, 3- and 4-bond correlations from one another, but usually this can be overcome by comparing with HSQC-TOCSY or COSY data. Another experiment that can be used to better differentiate between 2- and 3- bond couplings is the heteronuclear two-bond correlation (H2BC) experiment[32,33].

1.2.3 Structural analysis

1.2.3.1 Common trends in NMR spectroscopy of glycopyranosides

For the common pyranosides, such as glucose, galactose and mannose, some trends can be observed. For example, when not substituted: the bulk protons will usually be between 3 and 4 ppm, with carbon values for C2-C5 between 68 and 74 ppm and C6 around 60 ppm. If substituted, the ^{13}C chemical shift of a given position will move upwards of 10 ppm downfield, often making it stand out, while the ^{1}H shifts are harder to predict. This makes it possible to compare with data from the literature when attempting to determine substitution patterns. However, due to these fairly narrow chemical shift ranges, the bulk region is often prone to signal overlap, especially for more complex samples.

Similarly, the anomeric positions will be around 4.5-5.5 ppm in the proton frequency and 88-110 ppm for carbon, with α-anomeric positions having the more downfield ^{1}H and more upfield ^{13}C chemical shifts compared to β. Reducing ends will have similar proton chemical shifts, but the carbon shifts will be more upfield compared to a non-reducing anomeric position.

Additionally, because of these trends, database tools like CASPER have been developed, and can be used to predict the chemical shifts of different carbohydrate structures[34–36].

As for coupling constants, the $^{3}J_{^{1}H,^{1}H}$ is often around 10 Hz for H_{ax}-H_{ax}, unless it is from the anomeric proton where its often closer to 8 Hz, and if one of them is equatorial, the couplings usually get quite low and often it is not possible to distinguish H_{ax}-H_{eq} from H_{eq}-H_{eq}[37]. For this reason it can be quite difficult to determine the anomeric configuation of monosaccharides such as mannose and rhamnose, as the H2 is equatorial, and thus the $^{1}J_{^{1}H,^{13}C}$ coupling constant can be necessary, as it is a rather robust method of determining the anomeric configuration. For most glycopyranosides, the $^{1}J_{^{1}H,^{13}C}$ coupling constant is approximately 160 Hz for β- and 170 Hz for α-configurations[38].

The above trends are only for glycopyranosides, and there are exceptions to all of the statements, but as general rules of thumb they can be quite useful.

1.2.3.2 Example of procedure for assignment

When looking at the data, a good place to start is often the HSQC spectra, as even a quick glance will yield information about the number of different anomeric signals and if any 1,6-linkages are present, as the 6-positions for the common hexoses have quite distinct ^{13}C chemical shifts depending on their substitution as well as opposite phase to the remaining signals.

After having noted the peaks in the HSQC data, the next step is usually to attempt to assign each spin system, or monosaccharide, using spectra like COSY, TOCSY and HSQC-TOCSY. However, if the sugar is not of glucose configuration, and thus not having large $^{3}J_{^{1}H,^{1}H}$ couplings from H2-H5, it might often be necessary to also use HMBC and NOESY to assign all the positions. For example, galactose it can often be difficult to correlate the 5- and 6-position to the remaining positions, but using NOESY or HMBC it is often possible to correlate the 1- and 5-positions or the 3- and 5-positions, overcoming the problem with low couplings.

When having all the positions of a monosaccharide assigned, the last step of determining what monosaccharide it is will usually involve the coupling constants, but sometimes, due to either spectral overlap or inconclusive coupling constants, the comparison to literature assignments must suffice[37,39,40].

Finally, when all the positions of an oligo- or polysaccharide is assigned, the final step is the determination of the linkages between monosaccharides, and as mentioned earlier, this can often be done using HMBC or NOESY, or sometimes both, and can also be supported by chemical shift comparison with literature.

There is one important piece of structural information that NMR spectroscopy does not yield for carbohydrates: Whether the monosaccharides have L or D configurations, as enantiomers have identical chemical shifts. However, when comparing the chemical shifts of a disaccharide, or larger sugar, to a similar or identical molecule from literature, the shifts should only be the same if either all or none of the monosaccharides has the same L or D configuration as the corresponding reference.

1.3 Dissolution dynamic nuclear polarisation

A general problem of NMR spectroscopy is the low sensitivity, which is in part due to the low polarisation of nuclear spins - even at high magnetic fields. The spin polarisation can be expressed by the following equation:

$$P = \frac{n1-n2}{n1+n2} = \tanh\left(\frac{\hbar\omega}{2k_BT}\right) \;\&\; \omega = \gamma B_0$$

Where n1 and n2 represents of the population of the two spin states, $\hbar$ is the Planck constant, k_B is the Boltzmann constant, T is the temperature, γ is the gyromagnetic ratio, and B_0 is the magnetic field. This equation shows that for a 9.4 T NMR magnet, which corresponds to 400 MHz in the ^{1}H frequency, operating at 25 °C, only 32 and 8 ppm of ^{1}H and ^{13}C spins, respectively, are polarised. So even if going to the highest field currently possible of 1.2 GHz[41], only triple the polarisation can be achieved in thermal NMR spectroscopy. The equation is plotted in figure 1 for electron, proton and carbon spins at 3.35 T.

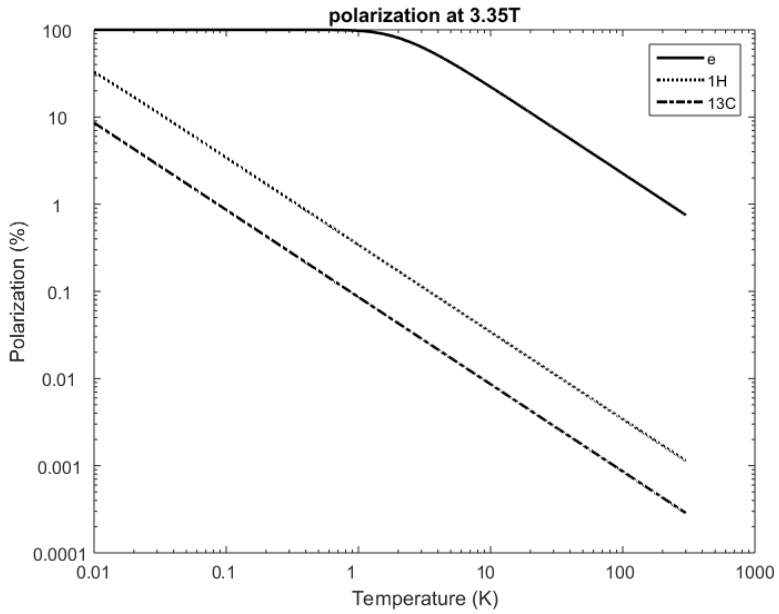


Figure 1 – Polarisation of e, ^{1}H and ^{13}C spins at 3.35 T as a function of temperature.

When looking into the equation, the only parts available for increasing the polarisation would be the magnetic field, the temperature and the gyromagnetic ratio. As already mentioned, even high field NMR spectroscopy still yields low polarisation, and in solution state NMR it is rarely an option to go to very low temperatures, in particular when investigating enzymes or other biological processes. This leaves the gyromagnetic ratio, and for electrons it is approximately 660 times higher than for

protons, and approximately 2620 times higher than that of ^{13}C. Therefore, if one was to transfer the polarisation from electron spin to nuclear spin, a much higher polarisation could be achieved.

These effects can be utilised in dynamic nuclear polarisation (DNP), a method first described in the early 1950's[22,42], in which the sample is mixed with a stable radical, cooled to low temperatures in a magnetic field, followed by polarisation transfer from electron spins to nuclei spins using microwaves. For example, at 1.4 K and 3.35 T the electron spins will reach approximately 92% polarisation, and this can be transferred to nearby nuclei by using microwaves at the electron Larmor frequency, which in the case of 3.35 T would be approximately 94 GHz. This polarisation can transfer to more remote nuclear spins by spin diffusion, and due to the slow relaxation of nuclei at low temperatures, while electrons still have rapid relaxation, a buildup of polarisation can be achieved. Eventually, the polarisation will reach a plateau when the relaxation and buildup rates balances out, and at 1.4 K and 3.35 T it is possible to reach >10% polarisation for ^{13}C.

However, the sample would then be at 1.4 K and in a glassy solid state, which as mentioned would not be particularly useful for liquid-state NMR at room temperature or 37° C, nor for magnetic resonance imaging (MRI). This was solved in 2003 by Ardenkjær-Larsen and co-workers[43] by making a rapid dissolution of the solid sample using heated water, followed by transfer to an NMR magnet. Using this method it was possible to achieve an SNR increase of four orders of magnitude, as exemplified in figure 2 with *o*-nitrophenyl β-galactopyranoside.

However, this great increase in SNR will disappear alongside the polarisation when the spins relax to thermal equilibrium with T_1 relaxation, somewhat limiting its applications to long T_1 probes and preferably fast chemical or biochemical systems. DNP is also used for solid state NMR[44], and other types of hyperpolarisation exists[45,46], but neither will be addressed further.

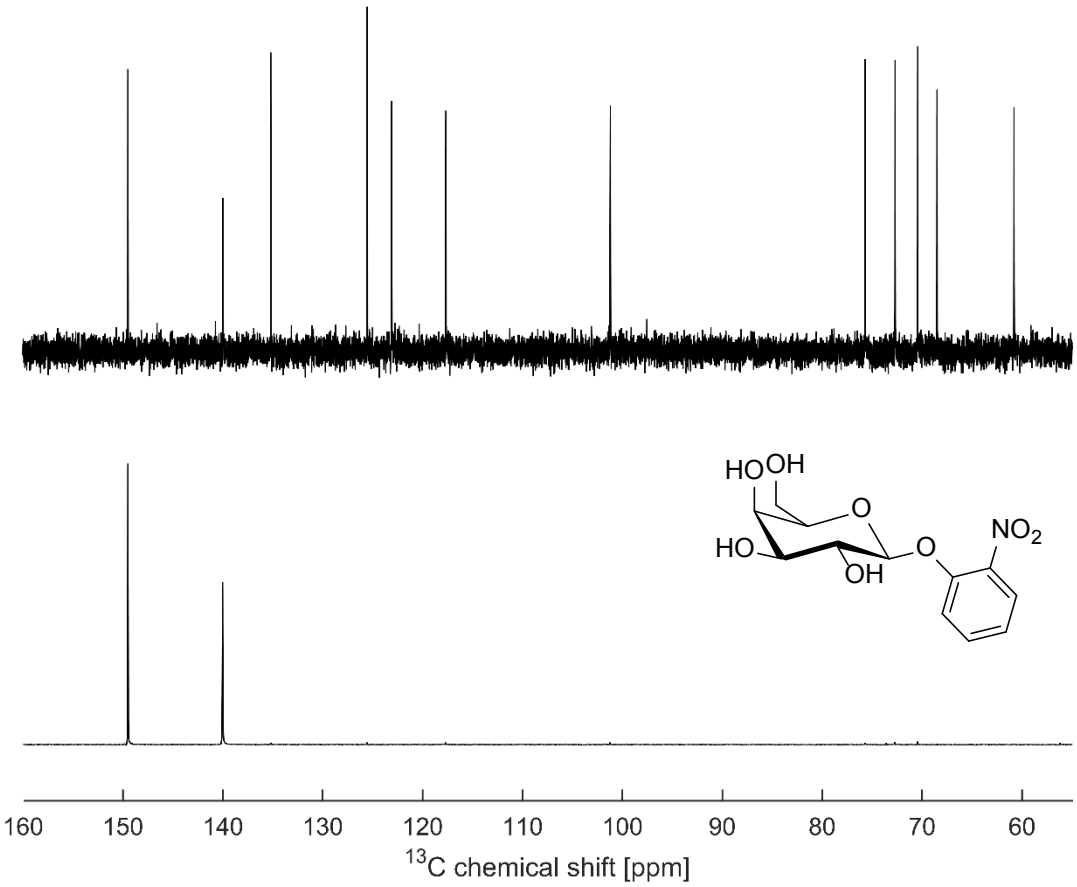


Figure 2 – Example of dDNP NMR compared to thermal NMR of the same sample containing o-nitrophenyl β-D-galactopyranoside. Bottom: First scan after dissolution with an SNR of 778. Top: 26000 scans of the same sample at thermal conditions with an SNR of 19. It would take approximately $44 \cdot 10^6$ scans to reach the same SNR, which would require almost four years of acquisition time at the conditions used. Note that only the two quaternary positions of the nitrophenyl ring can be observed.

In the 2003 paper[43] the DNP probe used was ^{13}C-labelled urea as it has very long T_1 relaxation time of approximately 60 seconds, making it very cumbersome to work with in thermal NMR and great for dDNP. Another commonly used probe for dDNP MRI is pyruvate, shown in scheme 2, as it is has favourable properties for dDNP, namely a long T_1 relaxation on the quaternary positions as well as forming a glassy solid when neat, allowing for high concentration and purity during polarisation build-up. Furthermore, it has important biological roles and can be injected in fairly large doses due to its low toxicity, making it an optimal probe for MRI[47,48]. Exploiting the Warburg effect[49,50], which results in increased concentrations of lactate in cancer cells, hyperpolarised pyruvate can be used to probe for tumors[51], and in 2013 the first human study using the dDNP MRI method was published in which it was used to probe for prostate cancer in patients[52].

Scheme 2 – Two examples of long T_1 molecules, pyruvate and lactate.

However, in chemistry and biochemistry, the use of dDNP NMR has been halted somewhat by the lack of a similar substrate to pyruvate, and many molecules of interest have much shorter T_1 relaxation times. Nevertheless, several examples exist in which similar molecules have been used[48,53–55]. An example of this could be acetate, which has been used to detect and quantify reaction intermediates in enzymatic conversion of acetate to acetohydroxamate[54]. In said study, the intermediates of interest was of a multistep reaction of two enzymes, acetyl-coenzyme A synthetase and citrate synthetase.

In most of the examples of reaction monitoring using ^{13}C dDNP NMR, ^{13}C-labelling have been used to further increase the sensitivity by another two orders of magnitude, but as shown by Hilty and co-workers it is indeed possible to suffice with natural abundance[56–60]. In the first of these studies, Bowen and Hilty (2008)[56] described an automated rapid transfer system from polariser to NMR magnet, and demonstrated its usefulness using a trypsin catalysed hydrolysis of benzoyl-L-arginine ethyl ester. Using the same reaction, another study showed that having a deuterated leaving group, d3-methanol, increased the T_1 of the carbonyl it was attached to from 9 to 11 seconds by reducing the remote polarisation[61]. A similar observation was made regarding fructose, in which deuteration increased the T_1 of the anomeric position from 17 to 35 seconds[48]. So not only is it preferable to have ^{13}C labelling, but ^{2}H labelling as well. Another fairly common trend in dDNP is the addition of a paramagnetic agent to the DNP sample, usually a Gd-chelate, as it can increase the total polarisation by increasing the electron relaxation rates[62]. However, the trade-off is a shortening of the nuclei T_1 time after dissolution, and in some cases[63], little to no polarisation boost is achieved.

Several examples of carbohydrates being used for dDNP NMR also exists, and is discussed further in chapter 3.1, where the hyperpolarised analyte presented in figure 2 is used for additional dDNP studies.

2 Structural elucidation of polysaccharides

In this chapter three different examples of NMR spectroscopy used for polysaccharide elucidation. In the first example, the elucidation of two capsular polysaccharides from Streptococcus pneumoniae are presented, and the second example is concerning different polysaccharides isolated from the chaga mushroom. Finally, the last example is about the structural determination of oligosaccharide mixtures obtained from polysaccharide degradation.

2.1 Structural analysis of a novel serogroup 7 *Streptococcus pneumoniae* serotype

This project was part of a collaboration with SSI Diagnostica and Public Health England, and my part of the work was the NMR assignment of the isolated capsular polysaccharides. The work of our collaborators can be found in paper 1, but will be presented briefly when appropriate.

2.1.1 Introduction

Streptococcus pneumoniae is a Gram-positive bacteria and a common human pathogen. Estimates for the mortality in children under the age of 5 ranges from around 5%[64] to 11%[65], and in 2005 WHO estimated that 1.6 million people die of the disease annually[66]. Historically, the pathogen has had severe impact and has caused much of the mortality in pandemic influenza, including the extraordinarily severe 1918-1919 outbreak known as the Spanish flu[67].

The high virulence of the pathogen can largely be attributed to the capsular polysaccharide camouflaging it from the immune system[68], and as a result of this, the capsular polysaccharide is a subject of great interest. The different *Streptococcus pneumoniae* serotypes are based on their capsules reaction with rabbit antisera, and at least 92 serotypes have been identified[4,69,70]. These serotypes are further classified in serogroups based on the Danish serotype classification system. Additional serotypes have been proposed based on genetics[70,71], but even with different genes they can still produce an identical capsular polysaccharide, as was the case for the putative serotype 6E[72], as NMR analysis of the capsular polysaccharide found that it was identical to that of 6B[73], and similarly for 23B[74], where several genetically distinct strains with different genetic lineages produce the same cps.

The currently used vaccines, the 23-serotype valent pneumococcal polysaccharide vaccine (PPV) which is mostly used for the elderly and 7-, 10- and 13-valent protein-conjugate vaccines (PCV) which is used for children[75,76], cover the most common serotypes and target the capsular polysaccharide. High levels of vaccination have caused an increase in infections from more rare

serotypes, and could conceivably cause novel serotypes to emerge as the pathogen evolves to circumvent the vaccines[77–80]. In order to better survey such effects, our collaborators developed a new serotyping tool called PneumoCaT[81], during which a novel serotype belonging to serogroup 7 was discovered. This work covers the structural analysis of the cps of this new serotype by NMR spectroscopy, and the work of our collaborators can be found in paper 1.

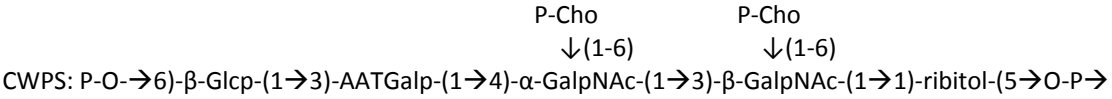
2.1.2 Serogroup 7

Four different serotypes had previously been classified into serogroup 7[4], and one of them, 7F, is quite an important serotype as it is a part of the 10- and 13-valent PCV vaccines administered to infants as well as the PPV vaccine. Serotype 7F is also the most common of the four known serotypes in serogroup 7, causing more than 95% of the serogroup 7 infections.

All the four capsular polysaccharides from previously known serotypes 7F, 7A, 7B and 7C as well as from the new serotype 7D were identified by NMR spectroscopy. Three of these serotypes had known structures, 7F[82], 7A[83] and 7B[84], while 7C had not previously been described, and 7D was a novel serotype.

2.1.2.1 CWPS structure

All the samples contain impurities of up to 5% w/w of pneumococcal cell wall polysaccharide (CWPS)[85], the structure of which has been described previously[86].



The repeating unit of CWPS can be either monocholinated or dicholinated (shown above). The monocholinated would only have a phosphocholine on the α-GalpNAc. A third possible structure also exists, where the β-Glcp is replaced by a β-Galp[87], but that does not correspond to the structures observed in the samples described here. The structure contains a fairly rare monosaccharide, 2-acetamido-4-amino-2,4,6-trideoxy-galactopyranoside (AATGalp), as well as ribitol and a phosphodiester as part of the repeating unit. When analysing the capsular polysaccharide spectra, care should be taken to not confuse signals from the CWPS with signals from the capsular polysaccharide of interest.

2.1.2.2 7F capsular polysaccharide structure

The repeating unit structure of the 7F capsular polysaccharide has previously been described[82] and our data support this repeating unit structure:

7F: →6)-α-Galp-(1→3)-β-Rhap2Ac-(1→4)-β-Glcp-(1→3)-β-GalpNAc-(1→
↑(1-2) ↑(1-4)
β-Galp α-GlcpNAc-(1→2)-α-Rhap

The repeating unit is a heptasaccharide containing two branches, one O-acetylation and two *N*-acetyl monosaccharides. A full NMR assignment can be found in the supplementary information of paper 1.

2.1.2.3 7A capsular polysaccharide structure

The repeating unit structure of the 7A capsular polysaccharide have previously been described[83] and our data supports this repeating unit structure:

7A: →6)-α-Galp-(1→3)-β-Rhap2Ac-(1→4)-β-Glcp-(1→3)-β-GalpNAc-(1→
↑(1-4)
α-GlcpNAc-(1→2)-α-Rhap

The 7A cps has the same repeating unit structure as 7F, except for the branched β-Galp which is absent, and can be explained by the inactivation of the glycosyltransferase responsible for the branching β-Galp[88,89]. A full NMR assignment can be found in the supplementary information of paper 1.

2.1.2.4 7B capsular polysaccharide structure

The repeating unit structure of the 7B capsular polysaccharide has previously been described[84] and our data support this repeating unit structure:

7B: P-O→6)-α-GlcpNAc- (1→2)-α-Rhap-(1→2)-β-Rhap-(1→4)-β-Glcp-(1→4)-α-Glcp-(1→O-P→
↑(1-3)
β-Ribf-(1→4)-α-Rhap

Unlike 7A, this structure is quite different from 7F, and it contains no O-acetyl groups but have a phosphodiester as part of the repeating unit backbone. A full NMR assignment of 7B can be found in table 1.

2.1.2.5 7C capsular polysaccharide structure

The repeating unit structure for the capsular polysaccharide of 7C was to the best of our knowledge previously not reported in literature and was identified, see section 2.1.3.1, as part of the project to identify the new 7D capsular polysaccharide structure as they had very similar spectra, and the structure of 7C was found to be:

7C: P-O→6)-α-GlcpNAc- (1→2)-α-Rhap-(1→2)-β-Rhap-(1→4)-β-Glcp-(1→4)-α-GalpNAc-(1→O-P→
↑(1-3)
β-Ribf-(1→4)-α-Rhap

It should be noted that the only difference between the repeating unit structures of 7B and 7C is the replacement of the disubstituted α-Glcp with an α-GalpNAc, which is also the position connected to the phosphodiester. Another interesting feature is the conservation of the α-GlcpNAc-(1→2)-α-Rhap through all the serotypes, even though it is part of the backbone in 7B and 7C while it is a side chain in 7F and 7A. Another similarity between 7F/7A and 7B/7C is the β-Rhap-1,4-β-glc present in all four structures, although in 7F and 7A it contains an O-acetyl group in the 2-position of the rhamnoside.

2.1.2.6 7D capsular polysaccharide structure

The novel serotype 7D was the main goal of the project, see section 2.1.3.2, and the capsular polysaccharide structure was determined to have the following two repeating units present in an approximately 5:1 ratio:

5x P-O→6)-α-GlcpNAc- (1→2)-α-Rhap-(1→2)-β-Rhap-(1→4)-β-Glcp-(1→4)-**α-GalpNAc**-(1→O-P→
↑(1-3)
β-Ribf-(1→4)-α-Rhap

1x P-O→6)-α-GlcpNAc- (1→2)-α-Rhap-(1→2)-β-Rhap-(1→4)-β-Glcp-(1→4)-**α-Glcp**-(1→O-P→
↑(1-3)
β-Ribf-(1→4)-α-Rhap

These two repeating units are those of 7C (top) and 7B (bottom), and as mentioned above, they were in a 5:1 ratio, making it some variant of hybrid serotype. Because of this, it was first necessary to fully assign the 7C cps to learn the structure the 7D cps. The only difference between the two structures present in 7D is highlighted.

2.1.3 Results and discussion

2.1.3.1 Serotype 7C capsular polysaccharide structural elucidation

The ^{1}H-NMR spectrum of 7C, shown in figure 3 with zoom of the anomeric region in figure 8, contains seven signals from anomeric protons with similar intensities. The signal at 5.42 has a $^{3}J_{^{1}H,^{31}P}$ coupling of 6 Hz which also explains its downfield ^{1}H chemical shift compared to the other anomeric positions. Furthermore, the ^{1}H-NMR spectrum shows five methyl group signals, three from C-Me signals around 1.3 ppm and two from *N*-Ac signals of similar intensities around 2 ppm.

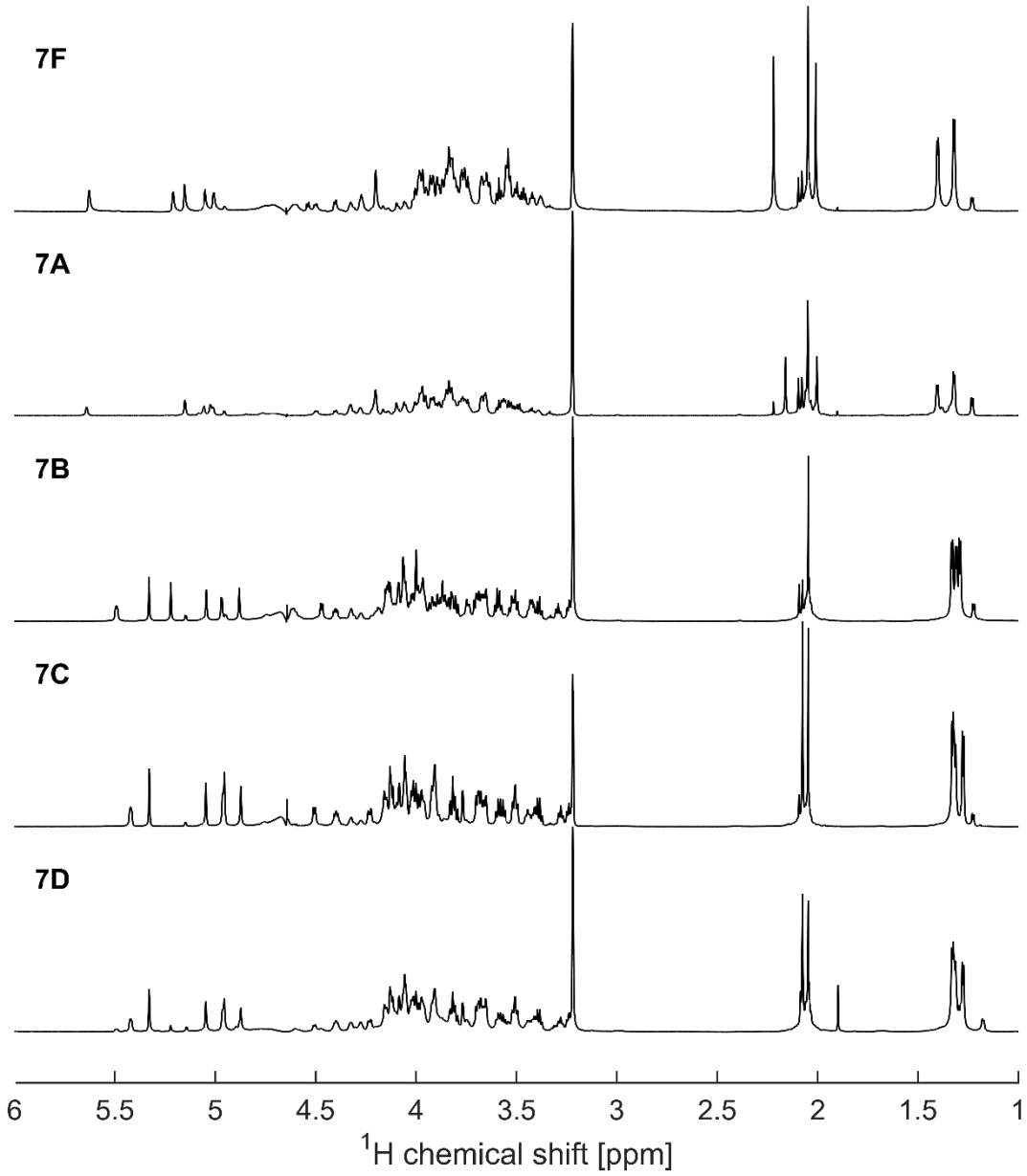


Figure 3 – ^{1}H spectra of the cps of serogroup 7. From top to bottom: 7F, 7A, 7B, 7C and 7D.

The ^{13}C-NMR signals of the anomeric positions were all assigned using the $^1J_{^1H,^{13}C}$ connection observed using HSQC, and the anomeric configurations were determined using both $^3J_{^1H,^1H}$, obtained primarily using DQF-COSY, and $^1J_{^1H,^{13}C}$ couplings, obtained using CLIP-HSQC[28]. A summary of the chemical shifts of the seven residues can be seen in table 1 and a summary of the coupling constants can be found in table 2. The spin systems were identified mainly through the use of DQF-COSY, TOCSY and HSQC-TOCSY, and the glycosidic linkages between the carbohydrate units were determined mainly by NOESY and HMBC, see table 3 for a summary, and were supported by chemical shift analysis.

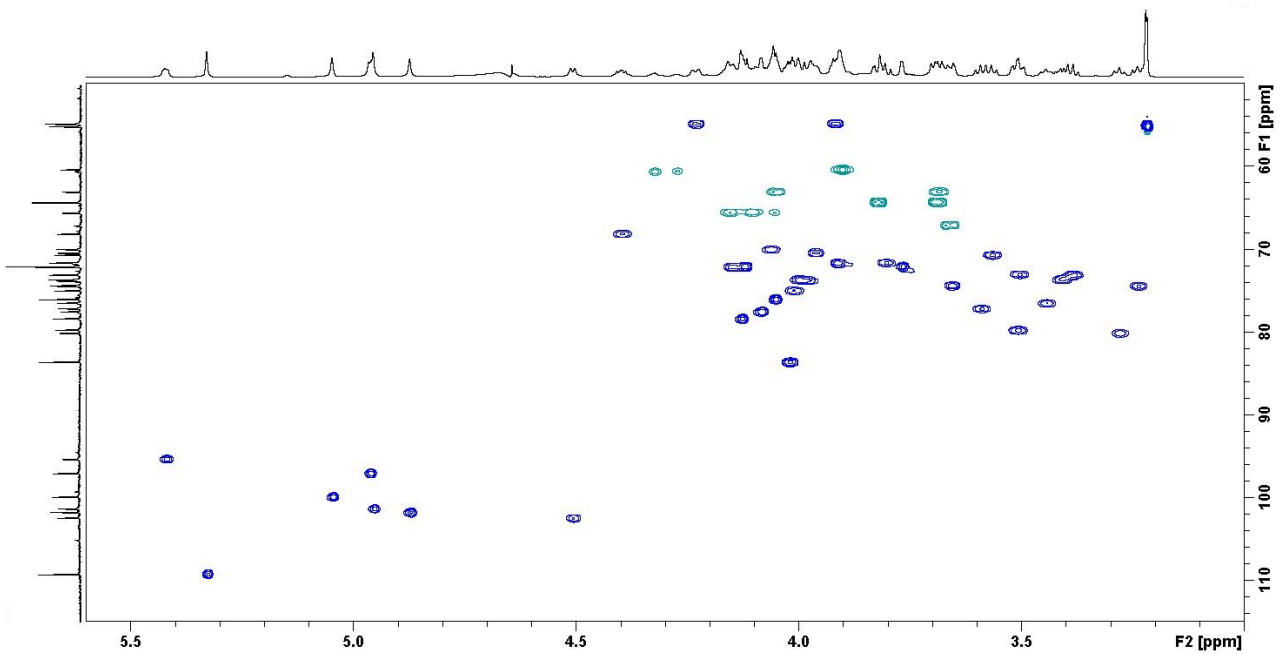


Figure 4 – HSQC spectrum of the cps of 7C. The projections are from separate 7C ^{1}H and ^{13}C experiments.

The seven monosaccharide units were named **A**-**G** in order of descending ^{13}C chemical shift of their anomeric position. A single furanoside, **A**, and six pyranosides, **B**-**G**, were identified. The anomeric configuration of three of the six pyranosides could be identified from the $^{3}J^{1}_{H1,^{1}H2}$ coupling constants as two with α-configuration with coupling constants of approximately 3 and 4 Hz for **F** and **G**, respectively, and one with β-configuration with a coupling constant of 8 Hz for **B**. As the three other pyranosides, **C**-**E**, had $^{3}J^{1}_{H1,^{1}H2}$ coupling constants of approximately 6 Hz, $^{3}J^{1}_{H2,^{1}H3}$ coupling constants of approximately 4 Hz and $^{3}J^{1}_{H3,^{1}H4}$ coupling constants of approximately 10 Hz it was evident the protons on the 2-positions were equatorial for all three and $^{1}J^{1}_{H,^{1}C}$ coupling constants[38] were required to determine the anomeric configuration. These were found to be 171, 170 and 159 Hz for **D**, **E** and **C**, respectively, corresponding to α-configuration for **D** and **E** and β-configuration for **C**.

The three pyranosides, **C**, **D** and **E**, with similar coupling constants were all identified as rhamnoses, which also explains the three C-Me signals mentioned earlier. Both the coupling constants, as mentioned above, and their chemical shifts were used for this determination. They were all three found to have a single substitution of another carbohydrate, two of them on their 2-position, **C** and **E**, and one on its 4-position, **D**. This was evident both by the downfield ^{13}C chemical shifts of the substituted position compared to unsubstituted methyl rhamnopyranosides[40] and by their HMBC and NOESY connectivities, see table 3.

The three other pyranosides, **B**, **F** and **G**, were identified as β-glucose, α-*N*-acetylglucosamine (GlcNAc), and α-*N*-acetylgalactosamine (GalNAc), respectively. It was evident that **G** had a phosphodiester on its 1-position as it showed a $^{3}J_{^{1}H,^{31}P}$ coupling and a $^{2}J_{^{13}C,^{31}P}$ coupling, both of approximately 6 Hz, measurable in 1D ^{1}H and ^{13}C spectra. Furthermore, using ^{1}H-^{31}P HMBC it was possible to identify four proton signals which correlated to the phosphodiester, the 1- and 2-position of the GalpNAc **G** and the 6-position of the GlcpNAc **F,** as well as possibly the 5-position of **F** as it overlaps with the 6-position.

The monosaccharide unit **G** was identified as a GalpNAc based on the $^{3}J_{^{1}H,^{1}H}$ from H2-H3, showing both protons were axial, the lack of any $^{3}J_{^{1}H,^{1}H}$ connectivity through the 4-position, which is often seen for galactopyranosides, and the chemical shifts. In particular the upfield ^{13}C chemical shift of the 2-position, supported by HMBC connectivity from H2 to the carbonyl of the *N*-acetyl group. Furthermore, the GalpNAc unit **G** was determined to have two substitutions, on the 3- and the 4-position, both based on the downfield ^{13}C chemical shifts of the substituted positions compared to unsubstituted methyl α-GalpNAc[40] and on NOESY and HMBC correlations. The methyl group of the *N*-acetyl group was assigned by HMBC correlation to the carbonyl which also had a HMBC connection to the 2-position.

The GlcpNAc unit **F** also had a connection to a phosphodiester from its 6-position as its only substitution, and being the only other part of the molecule with a linkage to a phosphodiester this was assumed to be the other end of the phosphodiester connected to GalpNAc unit G, which was further supported by ^{1}H-^{31}P HMBC. The unit was identified as a GlcpNAc based on its $^{3}J_{^{1}H,^{1}H}$ coupling constants being large from H2-H3, H3-H4 and H4-H5, meaning it had a glucose configuration, and the 2-position had a ^{13}C chemical shift of 54.8 ppm, which fits nicely with where a GlcpNAc 2-position would be expected[40]. Furthermore, there was an HMBC connection from the H2 to the carbonyl of the *N*-acetyl, which also had an HMBC correlation with the methyl group of the acetyl.

The glucopyranoside **B** was found to be substituted on its 4-position and was determined to be a glucose based mainly on its large $^{3}J_{^{1}H,^{1}H}$ coupling constants between all ring protons as well as the chemical shifts, where only the 4-position had a large downfield shift in ^{13}C chemical shift compared to methyl β-glucopyranoside[40].

The last carbohydrate unit, the furanoside **A**, was determined to be an unsubstituted ribofuranoside based mainly on its chemical shifts compared to methyl ribofuranoside, where it has quite similar chemical shifts to methyl β-ribofuranoside[40]. Compared to the chemical shifts of the other 1-O methylated pentofuranosides, arabinose, ribose, xylose and lyxose, squaring the differences

between each position of the furanoside **A** and the corresponding position in each of the methylated pentofuranosides, results in the β-ribofuranoside being the best match with over an order of magnitude to the remaining[40]. Furthermore, the chemical shifts of the β-Ribf are almost identical to those of the corresponding β-Ribf in 7B[84].

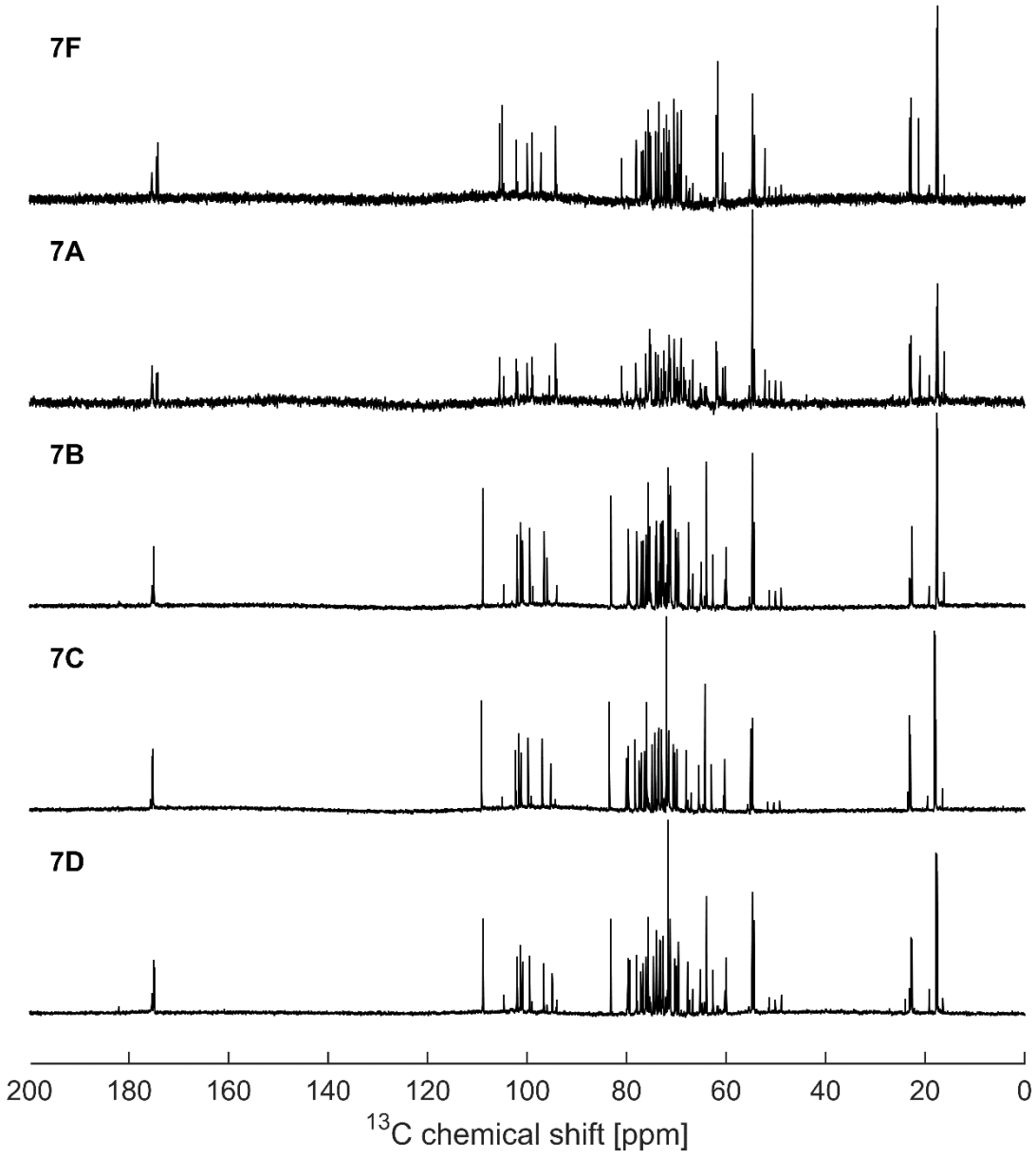


Figure 5 – ^{13}C spectra of the cps of serogroup 7. From top to bottom: 7F, 7A, 7B, 7C and 7D.

Table 1 – Chemical shifts for 7C, 7B and 7D table. 7D is a mixture of 7C and 7B in a ratio of 5:1.

7C table, intensity 5	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Ribf-(1-	5.330 108.84	4.054 75.63	4.124 71.64	4.024 83.17	3.691/3.824 63.77	N/A	N/A	N/A
-4)-α-Rhap-(1-	4.956 100.9	3.771 71.39	3.916 71.24	3.510 79.31	4.401 67.71	1.276 17.60	N/A	N/A
O-P-6)-α-GlcpNAc-(1-	4.964 96.63	3.921 54.39	3.808 71.18	3.571 70.25	4.154 71.72	4.113/4.154 65.17	2.049 22.60	- 174.99
-2)-α-Rhap-(1-	5.049 99.48	4.089 77.10	3.967 70.01	3.508 72.62	4.066 69.56	1.324 17.54	N/A	N/A
-2)-β-Rhap-(1-	4.874 101.33	4.131 77.95	3.659 73.93	3.39 72.69	3.414 73.19	1.317 17.77	N/A	N/A
-4)-β-Glcp-(1-	4.509 102.01	3.240 74.02	3.594 76.71	3.283 79.69	3.447 76.06	3.688/4.054 62.63	N/A	N/A
-3,4)-α-GalpNAc-(1-O-P-	5.422 94.94	4.234 54.45	4.014 74.56	3.976 73.37	4.002 73.25	3.909 59.94	2.076 22.78	- 174.87
7B table, intensity 1	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Ribf-(1-	5.330 108.85	4.062 75.61	4.146 71.61	4.022 83.10	3.699/3.837 63.89	N/A	N/A	N/A
-4)-α-Rhap-(1-	5.222 100.89	3.999 71.17	3.989 71.28	3.524 79.64	4.401 67.50	1.290 17.61	N/A	N/A
O-P-6)-α-GlcpNAc-(1-	4.968 96.56	3.933 54.35	3.804 71.11	3.596 70.13	4.145 71.67	4.188/2.048 64.99	2.048 22.56	- 174.96
-2)-α-Rhap-(1-	5.046 99.42	4.087 76.98	3.967 69.96	3.506 72.53	4.062 69.53	1.331 17.47	N/A	N/A
-2)-β-Rhap-(1-	4.881 101.24	4.130 77.87	3.657 73.90	3.385 72.64	3.421 73.14	1.330 17.54	N/A	N/A
-4)-β-Glcp-(1-	4.473 102.00	3.237 74.05	3.586 76.59	3.292 79.54	3.433 76.01	3.687/4.066 62.61	N/A	N/A
-3,4)-α-Glcp-(1-O-P-	5.493 95.95	3.746 73.26	4.003 75.29	3.869 72.81	3.967 72.86	3.893 59.90	N/A	N/A

Table 2 - 7C coupling constants. N/A is used to indicate when a part of the table does not apply for this carbohydrate unit and n.d. for when it was not possible to determine the coupling, usually due to high overlap in the DQF-COSY spectrum. All $^{1}J^{1}_{H,^{13}C}$ coupling constants were determined using CLIP-HSQC and all $^{3}J^{1}_{H,^{1}H}$ coupling constants determined using DQF-COSY.

7C couplings furanoside	**H1-C1**	**H1-H2**	**H2-H3**	**H3-H4**	**H4-H5**	**H4-H5'**	**H5-H5'**	N/A
β-Ribf-(1- **A**	177 Hz	3.1 Hz	5.2 Hz	7.9 Hz	7.4 Hz	3.8 Hz	12.2 Hz	N/A
7C coupling pyranosides	**H1-C1**	**H1-H2**	**H2-H3**	**H3-H4**	**H4-H5**	**H5-H6**	**H5-H6'**	**H6-H6'**
-4)-α-Rhap-(1- **D**	171 Hz	6.3 Hz	4.3 Hz	10.0 Hz	10.1 Hz	6.5 Hz	N/A	N/A
-P-O-6)-α-GlcpNAc-(1- **F**	171 Hz	4.3 Hz	11.2 Hz	8.3 Hz	10.1 Hz	n.d.	n.d	n.d
-2)-α-Rhap-(1- **E**	170 Hz	6.2 Hz	4.4 Hz	10.5 Hz	10.2 Hz	6.5 Hz	N/A	N/A
-2)-β-Rhap-(1- **C**	160 Hz	6.2 Hz	4.1 Hz	9.6 Hz	n.d.	5.9 Hz	N/A	N/A
-4)-β-Glcp-(1- **B**	162 Hz	7.7 Hz	11.1 Hz	7.5 Hz	12.0 Hz	7.8 Hz	5.0 Hz	11.9 Hz
-3,4)-α-GalpNAc-(1-O-P- **G**	178 Hz	2.7 Hz	10.2 Hz	n.d.	n.d.	6.7 Hz	N/A	N/A

After having assigned all the chemical shifts of the carbohydrate units, the next challenge was to assemble the fragments. As mentioned above, this was mainly done using HMBC and NOESY correlations, and a summary of the strong HMBC and NOESY correlations from all the anomeric protons can be found in table 3. The β-Ribf unit **A** was found to be attached to the 4-position of the α-Rhap **D**, as the only strong HMBC and NOESY connections to another monosaccharide from the 1-position in **A** is to the 4-position of **D**. The anomeric position of the α-Rhap **D** was found to be connected the 3-position of the α-GalpNAc **G**, as the only strong HMBC and NOESY connections to another monosaccharide was to this position. As the α-GalpNAc **G** is the only disubstituted of the seven monosaccharides, and the β-Ribf **A** the only terminal monosaccharide, there are no further branching points. The α-GalpNAc **G** is also the only monosaccharide with an anomeric linkage to the phosphodiester, which, as mentioned earlier, is also connected to the 6-position of the α-GlcpNAc **F**.

Table 3 - Summary of strong HMBC and NOE correlations from the anomeric positions of the cps of 7C.

Anomeric position	**Strong HMBC**	**Strong NOE**
β-Ribf **A**	A2, A3, A4, D4	D4
α-Rhap **D**	D2/D3, D5, G3	G3
α-GlcpNAc **F**	F3, F5, E2	E1, E2
α-Rhap **E**	E3, E5, C2	C2, F1
β-Rhap **C**	C2, B4	C3, C5, B4
β-Glcp **B**	G4	B5, G4, G5
α-GalpNAc **G**	G3, G5	none

The α-GlcpNAc is anomerically connected to the 2-position of the α-rhamnopyranoside **E**, as it is the only strong HMBC correlation as well as the 1- and 2-positions of **E** being the only strong NOESY correlations to protons of another unit. The α-Rhap **E** is connected to the 2-position of the β-Rhap **C**, as it is the only position from another monosaccharide unit that it has strong correlations to in HMBC and NOESY.

The β-Rhap **C** has an anomeric linkage to the 4-position of the β-Glcp **B**, as this is the only strong HMBC and NOESY correlations to a position not from its own carbohydrate. And lastly, the β-Glcp has an anomeric connection to the 4-position of the α-GalpNAc **G**, as this is the only strong HMBC correlation, and the 4- and 5-positions of **G** are the only strong NOESY correlations it has to anything not in its own monosaccharide.

These connectivities produces the structure as presented in section 2.1.2.5 and scheme 1. For this scheme, and because it was reported for 7B[84], it was assumed that the monosaccharides, with the exception of the rhamnoses, had D-configuration. As mentioned in section 2.1.2.5, the structure of 7C and 7B are very similar, with the only differences being the branched α-Glcp in 7B being replaced by an α-GalpNAc in 7C.

2.1.3.2 Serotype 7D capsular polysaccharide structure

The structure of 7D was found to be a mixture of 7B and 7C, with a 1:5 ratio between 7B/7C. Whether both polysaccharides are there, or if the two repeating units are interconnected, is currently not known. The ratio of 1:5 between the 7B/7C cps is supported by ^{1}H NMR (figure 8), ^{31}P NMR (figure 7) and HSQC (figure 6) data as well as with the agglutination test performed by our collaborators, described in paper 1.

In figure 3 the ^{1}H NMR spectra of serogroup 7 cps are compared, with a zoom of the anomeric regions of 7B, 7C and 7D shown in figure 8, and similarly in figure 5 the ^{13}C NMR spectra of the cps can be found. In these figures, it is evident that the peaks unique to the 7B repeating unit in 7D are much less intense than the peaks unique to the 7C repeating unit. In figure 6 a similar comparison is made with an overlay of HSQC spectra of the 7B, 7C and 7D and the same trend can be observed. It should be noted that some of the peaks originating from the CWPS[86] present in all three samples have shifted slightly in the 7D sample due to a higher salt concentration from the purification process. In table 1 the chemical shift assignment for the 7D cps can be found, and as it is a combination of the cps of 7B and 7C, the table is also a combination of both.

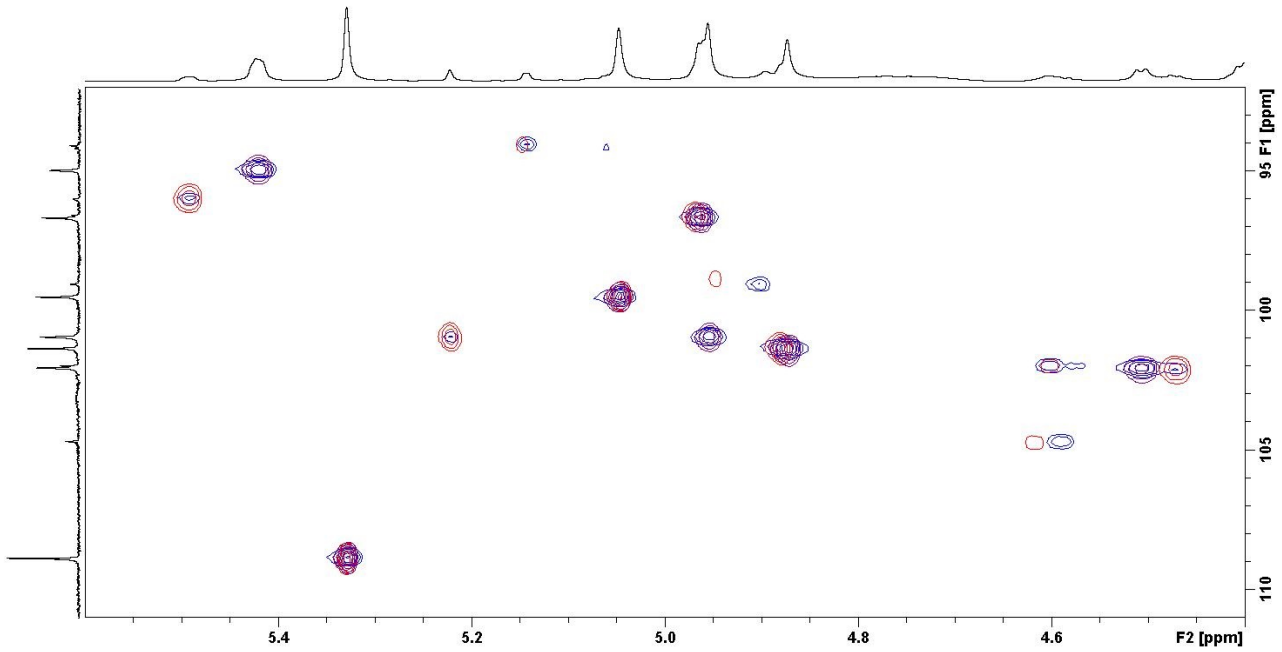


Figure 6 - HSQC spectrum overlay with 7B in red, 7C in purple and 7D in blue. Note that 7D is a mixture between 7B and 7C with a 1:5 ratio. The projections are from separate 7D ^{1}H and ^{13}C experiments.

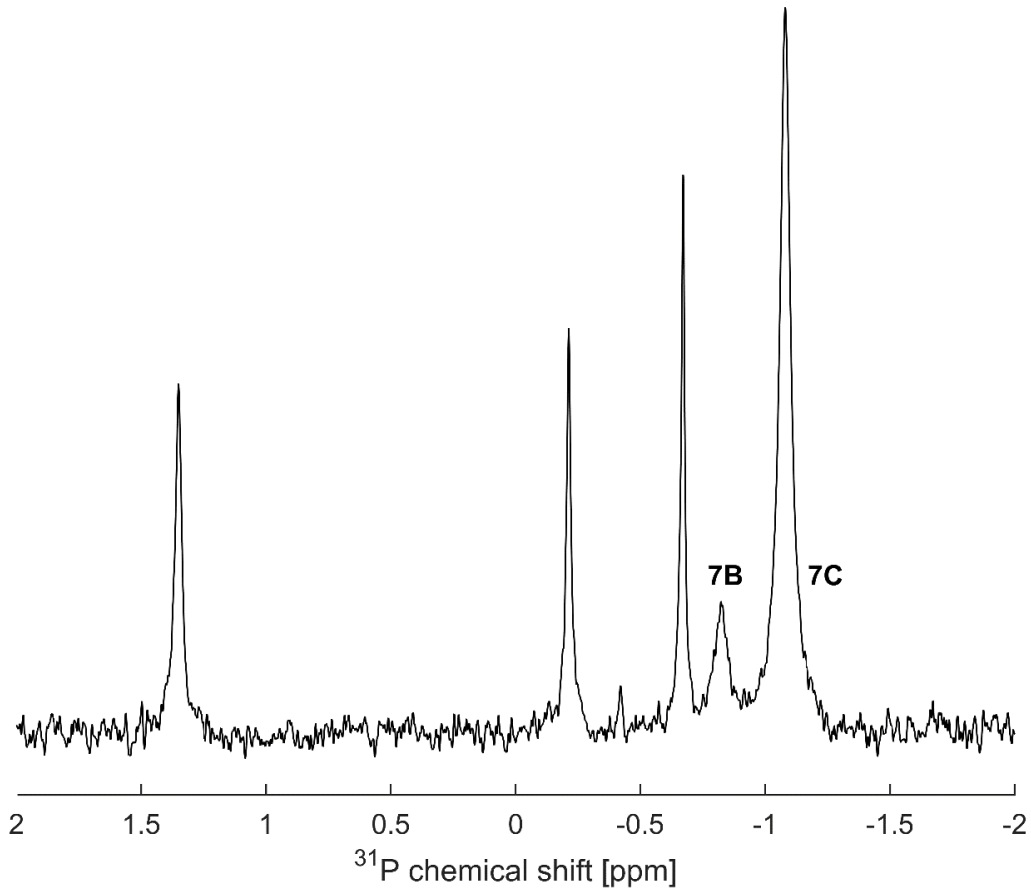


Figure 7 – ^{31}P NMR spectrum of sample 7D. The peaks from the phosphodiesters of the 7B and 7C repeating units are labelled 7B and 7C. The remaining peaks are from the CWPS.

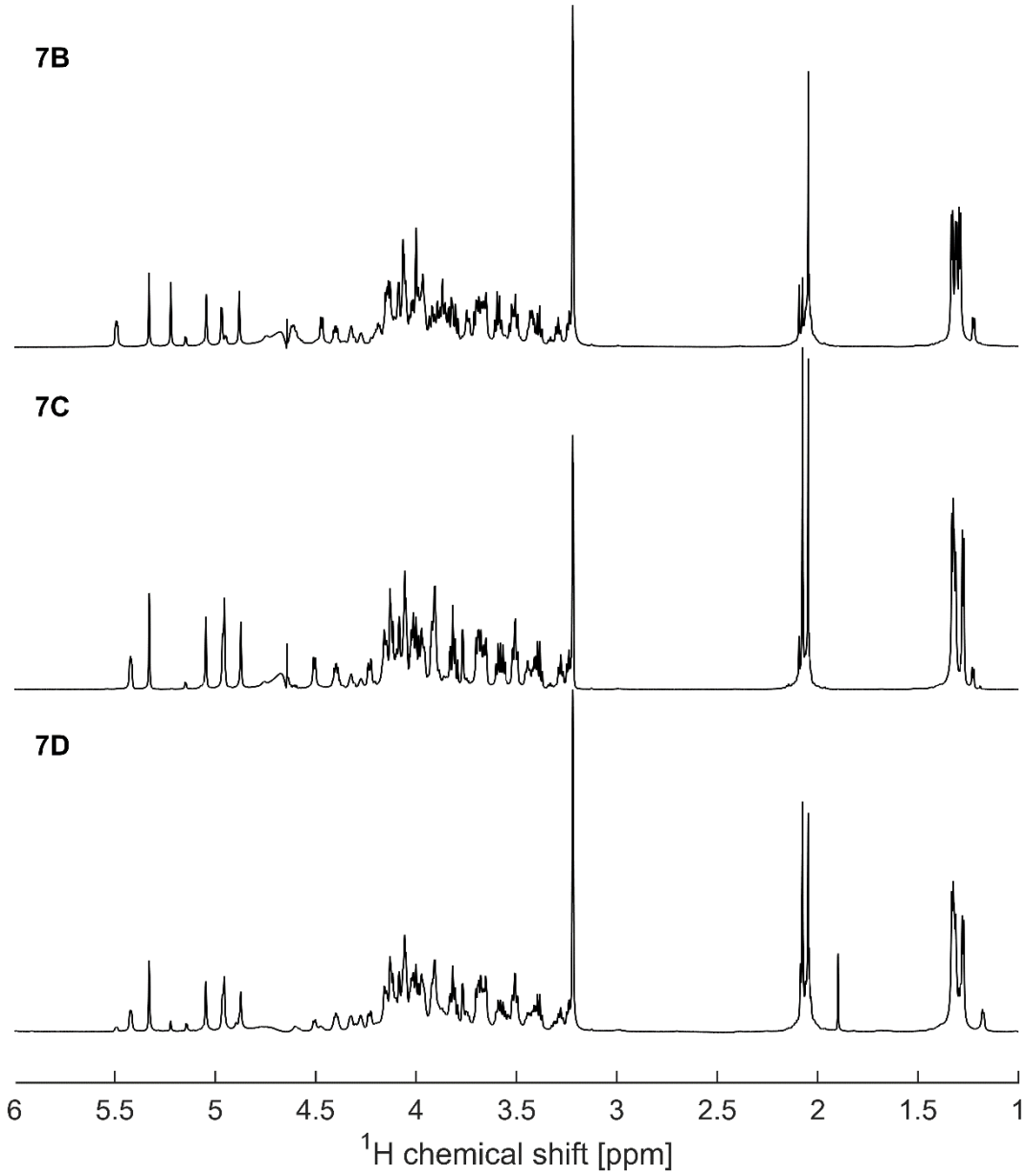


Figure 8 - 1H, spectrum zoomed on the anomeric region, comparison between 7B (top), 7C (middle) and 7D (bottom). Note that 7D is a mixture between 7B and 7C with a 1:5 ratio.

Other examples of hybrid serotypes have been reported[90–93], and similar cases to 7D have been reported in serogroup 6, an example of which could be 6F and 6G[92]. The capsular polysaccharide of 6F turned out to be a mixture of the capsular polysaccharides from 6A and 6C in a 3:1 ratio, while for 6G it was a 2:3 ratio of 6B/6D. The differences between capsular polysaccharides of 6A/6C and 6B/6D is a change from an α-Galp to an α-Glcp, and the two new serotypes 6F and 6G became combinations of these by having a mutations in the glycosyltransferases from 6A/6B which

adds UDP-Galp. In said enzyme, 6F had a single mutation on residue 150 from alanine to threonine, whereas 6G had two mutations, the same mutation of residue 150 as in 6F, as well as of residue 38 from aspartic acid to asparagine. These mutations resulted in conversions from mono-specific to bi-specific glycosyltransferases that could transfer both UDP-Gal and UDP-Glc, but at different ratios.

By comparison, using PneumoCaT our collaborators at Public Health England found that 7D was genetically similar to 7B, except for a novel codon on the capsular polysaccharide loci resulting in an F385L mutation in the glycosyltransferase. For comparison, in 7C this glycosyltransferase has a cysteine in position 385[81]. It would seem reasonable that the mutation in the 7B *cps* loci produced a bi-specific glycosyltransfaerase capable of using both UDP-Glc and UDP-GalNAc as donors, and consequently capable of making both the 7B and 7C repeating unit.

What exactly the hybrid nature of these serotypes capsular polysaccharides with multiple repeating units means for the virulence is not yet clear. However, as the capsular polysaccharide shields the pathogen from the host immune system, even small changes can moderate the pathogens virulence.

2.1.4 Conclusion

As the capsular polysaccharide is largely responsible for the virulence of pneumococcus[68], which in turn is responsible for numerous fatalities annually[64], knowledge of the structure is important to understand how the pathogen evolves and adapts.

The cps for 7C was elucidated, and it turned out to be very similar to the structure of 7B, see structures in section 2.1.2, with only a single monosaccharide in the repeating unit being different. Comparing 7B/7C to 7F/7A, there are two structural similarities: The α-GlcpNAc-(1→2)-α-Rhap and the β-Rhap-(1→4)-β-Glcp, although the former part is present as part of the backbone in 7B/7C and as a sidechain in 7F/7A, and the latter part contains a 2-*O*-acetylation in 7F/7A.

The cps of the novel serotype 7D was determined to be a hybrid of 7B and 7C, and contained the repeating units from both 7C and 7B in a 5:1 ratio. This suited quite well with the results obtained at SSI Diagnostica, discussed further in paper 1, in which it was observed that 7D reacted with antisera from rabbits immunised with either 7B, 7C or 7D. Furthermore, sera from rabbits immunised with 7D also reacted with 7B and 7C, further supporting the hybrid nature of the novel serotype. Additionally, it was also observed that 7D acted more like 7C than 7B, despite being almost genetically identical to 7B, also supporting the weighted difference between the two capsular polysaccharide repeating unit structures in 7D.

While no covalent linkage between the 7B and 7C repeating unit was observed, it would seem most reasonable that they are. The alternative would be that both cps were present but not connected, and that the bi-specific enzyme would have selectivity based on the previous repeat.

Even though tools are being developed to help better survey and understand the spread and evolution of pneumococcus[81], the genetics alone does not necessarily give all the desired information of the capsular polysaccharide, as different strains can produce the same capsular polysaccharide[73,74] or even multiple polysaccharides[90,92,93], which was also the case with 7D, highlighting the need to investigate the structures to understand the genetics.

2.1.5 Experimental

2.1.5.1 Purification performed by Sofie Slott at SSI Diagnostica

The 7D pneumococcus was grown in 5 L serum bouillon for 4 hours at 37 °C. The cells were harvested by centrifugation and lysed using pH 7.9 Tris buffer containing 0.1% deoxycholate. The solution was adjusted to pH 5 using acetic acid and the precipitate, deoxycholate and cell debris, was removed by centrifugation. Proteins were precipitated by adding $CaCl_2$ to reach 0.1% concentration followed by ethanol to reach 25% concentration and removed by filtration. Polysaccharides were precipitated by increasing the concentration of ethanol to 80% and recovered by centrifugation. Subsequently, the precipitate were dissolved in pH 5 acetate buffer (50 mM) and DNAse and RNAse were added to the solution. The pH was adjusted to 7.47 using NaOH and trypsin was added to decompose any remaining protein residues. Finally, the polysaccharides were freeze-dried after diafiltration. Following the purification, the samples were dissolved in D_2O and used for NMR.

The remaining serogroup 7 cps are commercially available (SSI Diagnostica), and are produced in a similar fashion.

2.1.5.2 NMR spectroscopy

For the NMR samples approximately 10 mg purified cps was dissolved in 500 µL D2O.

The NMR spectrometers used for this project were a Bruker 800 MHz with a cryoprobe, a Bruker 400 MHz with Prodigy probe and a Bruker 400 MHz. All experiments were recorded at 40 °C, with the exception of the CLIP-HSQC which was measured at 25 °C. Chemical shifts are reported in ppm using acetone (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively[16]) as internal reference. All NMR experiments, with the exception of CLIP-HSQC, were performed using Bruker standard pulse sequences. The spectra were acquired using TopSpin 2.1, with the exception of the CLIP-HSQC which was acquired using TopSpin 3.2, and processed using TopSpin 3.5. The

supplementary material to paper 1 contains all the spectra used for the cps structural elucidation of 7C and 7D as well as HSQC spectra of 7F, 7A and 7B as examples.

The following spectra were all recorded on a Bruker 800 MHz Avance spectrometer equipped with a 5 mm TCI ^{1}H/(^{13}C,^{15}N) cryoprobe: 1D ^{1}H-NMR spectra were acquired with water presaturation and 32768 data points zero filled to 65536 data points and a sweep width of 6393.9 Hz. 1D ^{13}C-NMR spectra were acquired with power-gated decoupling of ^{1}H, 32768 points zero filled to 65536 and a sweep width of 48076.9 Hz. DQF-COSY[94] was recorded with 4096x512 points zero filled to 8192x1024 and sweep widths of 6393.9 Hz in F2 and 6390.4 Hz in F1. NOESY[23] was recorded with presaturation of the water signal with 4096x512 points zero filled to 8192x1024 with a mixing time of 200 ms and sweep widths of 6393.9 Hz in F2 and 6390.4 Hz in F1. TOCSY[95,96] with 31887.8 Hz spinlock power using the MLEV17 spinlock sequence for mixing was recorded with 4096x512 points zero filled to 8192x1024 with a mixing time of 60 ms and sweep widths of 6393.9 Hz in F2 and 6390.4 Hz in F1. Multiplicity edited HSQC[25,27] was recorded with 2048x512 points and zero filled to 4096x1024 and sweep widths of 6393.9 Hz in F2 and 35354.2 Hz in F1. HSQC-TOCSY[29] with 31887.8 Hz spinlock power was recorded using DIPSI2 spinlock sequence with 1024x512 points zero filled to 4096x1024, a mixing time of 60 ms and sweep widths of 6393.9 Hz in F2 and 35354.2 Hz in F1. Long range optimised HMBC[31] was recorded with 4096x512 points zero filled to 8192x1024 and sweep widths of 6393.9 Hz in F2 and 44192.7 Hz in F1. The HMBC was optimised for 8 Hz couplings.

The following spectra were recorded on a Bruker 400 MHz Avance III spectrometer with a 5 mm Smartprobe: 1D ^{31}P was recorded with 65536 points zero filled to 131072 points and a sweep width of 12931.0 Hz. ^{1}H-^{31}P HMBC optimised for 8 Hz couplings was recorded with 4096x512 points zero filled to 8192x1024 and sweep widths of 2406.9 Hz in F2 and 1944.2 Hz in F1.

CLIP-HSQC[28] was measured on a Bruker 400 MHz Avance III HD spectrometer with a 5 mm prodigy cryoprobe with 2048x512 points zero filled to 4096x1024 and sweep widths of 4795.4 Hz in F2 and 16668.5 Hz in F2.

2.1.5.3 Additional experimental by collaborators

The experimental procedures carried out by our collaborators can be found in paper 1. This includes DNA extraction and sequencing, bioinformatic methods, assembly-based sequence analysis and serotyping.

2.2 Structucal characterisation of heteropolysaccharides from *Inonotus obliquus*

This project was performed in collaboration with Christian Winther Wold of the Kari Tvete Inngjerdingen group at the School of Pharmacy, University of Oslo. Christian extracted the different polysaccharides and visited our lab in the autumn of 2016. My part of this project was to assist with the NMR acquisition and analysis.

2.2.1 Introduction

The fungus *Inonotus obliquus*, more commonly known as the chaga mushroom, infects living broad-leaved trees as sterile conks of sclerotia[97,98]. It is native to northern latitudes[97] and has been used as a medicinal mushroom for centuries against various diseases including gastrointestinal disorders, cardiovascular diseases and cancer[11,99,100]. Compounds of medicinal interest from the fungus includes melanin, polyphenol, triterpenoids and polysaccharides[101], the latter of which is the focus of this study.

Some of the beneficial effects of the chaga mushroom are attributed, at least in part, to the polysaccharides found in the fungus[102–107]. However, not much is known of the structure of the polysaccharides from *Inonotus obliquus*, even though the structure is very important for any biological roles it might have. An example of a fungal polysaccharide used medicinally could be lentinan[9,108,109], extracted from *Lentinula edodes*, which has a β-glucan backbone of 1,3-β-Glcp-1,3 with β-Glcp-1,6 sidechains. The mechanism of action for medicinal β-glucan containing polysaccharides is binding to pattern recognition receptors[110], and as such the structure is very important for the activity[111].

2.2.2 Results and discussion

2.2.2.1 Summary of the purification and separation of polysaccharides performed by collaborators

In order to gain some structural information of the polysaccharides from chaga, our collaborators harvested a fresh specimen, split the black exterior from the brown interior, dried the samples and ground them into a fine powder. From these powders a number of samples were acquired by different methods of extraction and purification, as described in paper 2. In short, a total of eight samples were collected by different methods of extraction and purification of either the external or internal part of the fungus. From the external powder, two samples were obtained, one was extracted using neutral (IOE-WN) and the other with acidic conditions (IOE-WAc). From the internal powder, one sample was obtained by alkaline extraction (IOI-Al) as well as one from neutral extraction (IOI-WN) and one using acidic extraction (IOI-WAc). The acid extracted sample

was further separated into an additional three samples using size exclusion chromatography (IOI-WAcF1, IOI-WAcF2 and IOI-WAcF3 in that order), whereas the other samples did not separate. Following this, all the samples were lyophilised.

All of the samples were dissolvable in D_2O (5-10 mg), with the exception of IOI-AI, which could only be dissolved in basic conditions. Furthermore, all of the samples contained polysaccharide of heterogeneous nature and did not have a repeating unit. The following analysis describes observed trends and compares the data to the monosaccharide and linkage analysis performed by our collaborators, which can be found in paper 2.

2.2.2.2 Overview and comparison of the samples

The samples all had some trends in common, namely the signals arising from the anomeric positions highlighted with **A**, **B** and **C** in figure 9. Each of these three areas corresponds to several signals overlapping, for example: the signals at **A** contain 3-β-Glcp, 6-β-Glcp, and 3,6-β-Glcp attached to the 6-position of another β-Glcp. In addition, the peak at **A** also contains β-xylose signals. The signals at **B** are primarily β-Glcp attached to the 3-position of another β-Glcp and **C** is α-Galp and 3-*O*-Me α-Galp, both of which are linked to the 6-positions of other α-Galp. However, the relative intensities of these peaks all varied, as well as the minor signals in figure 9, as the samples all had different ratios between the monosaccharides and their linkages.

The spectrum shown in figure 9 is from the IOE-WN sample, and the IOI-WN sample was mostly quite similar to this sample, see supplementary to paper 2. However, all of the acid extracted samples had a high concentration of galacturonic acid and markedly less galactose. Lastly, the sample extracted using basic conditions contained mostly glucose, was only dissolvable in basic conditions and gave poor spectra. The basic extracted sample was the only sample not measured at DTU, but was analysed by our collaborators instead.

Assignments of the major monosaccharides present is shown in table 4, but due to the heterogeneous nature of the polysaccharides, all these values should be regarded as averages and some variation is expected between samples.

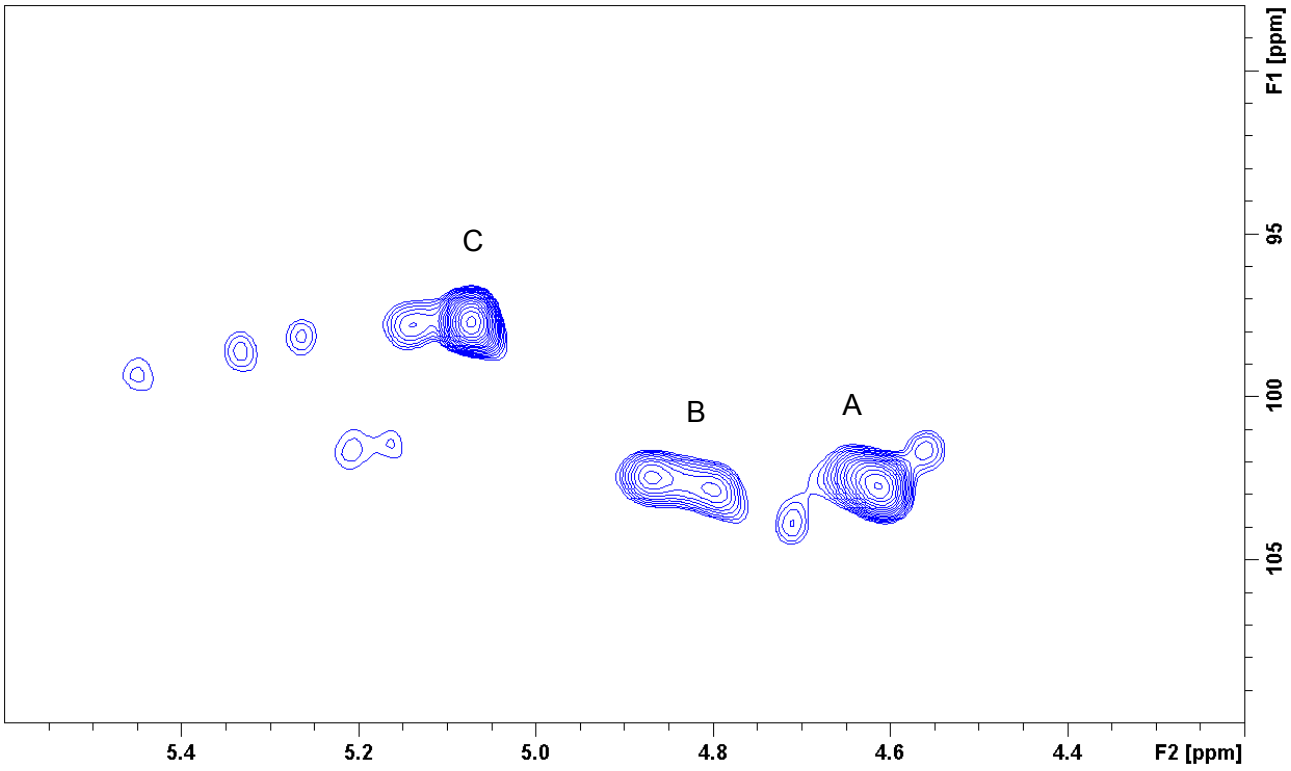


*Figure 9 – HSQC spectrum of the IOE-WN sample zoomed to the anomeric region with signal clusters of **A**, **B** and **C** highlighted (see the text for further clarification).*

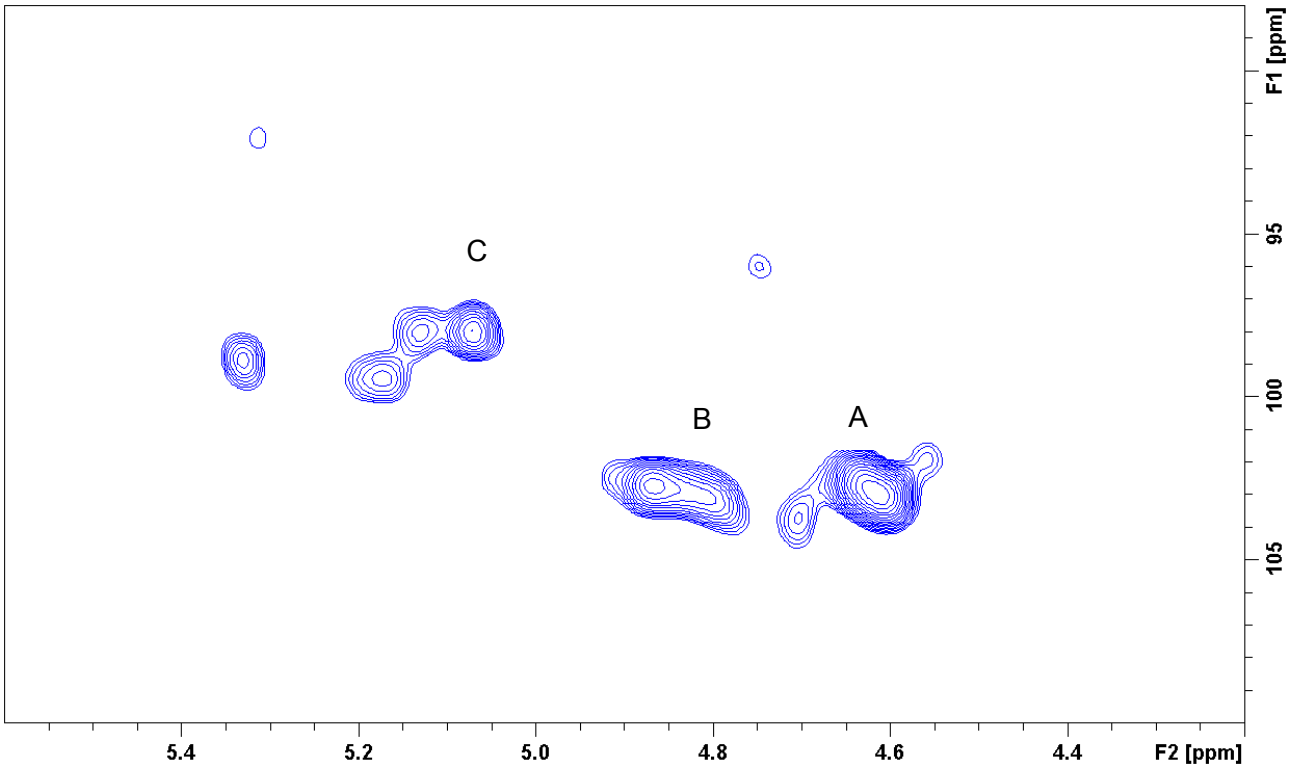


*Figure 10 – HSQC spectrum of the IOI-WAcF2 sample zoomed to the anomeric region with signal clusters of **A**, **B** and **C** highlighted (see the text for further clarification).*

Table 4 – NMR assignments of the major signals in the water-extracted polysaccharides. The shifts can vary slightly between samples. Due to spectral overlap or weak intensity, some species are only partially assigned.

Type of monosaccharide w/linkage		1	2	3	4	5	6	3-*O*-Me
(1→3)-β-Glc-(1→3)								
	δ^{13}C	102.48	73.17	84.59	68.17	75.44	60.80	
	δ^{1}H	4.87	3.65	3.85	3.61	3.6	3.83/4.03	
1→6)-β-Glc-(1→3)								
	δ^{13}C	102.92	73.35	75.42	69.62	74.59	68.85	
	δ^{1}H	4.80	3.47	3.62	3.51	3.80	3.96/4.31	
(1→6)-β-Glc-(1→6)								
	δ^{13}C	102.72	73.05	75.62	69.52	74.82	68.85	
	δ^{1}H	4.61	3.42	3.59	3.55	3.72	3.96/4.31	
(1→3)-β-Glc-(1→6)								
	δ^{13}C	102.40	72.58	84.79	68.11	75.44	60.80	
	δ^{1}H	4.63	3.61	3.84	3.60	3.58	3.83/4.03	
T-β-Glc-(1→6)								
	δ^{13}C	102.72	73.05	75.62	69.52	74.82	60.80	
	δ^{1}H	4.61	3.42	3.59	3.55	3.72	3.83/4.03	
(1→4)-β-Xyl								
	δ^{13}C	101.87	74.21			62.9		
	δ^{1}H	4.55	3.34	3.66	3.82	3.46/4.18		
T-β-Xyl								
	δ^{13}C	101.69				65.16		
	δ^{1}H	4.56	3.38			3.4/4.06		
(1→6)-α-Gal-(1→6)								
	δ^{13}C	97.72	67.95	69.83	69.23	69.11	66.76	
	δ^{1}H	5.08	3.93	3.97	4.11	4.26	3.78/4.01	
(1→6)-α-3-O-Me-Gal-(1→6)								
	δ^{13}C	97.72	67.42	78.77	65.1	68.77	66.76	56.02
	δ^{1}H	5.07	3.97	3.63	4.38	4.27	3.78/4.01	3.53
(1→4)-α-GalA-(1→4)								
	δ^{13}C	98.92	68.02	68.62	77.72	71.12	174.70	
	δ^{1}H	5.17	3.85	4.09	4.52	4.85		
(1→3)-α-Man								
	δ^{13}C	102.1	70.38	78.43			62.84	
	δ^{1}H	5.21	3.86	4.12			3.82/3.98	
(1→4)-α-GlcA								
	δ^{13}C	98.9						
	δ^{1}H	5.34		4.21				
(1→2)-α-Rha								
	δ^{13}C	102.3					16.93	
	δ^{1}H	5.09	4.18		3.5	3.9	1.33	

2.2.2.3 Glucose species

As all of the samples contained glucose as the major component, this also gave rise the some of the most intense signals in all the spectra. In figure 9 and figure 10, the signals labelled **A** and **B** are mostly from glucose, and as mentioned above, **A** was from the anomeric position in β-Glcp-1,6 and **B** from β-Glc-1,3. Both of these observations are supported by HMBC and NOESY data, see supplementary material for paper 2, and the assignments were supported by CASPER[34–36] calculations and literature[112,113]. Other species of glucose might also exist in these peaks, but are too low-populated or simply overlap too severely to be distinguished.

In **A**, the signals from the different species overlap quite extensively, and distinguishing the different species from this peak alone becomes quite difficult, whereas the signals in **B** clearly splits up. As it turned out, the downfield peak, in terms of proton chemical shift, in **B** originated from 3-β-Glcp-1,3 whereas the upfield peak was from 6-β-Glcp-1,3.

The samples contained four major species of β-Glcp, 6-β-Glcp-1,6, 3-β-Glcp-1,6, 6-β-Glcp-1,3 and 3-β-Glcp-1,3, as well as several minor species that were difficult to discern from the aforementioned. However, some of the signals from the HSQC spectra were assumed to originate from other species of glucose, such as disubstituted 3,6-β-Glcp. Specifically, a peak approximately 0.07 ppm downfield in the ^{1}H frequency, but at the same ^{13}C frequency, from the 4-position of 3-β-Glcp was assigned to said species.

Several β-Glcp species were discernible in the spectra, but due to severe overlap some were unassignable. Table 4 summarises all the fully assigned glucose species. No correlations from any glucose species to any non-glucose were detectable.

2.2.2.4 Galactose species

The signals from the galactose present in the polysaccharide were very homogenous, being almost exclusively from two apparent species: 6-α-Galp-1,6 and 3-*O*-Me 6-α-Galp-1,6, the anomeric signals of which are both in peak C of figure 9 and figure 10. The methyl group had strong HMBC correlation to the 3-position and was thus assigned to be attached to the 3-position. This was supported by NOESY. Even though it was not possible to observe correlation between H3 and the Me as they were too close in chemical shift, there was a clear NOESY correlation to H4. Depending on the sample, the degree of methylation varied in the samples, but in most of them it was approximately one third.

Both the species were determined to have galactose configuration as the H4 was determined to be equatorial due to weak couplings with adjacent protons. The only unexpected correlations in

HMBC and NOESY observed from the anomeric positions of the two galactose species were to 6-positions of galactose, and no signals corresponding to any other substitution pattern were evident from the spectra. This finding was consistent with the linkage analysis performed by our collaborators, in which only small amounts of 2-, 3- or 4-substitution of galactose were observed in the samples.

2.2.2.5 Xylose species

Xylose was present in all the samples, in amounts varying from around 5% up to 20%, and thus in some samples it was quite difficult to discern the signals. Furthermore, unlike glucose and galactose, the xylose seemed to have more variety across the samples, suggesting different chemical environments. For example, in IOE-WN only one 4-substituted β-Xylp was apparent, whereas in IOI-WN a second, but quite similar, 4-substituted β-Xylp was observable as another 5-position was observable in HSQC. It was possible to differentiate terminal β-Xylp from 4-substituted β-Xylp as the substitution caused the 5-position ^{13}C chemical shift to move downfield. No HMBC or NOESY correlations could be assigned to the anomeric position of the xylose, possibly due to the severe overlap with the glucose signals.

2.2.2.6 Galacturonic acid species

As mentioned earlier, galacturonic acid was primarily present in the acidic extracted samples, and similarly to the galactose, the signals from the α-GalpA were quite homogeneous. The galactose configuration was confirmed by the weak $^{3}J_{^{1}H,^{1}H}$ coupling constant from H4 to H3 and H5. As the anomeric position and the 4-position of the α-GalpA had HMBC correlations it was determined to primarily be connected as such, giving 4-α-GalpA-1,4-, which is supported by the GC-MS linkage data described in paper 2.

2.2.2.7 Other species

Smaller species observed in the GC-MS linkage analysis of the samples, such as arabinose, rhamnose, mannose, fucose and glucuronic acid, were not fully assignable, and in some cases they were not even possible to confirm the presence of through NMR spectroscopy. However, some observations regarding signals possibly originating from these species were made:

For example, while no rhamnose species could be fully assigned, some signals that most likely originated from an α-Rhap were observed, based on the chemical shifts of the positions which could be linked to the methyl of the 6-position, but not to any anomeric position. Compared to most of the other species discussed above, the signals expected to come from rhamnose species were even more varied in chemical shift, and several species might have been present in each sample. For example, in the IOI-WAcF3 sample the methyl groups at approximately 1.3-1.4 and 16.5 ppm

in ^{1}H and ^{13}C chemical shift, respectively, had COSY correlations to at least four different H5 signals around 3.8-4.0 ppm as well as TOCSY correlations to several protons around 3.5, 3.8, 4.0 and 4.2 ppm, which could be from the remainder of the α-Rhap protons. Furthermore, through HMBC several ^{13}C signals around 67, 70, 72 and 80 ppm could be correlated to the aforementioned methyl groups. However, due to spectral overlap and low concentration of each of these species, proper assignments were not possible.

2.2.3 Conclusion

While it was not possible to fully assign the polysaccharides, some general trends were observed between the samples: As glucose was the most abundant monosaccharide in the samples, it would make sense if it was the main component of the backbone. According to a Smith degradation analysis performed by our collaborators, it would seem that most of the 3-β-Glcp-1,3- stretches were no longer than five units long before a kink or a terminal glucose would occur. Additionally, the galactose, and in case of the acidic extracted samples also galacturonic acid, were primarily bound to other galactose or galacturonic acid, respectively, with few branches or kinks. Lastly, xyloses, rhamnoses and other less populated species were likely spread across the polysaccharide in different environments, which would explain their weak and more diverse signals.

Furthermore, although no NOE or HMBC correlations between different species were observed, it is unlikely that they would not be part of the same heteropolysaccharide rather than being a mix of more heterogeneous polysaccharides that were copurified. The lacking of these correlations is likely due to the relative rarity within the polysaccharides, meaning they would likely either disappear in the noise or simply be indistinguishable because of spectral overlap.

Using the structural information, it was attempted to determine a structure-function relationship between the different polysaccharides and their immunomodulating effects, but it was not possible to determine any clear trends. These results are discussed further in paper 2.

2.2.4 Experimental

Extraction methods, monosaccharide composition, linkage analysis, Smith degradation and more performed by our collaborators are described in paper 2.

For NMR spectroscopy, the polysaccharides (5-10 mg) were dissolved in 500 µL D_2O (99.9%, Sigma). All NMR experiments were performed on an 800 MHz Bruker Avance III (799.90 MHz for ^{1}H and 201.14 MHz for ^{13}C) equipped with a 5 mm TCI cryoprobe. Acetone was used as internal reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively). The following experiments were

used for assignments and were all acquired at 313 K: 1H with presaturation of the water signal, 13C with power-gated decoupling, DQF-COSY, TOCSY with 60 ms mixing time using the MLEV17 sequence, NOESY with 200 ms mixing time, HSQC with multiplicity editing, HSQC-TOCSY with 60 ms mixing time using MLEV17 sequence and HMBC optimised for 10 Hz long range couplings. Spectra were recorded and processed using TopSpin 3.5 (Bruker).

2.3 Structural elucidation of polysaccharide fragments for β-lactoglobulin binding

This project was part of a collaboration Johnny Birch and Emil Stender of the Birte Svensson group at DTU Bioengineering, within which they had obtained oligosaccharide mixtures for binding studies in protein NMR titration with the Birthe Kragelund group at the University of Copenhagen. My part of the project was to analyse and assign the oligosaccharide mixtures. The results using the alginate oligosaccharide mixture are described in paper 3, while the results of the LY03 and 1187 oligosaccharide mixtures will be published at a later time.

2.3.1 Introduction

β-lactoglobulin is suggested to be a transport protein in the lipocalin family[114] capable of binding a wide array molecules. While it is one of the most abundant and studied proteins from bovine milk[115], little is known about the abilities of β-lactoglobulin to transport carbohydrates. The protein is known to interact with polysaccharides[116–119], but where and how it binds them is less clear.

To further understand how β-lactoglobulin binds carbohydrates, three different polysaccharides were chosen for binding studies and subsequently broken down into smaller oligosaccharides, either enzymatically or by acid hydrolysis. The first polysaccharide was LY03, a hetero exopolysaccharide from *Streptococcus thermophillus*[120], the second was the hetero exopolysaccharide 1187 from *lactobacillus delbrueckii bulgaricus*[121] and the last of the three polysaccharides was alginate[122] from *Phaeophycea* (brown algae).

In order to understand how and why these different oligosaccharide mixtures bind to β-lactoglobulin, it was also necessary to determine the structures.

2.3.2 Results and discussion

In the following sections, the NMR characterisation of several oligosaccharides will be discussed, and the purification of the polysaccharides, subsequent degradation to oligosaccharides and purification thereof were all performed by our collaborators at DTU Bioengineering. All of the samples contains both glycerol from the purification process as well as 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as internal standard.

2.3.2.1 LY03, a hetero exopolysaccharide from Streptococcus thermophillus

The LY03 polysaccharide obtained from *Streptococcus thermophillus* was subjected to acid hydrolysis by our collaborators, and the resulting oligosaccharide mixture was further separated by

high-performance liquid chromatography (HPLC) producing a sample containing only tetrasaccharides.

The repeating unit structure of LY03 has been described previously[120] and is shown in figure 11, and the sample was found to contain three tetrasaccharides, one major and two minor species in an approximately 8:1:1 ratio. The three tetrasaccharides were named LY03-1, LY03-2 and LY03-3, and how they originate from the repeating unit is shown in figure 11. The NMR assignment of the major species, LY03-1, is shown in table 5.

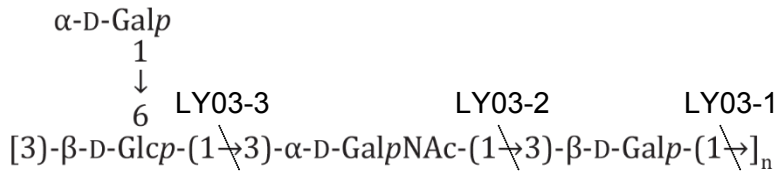


Figure 11 – Repeating unit of the Streptococcus thermophiles LY03 polysaccharide with the hydrolysed linkages resulting in each of the three tetrasaccharides LY03-1, LY03-2 and LY03-3 highlighted.

From the HSQC spectrum, the anomeric region being shown in figure 12, five distinct anomeric protons of similar intensity, as well as several of lesser intensity can be seen, and they were named **A**-**E** in order of descending ^{13}C chemical shift.

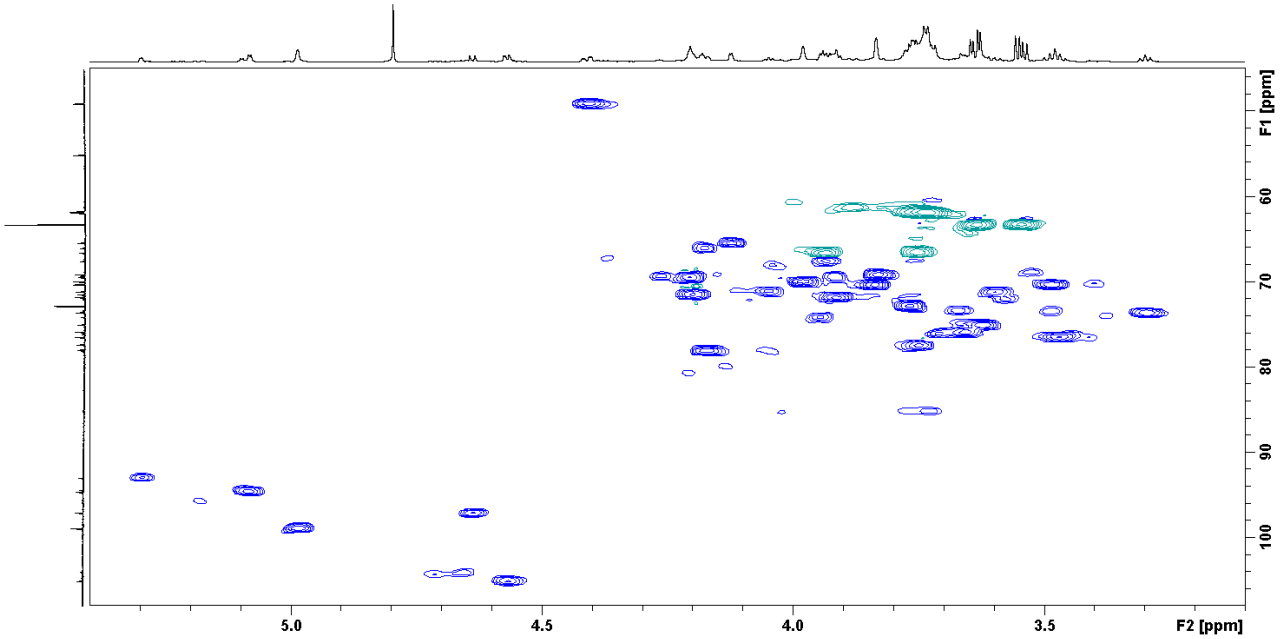


Figure 12 - HSQC spectrum of the LY03 sample. The projections are from separate experiments.

The first of these, **A**, was determined to be a β-glucose, based on the high $^{3}J_{^{1}H,^{1}H}$ couplings that could be measured as well as the chemical shifts. The only chemical shift from the β-Glcp that diverted from what would be expected was the 6-position[40].

The second monosaccharide, **B**, was determined to be from an unsubstituted α-galactose with low $^{3}J_{^{1}H1,^{1}H2}$, $^{3}J_{^{1}H3,^{1}H4}$ and $^{3}J_{^{1}H4,^{1}H5}$ coupling constants.

The third of the anomeric positions, **C**, was from a reducing β-Galp which, similarly to **B**, had low coupling constants from the 4-position proton to the adjacent protons. Additionally, from the chemical shift of the reducing β-Galp, the 3-position was suspected to be substituted.

The fourth anomeric position, **D**, belonged to an α-GalpNAc, which again had low couplings from neighbouring protons to H4 and chemical shifts corresponding to a 3-substitution.

Lastly, **E** was found to be the other α-configuration of the galactose reducing end, and also had chemical shifts corresponding to a 3-substitution.

Table 5 - Assignment of the major LY03 tetrasaccharide LY03-1. Most of the 6-positions were rather indistinguishable as due to major overlap, but the ^{1}H chemical shifts were determined using DQF-COSY.

LY03-1	**1**	**2**	**3**	**4**	**5**	**6**
α-Galp B	4.986 98.88	3.838 70.25	3.832 69.05	3.981 69.85	3.916 71.74	3.825 61.73
β-Glcp A	4.57 105.07	3.298 73.51	3.472 76.4	3.982 69.85	3.624 74.96	3.938/3.755 66.51
α-GalpNAc	5.086 94.53	4.408 48.95	4.174 77.97	4.207 69.33	4.202 71.38	3.736 61.73
reducing α-Galp	4.638 97.07	3.599 71.06	3.752 77.37	4.123 65.31	3.659 75.8	3.856 61.73
reducing β-Galp	5.298 92.93	3.947 74.03	4.178 65.91	4.124 65.31	4.049 71.02	3.808 61.73

Finally, using HMBC and ROESY data, a summary of which can be seen in table 6, the substitutions of and linkages between the monosaccharides were confirmed to match the following structure: α-Galp-1,6-β-Glcp-1,3-α-GalpNAc-1,3-Gal, which corresponds the 1,3-linkage between the β-Galp and β-Glcp in the repeating unit backbone, shown in figure 11, to have been hydrolysed.

Table 6 - HMBC and ROESY correlations for the anomeric positions in LY03-1. [a] overlapping signals.

Anomeric position	HMBC correlations	ROESY correlations
α-Galp - **B_1**	**B_2**, **B_5**, **A_6**	**A_6**, **B_2**
β-Glcp - **A_1**	**A_2**, **D_3**	**A_3**, **A_5**, **D_3**
α-GalpNAc - **D_1**	**C_3**, **D_3**, **D_5**	**C_3**, **C_4**, **E_3**, **E_4**/**D_3**[a]
reducing α-Galp - **E_1**	**E_5**	**E_2**
reducing β-Galp - **C_1**	**C_2**	**C_2**, **C_3**, **C_5**

However, as mentioned earlier, and as evident from the spectrum shown in figure 12, some minor species were also present, and based on the chemical shift of the weaker anomeric signals, they were determined to be the products from hydrolysis of the α-GalpNAc-1,3-β-Galp and the β-Glcp-1,3-α-GalpNAc linkages in the backbone, and the unique signals from the two species were both in approximately 12% intensity of the signals from the major species LY03-1. The hydrolysed linkages are highlighted in figure 11 and the resulting structures are:

B **A** **D** **C/E**
LY03-1 (~80%): α-Galp-1,6-β-Glcp-1,3-α-GalpNAc-1,3-Gal

α-Galp
↓(1,6)
LY03-2 (~10%): β-Galp-1,3-β-Glcp-1,3-GalNAc

α-Galp
↓(1,6)
LY03-3 (~10%): α-GalpNAc-1,3-β-Galp-1,3-Glc

2.3.2.2 1187, a hetero exopolysaccharide from lactobacillus delbrueckii bulgaricus

The polysaccharide 1187, obtained from *Lactobacillus delbrueckii bulgaricus*, was subjected to partial acid hydrolysis by our collaborators. The subsequent HPLC purification resulted in a sample containing only tetrasaccharides. The repeating unit has been reported previously[121], and is shown in figure 13.

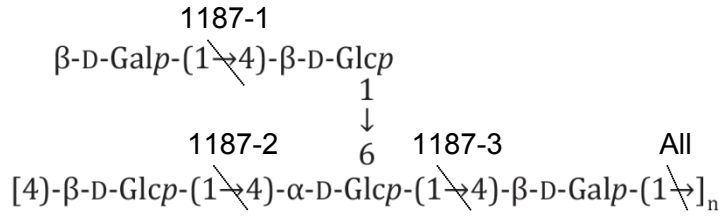


Figure 13 - Repeating unit structure of the 1187 polysaccharide. The hydrolysed linkages resulting in each of the three tetrasaccharides are highlighted.

The sample was found to contain three different tetrasaccharides, and unlike the LY03 sample the ratio was not as weighed towards mainly one of them, but rather approximately 6:3:1. Furthermore,

as the repeating unit of the 1187 polysaccharide is a pentasaccharide, each of the tetrasaccharide has two cuts from the repeating unit, shown in figure 13.

The HSQC spectrum from the 1187 sample, shown in figure 14, showed six different anomeric peaks of large, but different, intensities. Two of these six signals arise from a reducing end, and there is an additional set of peaks from another reducing end next to them, albeit at approximately 10% of the intensity. Additionally, it was evident that the sample contained two different substituted 6-positions as evident by the downfield methylene positions.

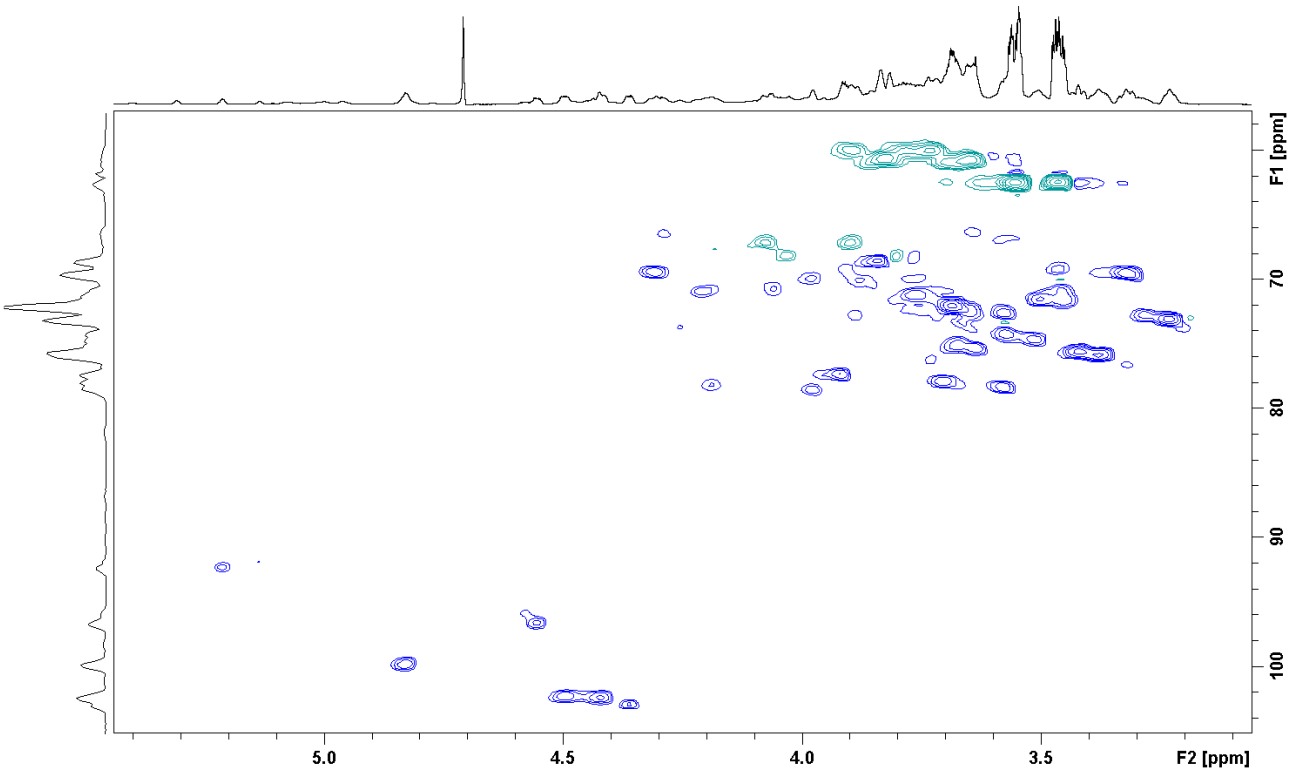


Figure 14 – HSQC spectrum of the 1187 sample.

The six peaks were dubbed **A**-**F** in order of descending ^{13}C chemical shift. Furthermore, the four non-reducing peaks originated from more than four signals, and thus the peaks **B** and **D** were further divided into **B1**/**B2** and **D1**/**D2**. The assignments of **A**, **B1**/**B2**, **C**, **D1**/**D2**, **E** and **F** can be found in table 7.

Starting from **A**, this peak had the lowest intensity of the non-reducing positions, and even lower than the β reducing end **E**. The signal was determined to arise from a terminal β-Galp, as determined partly from comparing the chemical shifts with literature[40] and also because of the weak $^{3}J_{^{1}H,^{1}H}$ couplings from H4 to adjacent protons.

The peak **B** had the highest intensity of any of the anomeric peaks, and it turned out to originate from two different β-Glpc, where **B1** was terminal and **B2** was 4-substituted. The peak **C** had a very similar shifts to **B1**, and was also determined as a terminal β-Glcp.

The last peak arising from non-reducing anomeric positions was **D**, and similar to **B** it was from two different glucopyranosides, but of α-configuration. Based on the downfield shifts of the 4- and 6-position **D1** was determined to be disubstituted on those positions, and **D2** only had the similar downfield shift of the 6-position and was thus determined as monosubstituted.

Finally, the reducing end, **E**/**F**, was determined as 4-substituted galactopyranosides, based on the downfield 13C chemical shift of the 4-position as well as the weak $^{3}J_{^{1}H,^{1}H}$ between H4 and the neighbouring protons.

From ROESY and HMBC spectra the following linkages could be determined: From the anomeric position of the β-Galp **A** there was correlations to the 4-position of the β-Glcp **B2**, and from the anomeric signals of **B1** and **B2** there were correlations to the 6-positions of **D1** and **D2**, respectively. From the anomeric positions of the α-Glcp **D**, correlations to the galactose reducing end **E** and **F** 4-positions could be observed. Lastly, from the anomeric position of the β **C** there were correlations to the 4-position of **C**. For a complete overview of HMBC and ROESY correlations see table 8.

Table 7 – Assignment of the major signals of the 1187 tetrasaccharides as well as partial assignments of additional reducing end. nd means not detected.

	1	2	3	4	5	6 (6a/6b)
A - β-Galp	4.443 103.8	3.541 71.66	3.658 73.22	3.922 69.2	3.720 76.04	3.774/3.774 61.66
B1 - β-Glcp	4.504 103.15	3.314 73.70	3.499 76.28	3.400 70.16	3.461 76.60	3.731/3.906 61.31
B2 - β-Glcp	4.518 103.03	3.362 73.38	3.594 75.31	3.661 79.01	3.655 74.83	3.814/3.979 60.67
C - β-Glcp	4.577 103.09	3.318 73.81	3.499 76.28	3.400 70.16	3.461 76.60	3.731/3.906 61.31
D1 - α-Glcp	4.911 100.58	3.582 72.26	3.849 71.77	3.785 78.69	4.393 70.00	3.982/4.157 67.75
D2 - β-Glcp	4.905 100.58	3.459 70.95	3.726 72.92	3.543 69.84	4.291 71.61	3.815/4.046 68.18
E - β-Galp red	4.634 97.36	3.531 72.26	3.729 72.89	4.001 78.05	3.757 75.8	3.758/3.892 60.42
F - α-Galp red	5.293 93.01	3.849 71.77	3.922 69.2	4.059 79.34	4.141 71.45	3.814/3.814 60.67
Reducing β-Glc	4.666 96.65	3.288 nd	3.618 nd	nd	nd	nd
Reducing α-Glc	5.857 92.06	3.572 nd	3.853 nd	3.592	nd	nd

The relative ratios between the three polysaccharides was determined using the integrals of the anomeric HSQC signals. As the repeating unit was a pentasaccharide, but the oligosaccharides were tetrasaccharides, each of them was the result of two different bonds having been hydrolysed. As there were only two reducing ends, two of the tetrasaccharides evidently had identical reducing ends, and furthermore, the minor reducing end had an approximately 1:9 ratio to the major.

As mentioned above, the peak corresponding to the anomeric positions in the two β-Glcp **B** had the highest integral of all the anomeric positions, this would likely be the common sugar between the three. Connected to **B** was the terminal β-Galp **A**, which had a much smaller integral, as well as the α-Glcp **D**, which only had a slightly smaller integral. To **D** was attached, in the case of **D1**, the terminal β-Glcp **C**, which had a slightly smaller integral than **A**. Using this information, the following three tetrasaccharide structures were proposed:

B1
β-Glcp
C ↓(1,6) **E/F**
1187-1 (~60%): β-Glcp-1,4-α-Glcp-1,4-Gal
D1

A **B2** **D2** **E/F**
1187-2 (~30%): β-Galp-1,4-β-Glcp-1,6-α-Glcp-1,4-Gal

A **B2**
β-Galp-1,4-β-Glcp
↓(1,6)
1187-3 (~10%): β-Glcp-1,4-Glc
C **G/H**

The corresponding hydrolysed linkages are highlighted in the repeating unit structure shown in figure 13.

Table 8 - Summary of the strong HMBC and ROESY correlations observable in the 1187 sample. nd means not detected.

Unit	HMBC correlations	ROESY correlations
A	B2-4	B2-4, A-3, A-5
B1	D1-6	D1-6a, D1-6b, B1-3, B1-5
B2	D2-6	D2-6a, B2-5, B2-3
C	D1-4	D1-4, C-3, C-5
D1	D1-5, reducing β-Gal-4	Reducing β-gal-4
D2	Reducing α-Gal-4	Reducing α-gal-4
E	nd	Reducing β-gal 5, reducing β-Gal-3
F	nd	nd
Reducing β-Glc	nd	nd
Reducing α-Glc	nd	nd

2.3.2.3 Alginate fragments

Our collaborators produced alginate trisaccharides by degrading alginate polysaccharides using an endo-acting alginate lyase[119,123] to give di- and trisaccharides which were subsequently separated by HPLC, as described in paper 3. As alginate is made up entirely of β-D-mannuronic acid and α-L-guluronic acid, which are C5-epimers, the resulting non-reducing end of the β-elimination will be identical no matter what monosaccharide it was before the elimination.

The mixture consisted of only two trisaccharides in a 3:1 mixture, shown in figure 15, where the reducing end was all guluronic acid, the middle monosaccharide was guluronic acid in the major species and mannuronic acid in the minor species and the non-reducing ends were identical. An almost full assignment, as well as the $^{1}J_{^{1}H1,^{13}C1}$ coupling constants, of both trisaccharides can be found in table 9. The major product, named ALG-1, was identified as (4-deoxy-α-L-erythro-hex-4-enopyranosyluronate)-(1,4)-(α-L-gulopyranosyluronate)-(1,4)-(L-gulopyranosyluronate) and the minor product, named ALG-2, as (4-deoxy-α-L-erythro-hex-4-enopyranosyluronate)-(1,4)-(β-D-mannopyranosyluronate)-(1,4)-(L-gulopyranosyluronate).

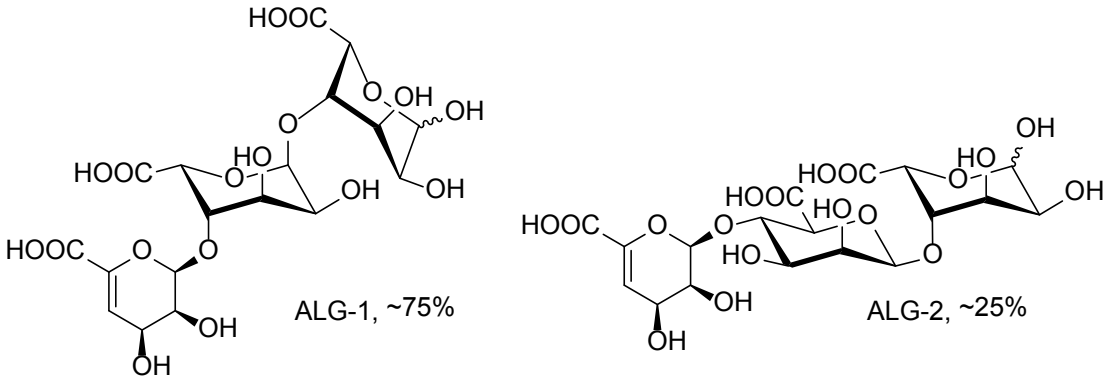


Figure 15 – Structures of the trisaccharides in the sample. The left structure is the major component in the mixture, and the only difference between the two structures is whether the middle carbohydrate is an α-L-guluronic acid (ALG-1) or a β-D-mannuronic acid (ALG-2).

The spectra contained seven distinct anomeric signals, dubbed **A**-**G** in order of descending ^{13}C chemical shift, see figure 16. The spectra were assigned using DQF-COSY, ROESY, HSQC, HSQC-TOCSY and HMBC. Lastly, as some of the monosaccharides had mannose configuration, $^{1}J^{1}_{H1,}{}^{13}_{C1}$ coupling constants[38] were required to determine the anomeric configuration, and to this end, CLIP-HSQC[28] was used.

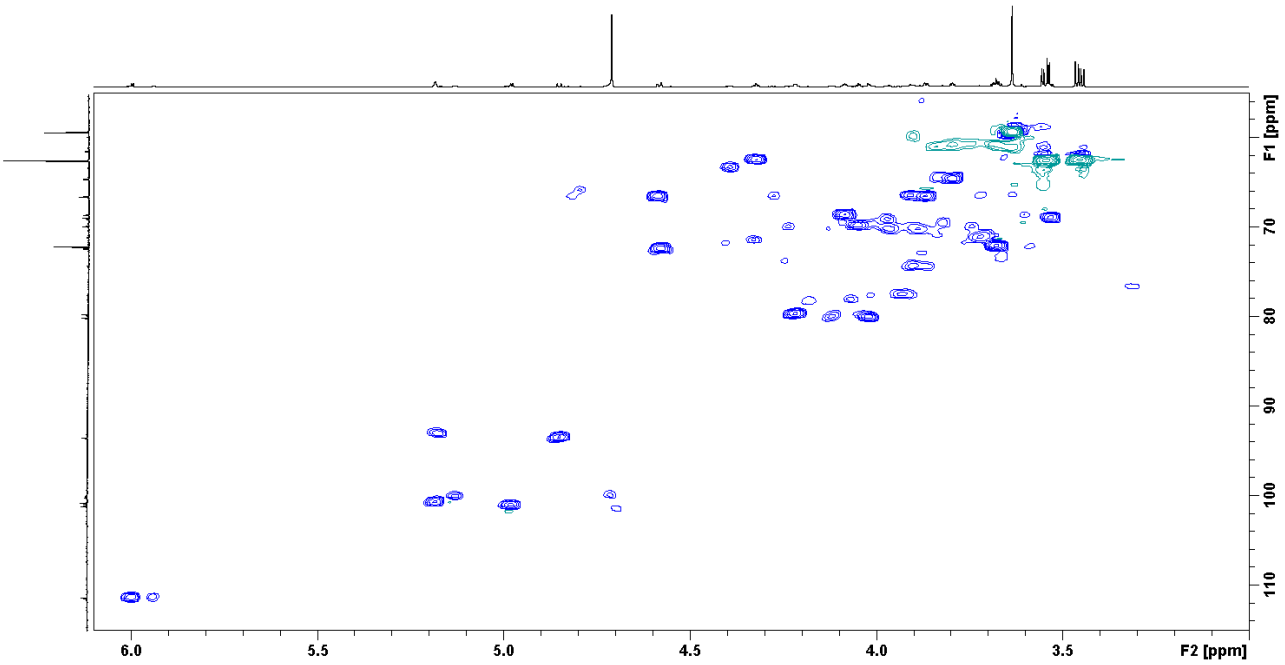


Figure 16 – HSQC spectrum of the of the alginate sample.

The first of these seven signals, **A**, had the lowest intensity of any of the anomeric positions and was determined to be from a mannuronic acid with β-configuration, based on the $^{1}J^{1}_{H1,}{}^{13}_{C1}$ coupling constants. They were determined as having mannose configuration as the $^{3}J^{1}_{H1,}{}^{1}_{H2}$ was quite weak, but the remaining $^{3}J^{1}_{H,}{}^{1}_{H}$ coupling constants were strong. Similar to **A** was **E**, which had very similar

chemical shifts and was also found to be of mannose configuration. Both **A** and **E** were weak signals, the two weakest of the seven anomeric signals, with **E** being the major of the two.

The second anomeric signal **B** was, unlike **A** and **E**, determined to have an equatorial proton based on the large $^{1}J^{1}_{H1,^{13}C1}$ coupling constant, and was determined to be of gulose configuration due to the low $^{3}J^{1}_{H3,^{1}H4}$ coupling constant of 3.6 Hz. This also meant that it was of α-configuration.

The next two anomeric positions, **C** and **D**, were, much like **A** and **E**, quite similar. They were both determined to have a double bond between their 4- and 5-positions, and thus quite unique chemical shifts, and the only difference between the two species were whether they were attached to a guluronic acid or mannuronic acid[124,125].

The final two of these seven signals, **F** and **G**, originated from a reducing end, meaning there was only a single monosaccharide as the reducing end. As one of these reducing ends, **G**, had a large $^{3}J^{1}_{H1,^{1}H2}$ of 8.4 Hz, the H2 could not be in equatorial position, and thus the reducing end could not be of mannose configuration[37]. This was from the β-reducing end, and had approximately twice the integral of the α-reducing end. Due to the low intensity of the α-reducing end it was not possible to fully assign, and similarly, due to the low concentration of the 6-position carbonyl in all the monosaccharide it was also not possible to assign.

Lastly, using HMBC and ROESY, the interconnections between the monosaccharides could be determined, and from the non-reducing ends **C** and **D** correlations to the 4-positions of **B** and **A**/**E**, respectively, was observed. Similarly, from the anomeric positions in the middle monosaccharides **B** and **A**/**E** correlations to the reducing end 4-positions were observed.

Table 9 – Assignment of the trisaccharides. na means not applicable and nd means not detected.

ALG-1β	$^{1}J^{1}_{H1,\ ^{13}C1}$	**1**	**2**	**3**	**4**	**5**	**6**
Term **C**	171 hz	5.185 100.73	3.867 66.43	4.322 62.35	5.995 111.29	na 141.73	na nd
α-GulA (mid) **B**	171 hz	4.978 101.01	3.798 64.45	4.086 68.55	4.217 79.55	4.586 66.46	na nd
β-GulA (red) **F**	164 hz	4.856 93.57	3.534 68.87	4.049 69.78	4.02 79.99	4.577 72.28	na nd
ALG-1α	$^{1}J^{1}_{H1,\ ^{13}C1}$	**1**	**2**	**3**	**4**	**5**	**6**
Term **C**	171 hz	5.185 100.73	3.867 66.43	4.322 62.35	5.995 111.29	na 141.73	na nd
α-GulA (mid) **B**	171 hz	4.978 101.01	3.798 64.45	4.086 68.55	4.217 79.55	4.586 66.46	na nd
α-GulA (red) **G**	172 hz	5.180 93.03	3.823 69.39	ND ND	4.123 79.97	4.794 65.86	na nd
ALG-2β	$^{1}J^{1}_{H1,\ ^{13}C1}$	**1**	**2**	**3**	**4**	**5**	**6**
Term **D**	171 hz	5.130 99.95	3.907 66.42	4.394 63.27	5.94 111.32	na 141.73	na nd
β-ManA (mid) **E**	159 hz	4.717 99.85	3.968 70.17	3.726 71.02	3.935 77.47	3.904 74.29	na nd
β-GulA (red) **F**	164 hz	4.856 93.57	3.534 68.87	4.049 69.78	4.02 79.99	4.577 72.28	na nd
ALG-2α	$^{1}J^{1}_{H1,\ ^{13}C1}$	**1**	**2**	**3**	**4**	**5**	**6**
Term **D**	171 hz	5.130 99.95	3.907 66.42	4.394 63.27	5.94 111.32	na 141.73	na nd
β-ManA (mid) **A**	163 hz	4.696 101.54	3.972 69.05	3.726 71.02	3.935 77.47	3.87 74.36	na nd
α-GulA (red) **G**	172 hz	5.180 93.03	3.823 69.39	nd nd	4.123 79.97	4.794 65.86	na nd

From the assignment it would seem that the alginate polysaccharide was primarily made up of guluronate, but as only the trisaccharides were analysed, it is not possible to draw far reaching conclusions on the overall alginate structure. Furthermore, the terminal end of the sugars could arise from both guluronate and mannuronate.

2.3.3 Conclusion

The structures of LY03 and 1187 fragments were assigned. Challenges arose due to the high similarity of the monosaccharides in each tetrasaccharide, the different tetrasaccharides present in the samples, as well as low concentration. Using the knowledge of the determined structures it is possible to build the different tri- and tetrasaccharides for future docking studies that are to be compared to the results of the β-lactoglubolin titration results.

2.3.4 Experimental

2.3.4.1 Summary of LY03 and 1187 acid hydrolysis and purification by our collaborators

Our collaborators performed the hydrolysis by treating the polysaccharides with 0.1 M trifluoroacetic acid (TFA) at 90° C for 20 minutes followed by neutralisation of the pH and filtration to collect oligosaccharides with a molecular mass below 3000 g/mol. Larger sugars were further hydrolysed by repeating the acid hydrolysis described above, and it was repeated until little to no >3000 g/mol sugars were left. Using a Bio-Gel P-2 column (Bio-Rad), oligosaccharides in the sizes from tri- to heptasacchararides were collected and further purified using HPLC fitted with a TSKgel amide-80 column and a refractive index detector.

2.3.4.2 LY03 sample

The sample had been used for titration studies with β-lactoglobulin studies and contained DSS as it was used as internal standard for those studies, as well as glycerol from the ensuing purification of the oligosaccharides by filtration. The sample amount was 1.2 mg of this mixture and was dissolved in 500 µL D_2O.

2.3.4.3 1187 sample

Similarly to the LY03 sample, this sample had also been used for titration studies, but the sample amount here was only 0.5 mg and also dissolved in 500 µL D_2O. It was attempted to concentrate the sample and redissolve in a smaller amount of solvent using a 3 mm tube. Presumably due to the high salt concentration of the sample, this approach did not improve the sensitivity.

2.3.4.4 Alginate sample

The alginate trisaccharide sample was prepared as described previously[119] and in paper 3. Similarly to the other two samples, it had been used for titration studies, described in paper 3, and the sample amount was 2.1 mg dissolved in 500 µL D_2O.

2.3.4.5 NMR experiments

All spectra were recorded on an 800 MHz Bruker Avance III (799.75 MHz for ^{1}H and 201.10 MHz for ^{13}C) equipped with a 5 mm TCI cryoprobe at 25° C. DSS was used as internal reference (0.00 ppm ppm and -2.00 ppm for ^{1}H and ^{13}C, respectively. The following experiments were used for the elucidations: 1H with presaturation, DQF-COSY, ROESY, HSQC, HSQC-TOCSY and HMBC, and they were all carried out under similar conditions as those described in section 2.2, with exception of the ROESY experiment used instead of NOESY. The ROESY mixing time was always 200 ms. Spectra were recorded using TopSpin 3.5 and processed using TopSpin 3.5.

3 Enzymatic synthesis of carbohydrates

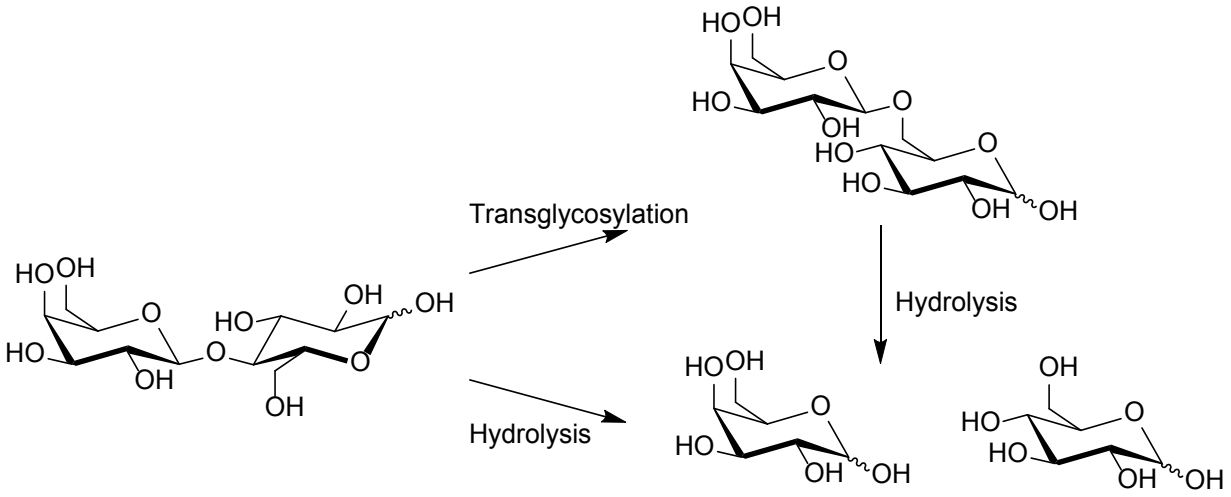
3.1 Investigation of the *lacZ* β-galactosidase using dDNP NMR

This project was primarily carried out at the Center for Hyperpolarization in Magnetic Resonance at DTU Electro.

3.1.1 Introduction

Dissolution dynamic nuclear polarization (dDNP[43] NMR) spectroscopy has been used for investigations for a number of chemical, biochemical and biological systems, including the detection of enzymatic intermediates[54]. As mentioned in section 1.3, the polarisation is lost with the T_1 relaxation, and consequently, nuclear spins with long T_1 times are preferred. For organic molecules, the two most relevant spin-½ nuclei are ^{1}H and ^{13}C. As most ^{1}H nuclei have short relaxation times of less than 2 seconds, ^{13}C spins at non-protonated sites is often the better option, as these sites can have T_1 relaxation times of up to more than a minute[53]. Other nuclei could also be useful, such as ^{15}N and ^{31}P, but for general use with carbohydrates, ^{13}C is the most appropriate choice.

In order to properly determine the usefulness dDNP NMR for investigating glycosidases, the enzyme *lacZ* β-galactosidase from *E. coli* (E.C. number 3.2.1.23) was chosen as the model enzyme, as it is very well described[2]. The enzyme is a tetramer with subunits of 1023 amino acids, and the structure was elucidated by crystallography in the mid 1990s[126]. It is a retaining glycosidase, and is generally believed to utilise the double displacement mechanism[127–129], with lactose most likely being the natural substrate[130]. The enzyme is well known for making transglycosylation to the 6-position of the glucose to form allolactose[131], shown in scheme 3, a reaction that is notorious as *lacZ* part of the lac operon system used for regulation of gene expression[132]. Other nucleophiles have also been used as acceptors[133], but for carbohydrate acceptors only 1,6-transglycosylation has been reported[131,134,135]. Furthermore, using computational modelling, Brás *et al.* (2010)[136] predicted that the enzyme should give approximately 100% 1,6-transglycosylation with glucose as the acceptor, and that 1,4-linked product would be theoretically more favourable than 1,3-linked product, even though both were very unlikely. From an industrial point of view, β-galactosidases are of great interest, as they can be used to make galacto-oligosaccharides from lactose[137–139], or simply be used in dairy production to remove lactose[140,141].



Scheme 3 – Lactose hydrolysis and transglycosylation to allolactose by β-galactosidase.

Using conventional NMR, it is possible to use study enzymes at equilibrium[142], and for glycosidases it is also possible to determine whether they have a retaining or inverting mechanism[143]. However, for quantitative NMR the delay between each scan should be at least five times the T_1 to ensure reliable quantification, which - together with signal averaging - results in low temporal resolution. Consequently, a tradeoff between sensitivity and temporal resolution has to be accepted , as acquiring fewer FIDs per spectrum will give lower sensitivity, but also shorter acquisition times. Upwards of several minutes between each spectrum is not uncommon in NMR monitoring of enzymes using ^{1}H NMR. Due to the narrow 1H chemical shift range, HSQC monitoring can be applied to utilise the much broader range of ^{13}C, but this does not improve the acquisition time compared to ^{1}H monitoring[144]. To increase both the sensitivity and the temporal resolution, dDNP NMR seemed like the optimal solution.

Several examples of dDNP of carbohydrates exists, and the most commonly used is glucose, but as it has no quaternary positions, 2H labelling is required. Even so, it has been used for several studies[145–150], both *in vitro* and *in vivo*, but only one of these was an enzymatic study, namely on glucose phosphorylation by hexokinase[149]. The remaining glucose studies mainly focused on metabolism in either yeast[146], bacteria[145,147] or mouse tumors[148], with the exception of Meier *et al.*[150] that described the detection of intermediates in molybdate-catalysed glucose epimerization. So while examples of dDNP NMR of aldoses exist, the method has only been applied using commercially available, labelled glucose substrate. Unlike aldoses, ketoses like fructose have a quaternary anomeric position with long T_1 times on the order of 30 seconds, and consequently fructose was the first non-carbonyl used as a metabolic probe for *in vivo* experiments[48].

Much like glucose, its C4 epimer galactose does not contain any quaternary carbon, and ^{2}H labelling is required to increase the T_1 time to a usable level. As the anomeric carbon is the position of interest, having the deuteration at this position would be optimal. Additionally, having a ^{13}C label would further increase the sensitivity by an additional two orders of magnitude. However, doubly labelled galactose is less commonly used, and harder to obtain, than glucose, and originally it was planned to synthesise it from glucose by epimerisation of the 4-position[151]. Luckily, D-[1-^{13}C;1-^{2}H]galactose became commercially available during this project, thus simplifying the β-galactosidase substrate synthesis significantly.

Finally, while lactose is likely the natural substrate, β-galactosidase is fairly promiscuous when it comes to substrates, as it will hydrolyse most β-galactopyranosides[2]. As the hydrolysis of lactose is quite cumbersome to monitor[131], the use of *o*- or *p*-nitrophenyl β-D-galactopyranoside (Galp-onp and Galp-pnp, respectively) as model substrates are commonplace. This also happens to be quite useful for dDNP NMR, as these substrates contain quaternary positions on the aglycon, which allows natural abundance dDNP experiments to be performed to determine proper conditions, both for the polarisation buildup and the ensuing enzyme catalysed turnover.

3.1.2 Results and discussion

3.1.2.1 Natural abundance experiments

As mentioned above, natural abundance Galp-onp was used to determine the experiment setup, made possible by the fairly long T_1 relaxations of the two quaternary positions of the nitrophenyl aglycon. The natural abundance substrate was used to determine the appropriate conditions to be used for the labelled experiments, including the formulation of the DNP mixture, the temperature of the dissolution buffer and the enzyme to substrate ratio.

As the Galp-onp is poorly soluble in water, dimethyl sulfoxide (DMSO) was used as the solvent for the DNP samples, and even at 1.4 M concentration it gave quite nice glasses when cooled. The addition of Gd(III) containing compounds to boost the total polarisation[62,152] was attempted, but no significant increase in polarisation was observed, whereas a shortening of the T_1 was found. The experiments led to the conclusion that using Galp-onp in a final concentration of approximately 12 mM would be favourable with six units of β-galactosidase per mM of onp-Galp.

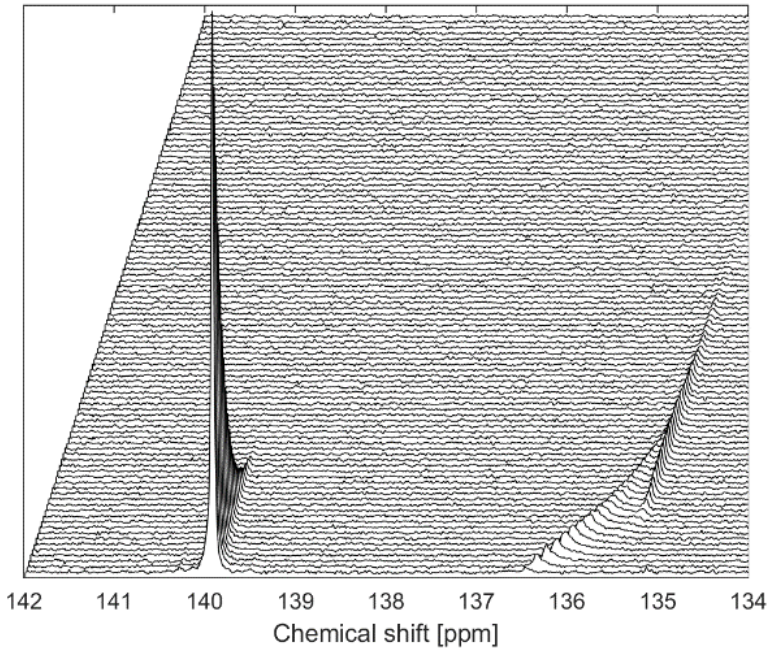


Figure 17 – Hyperpolarised Galp-onp (35 mM) hydrolysis by β-galactosidase (150 u). 100 spectra with a repetition time of 2 seconds. Each subsequent spectrum is shifted 0.02 ppm upfield.

At high substrate concentrations of the unlabelled substrate, around 35 mM, a migration of the released o-nitrophenol signals were observed, as well as several species of *o*-nitrophenol present at the same time, as shown in figure 17. This migration is over more than an entire ppm, and before the entire signal falls into a single peak, there are multiple species of the released *o*-nitrophenol at the same time.

However, what exactly causes the migration, or the concurrent species, is currently unknown. It could be due to temperature, pH, hydrophobic effect, binding to the enzyme following release or a combination of the above. This migration does seem to have some correlation to the concentration of substrate, and thus of released nitrophenol, as lower substrate concentration lead to less migration, see figure 18. Comparing the peak migration in figure 17 and figure 18, it is evident that the peak migrates a lot less at lower substrate concentration, as in both examples the signal start out at the same chemical shift, but in the lower concentration sample reaches an endpoint at 136.0 ppm, while in the high concentration sample it reaches an endpoint at 135.3 ppm.

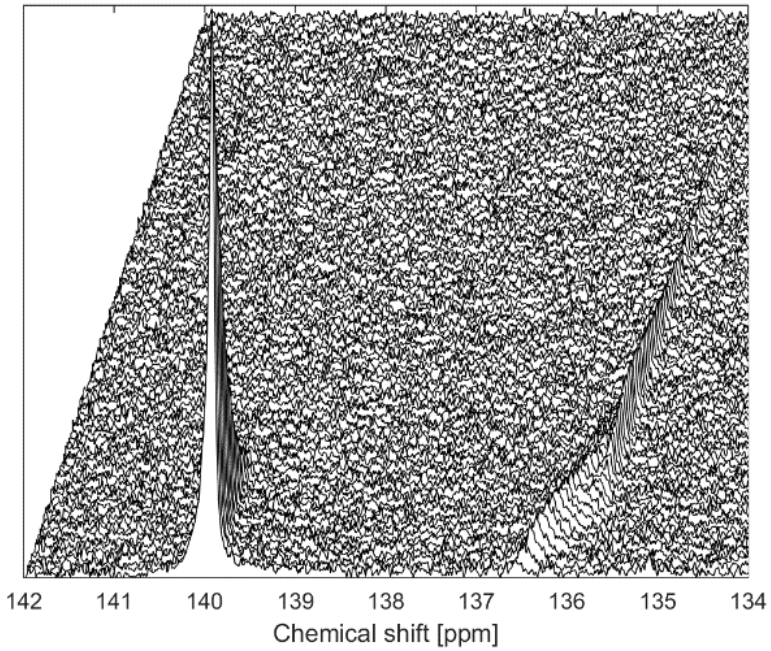


Figure 18 - Hyperpolarised Galp-onp (12 mM) hydrolysis by β-galactosidase (50 u). 100 spectra with a repetition time of 2 s. Each subsequent spectrum is shifted 0.02 ppm upfield.

3.1.2.2 Isotope labelled experiments

The first observations from using the doubly labelled substrate included that it had a shorter T_1 time compared to the quaternary carbons of the nitrophenyl group. Consequently, the timeframe for performing the experiment was shorter, and thus the amount of enzyme and the NMR flip angles was adjusted to better capitalise on the short-lived signal.

After having determined suitable conditions for the DNP experiment and subsequent enzyme hydrolysis, the experiment was performed with the doubly labelled substrate, shown in figure 19. In the spectra, the rapid formation of β-galactopyranoside and the formation of less populated species can be observed, but as the concentration of acceptor is low, the signals corresponding to transglycosylation products are quite small.

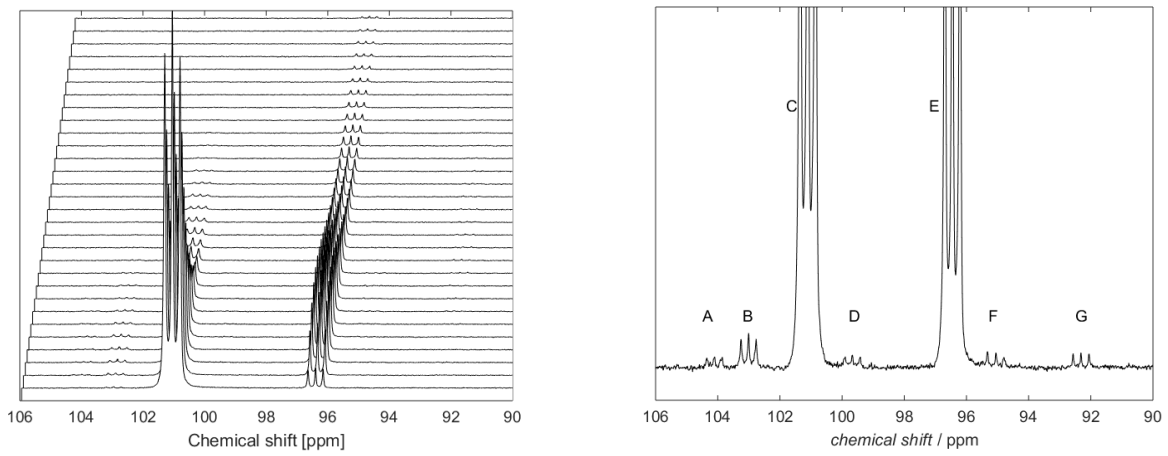


*Figure 19 - Left: The first 30 spectra of the hyperpolarised doubly labelled Galp-onp (10 mM) hydrolysed by β-galactosidase (87.5 u). Each subsequent spectrum is shifted 0.06 ppm. Right: The sum of spectra 5 to 10, and the labels correspond to **A**: β-Galp-1,3/4-Gal, **B**: β-Galp-1,6-Gal, **C**: β-Galp-onp, **D**: β-Galp-1,1-Gal, **E**: β-Galp, **F**: α-Galf and **G**: α-Galp*

In order to determine if these signals were from transglycosylation products, and to increase the transglycosylation production, excess galactose was added as the acceptor. As seen in figure 20, some of the signals that were barely detectable in the experiment without added acceptor, figure 19, are now much more intense, while those corresponding to mutarotation of the released β-galactopyranoside have similar intensity.

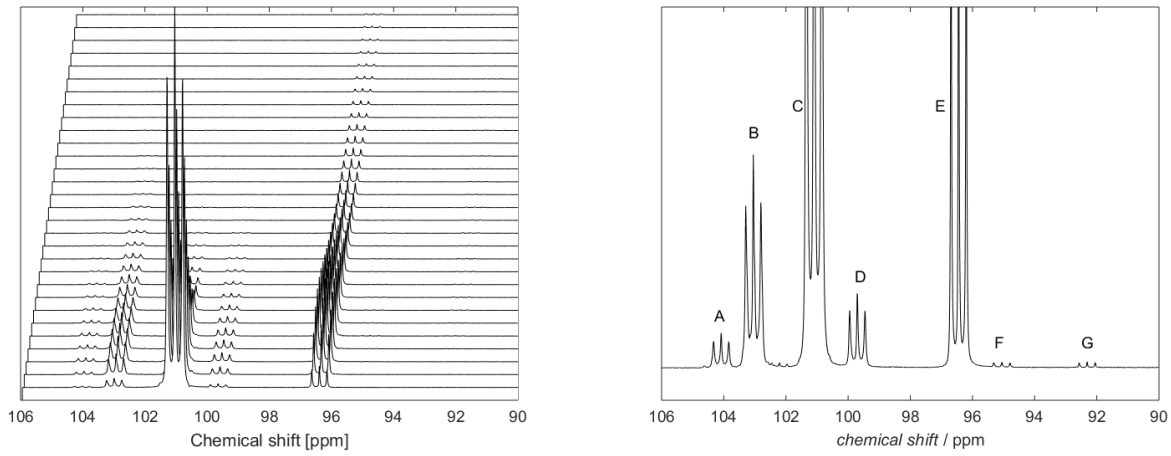


*Figure 20 - Left: The first 30 spectra of the hyperpolarised doubly labelled Galp-onp (10 mM) hydrolysed by β-galactosidase (87.5 u) with galactose (170 mM) added as acceptor. Each subsequent spectrum is shifted 0.06 ppm. Right: The sum of spectra 1 to 10, and the labels correspond to **A**: β-Galp-1,3/4-Gal, **B**: β-Galp-1,6-Gal, **C**: β-Galp-onp, **D**: β-Galp-1,1-Gal, **E**: β-Galp, **F**: α-Galf and **G**: α-Galp.*

The three transglycosylation products were identified by comparison to reported literature chemical shift values, and the largest product, labelled **B**, was identified as β-Galp-1,6-Gal[139], which was also the expected product, as the enzyme is only known to make 1,6-transglycosylation[2,131,135,136].

The second largest transglycosylation signal, labelled **D**, was determined as β-Galp-1,1-β-Galp, as this would be the best explanation for the upfield shift of the β-Galactopyranoside[153]. To further verify this 1,1-linkage, 1-*O*-methyl galactopyranoside was added as acceptor instead of galactose, and the peak almost disappeared, shown in supporting information for paper 4. Furthermore, as a fairly large difference between the chemical shifts of the polarised β-Galp in β-Galp-1,1-β-Galp and β-Galp-1,1-α-Galp[153], and in the experiments with no added acceptor, mostly β-galactose would be available until equilibrium could be achieved through mutarotation. The last of the three signals corresponding to a transglycosylation product, labelled **A**, was determined to be from either β-Galp-1,3-Galp or β-Galp-1,4-Galp, or both, as they have very similar chemical shift[154]. The final two minor species, labelled **F** and **G**, were assigned by comparing with thermal NMR spectra of the doubly labelled galactose, and were assigned as α-galactofuranoside and α-galactopyranoside. The β-galactofuranoside is not observed at it overlaps with the Galp-onp signal.

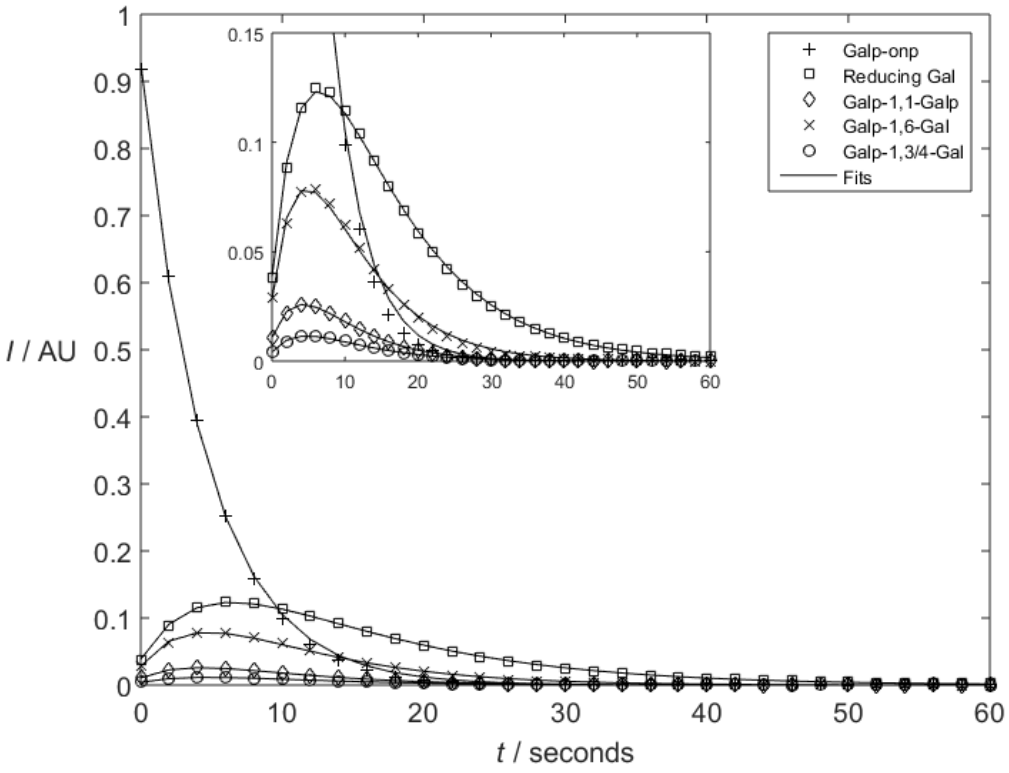


Figure 21 – Signal integrals of the five species used in the kinetic model of scheme 4 and scheme 5 plotted over time alongside the fits generated by said model. The signal intensity units are arbitrary and are normalised in the first spectrum The insert is a zoom on the product signals.

The relative transglycosylation rates of the different products could be determined fairly simply by comparing the peak signal integrals between them, easiest seen in figure 21, and the 1,6-linkage was formed 3.0 times faster than the 1,1-linkage and 6.9 times faster than the 1,3/4-linkage.

However, the relative hydrolysis rates were more difficult to approximate, as the signal loss due to hydrolysis and polarisation loss have the same functional form. To account for this problem, a kinetic model of the enzymatic reactions was developed based on the seven reactions shown in scheme 4. This relatively simplified model does not account for formation of one transglycosylation product from another, as this is an equilibrium the model would not be able to properly determine. Additionally, as it would not be possible to differentiate between what types of galactose an anomeric position was linked to, be it a reducing end or not, those are also indirectly incorporated into the model, as only the hyperpolarised labelled positions are detectable.

$$\text{*Galp-onp} + H_2O \xrightarrow{k1} \text{*Gal} + \text{onp}; \quad \text{*Galp-onp} + \text{Gal} \xrightarrow{k2} \text{*Galp-1,1-Galp} + \text{onp};$$

$$\text{*Galp-onp} + \text{Gal} \xrightarrow{k3} \text{*Galp-1,6-Gal} + \text{onp}; \quad \text{*Galp-onp} + \text{Gal} \xrightarrow{k4} \text{*Galp-1,3/4-Gal} + \text{onp};$$

$$\text{*Galp-1,1-Galp} + H_2O \xrightarrow{k5} \text{*Gal} + \text{Gal}; \quad \text{*Galp-1,6-Gal} + H_2O \xrightarrow{k6} \text{*Gal} + \text{Gal}; \quad \text{*Galp-1,3/4-Gal} + H_2O \xrightarrow{k7} \text{*Gal} + \text{Gal}$$

*Scheme 4 – The reactions used for the kinetic model. Anomeric configurations have not been noted, but are β for all non-reducing ends. The * marks the labelled detectable galactose. All reducing ends could also be labelled, but it would not be possible differentiate between free and substituted reducing galactose due to overlap.*

From these seven reactions, the five differential equations shown in scheme 5 were devised, and by minimizing least squares, fits were acquired. The apparent T_{1a} of the reducing galactose was constrained by measuring it in the spectra where the substrate signals were gone, but the reducing galactose signals were still present, and in all samples it ended up being between 10.5 and 11.4 seconds. The T_{1a} for the four enzyme substrates were constrained between 7.5 and 8.5 seconds, based on comparison with experiments without addition of enzyme.

$$d[\text{Galp-onp}]/dt = -[\text{Galp-onp}]\cdot(1/T_{1a}1+k1+k2+k3+k4)$$

$$d[\text{Galp-1,1-Galp}]/dt = [\text{Galp-onp}]\cdot k2 - [\text{Galp-1,1-Galp}]\cdot(1/T_{1a}2+k5)$$

$$d[\text{Galp-1,6-Gal}]/dt = [\text{Galp-onp}]\cdot k3 - [\text{Galp-1,6-Gal}]\cdot(1/T_{1a}3+k6)$$

$$d[\text{Galp-1,3/4-Gal}]/dt = [\text{Galp-onp}]\cdot k4 - [\text{Galp-1,3/4-Gal}]\cdot(1/T_{1a}4+k7)$$

$$d[\text{Gal}]/dt = [\text{Galp-onp}]\cdot k1+[\text{Galp-1,1-gal}]\cdot k5+[\text{Galp-1,6-gal}]\cdot k6+[\text{Galp-1,3/4-gal}]\cdot k7 - [\text{Gal}]\cdot 1/T_{1a}5$$

Scheme 5 – The differential equations used for the model. T_{1a} is the apparent T_1 relaxations of each species, including polarization loss due to NMR pulsing.

After having obtained the model fits, those can be plotted without the polarisation loss to give an approximate kinetics plot, shown in figure 22. From this fit it is evident that the 1,1-linked galactoside is hydrolysed at a far higher rate compared to the 1,3/4- and 1,6-linked, but not at nearly as high a rate as the Galp-onp. The 1,3/4- and the 1,6-linked transglycosylation products had similar hydrolysis rates, which fits with similar comparison between lactose and allolactose hydrolysis rates, in which allolactose was found to be at least as good a substrate than lactose[155].

The transglycosylation rates from the model were quite similar to those determined from peak integrals.

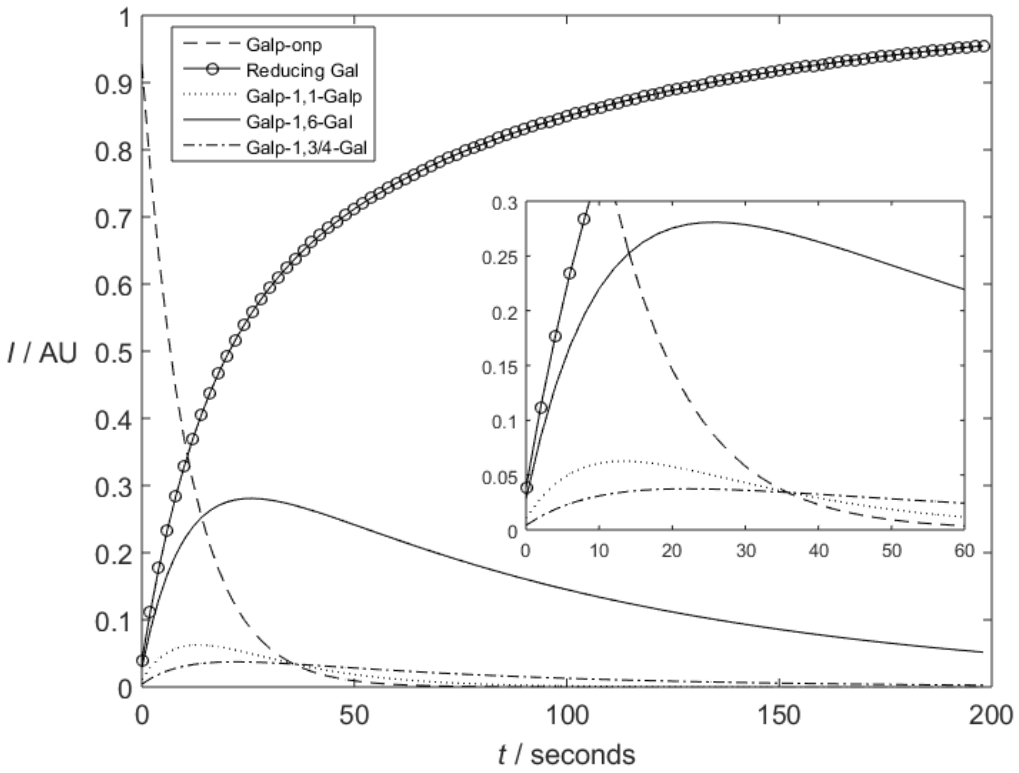


Figure 22 - The modelled fits without polarisation loss. The insert is a zoom.

When looking into smaller details in the spectra from figure 20, a zoom of which is shown in figure 23, the products formed by mutarotation show up after several spectra, which translates to several seconds after release of β-Galp. Additionally, the initial formation of α-Galf is slightly faster than α-Galp, but it also reaches its peak intensity faster. Lastly, there is also a small species showing up around 102.3 ppm, but it is very low populated. It could be from a 1,2-linkage, the β-Galp-1,1-α-Galp species or even one of the assigned signals that have shifted due to additional substitution.

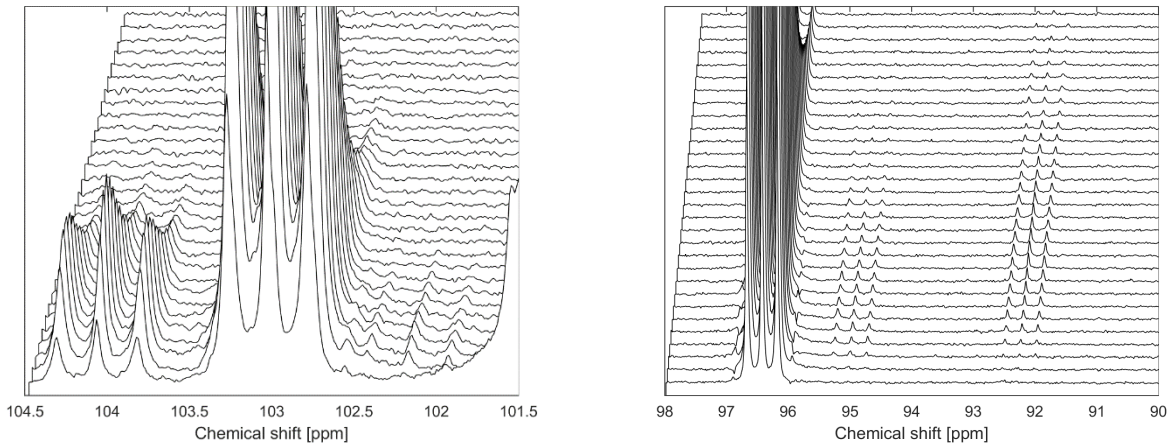


*Figure 23 - Zoom of the spectra shown in Figure 20. Left: The 1,3/4-linked product, the 1,6-linked product and to the far right is the edge of the Galp-onp. The signal around 102.3 ppm is unassigned. Right: the three signals are from β-galactopyranoside, α-galactofuranoside and α-galactopyranoside, **E**, **F** and **G**, in order of descending chemical shift.*

In experiments with higher enzyme to substrate ratio, another signal was observed at 103.7 ppm, shown in Figure 24, and similar to the abovementioned minor signals, it is unassigned and could arise from a number of species. Compared to the other experiments, the acquisition was started slightly slower and as a results in the first spectrum was acquired a few seconds after mixing.

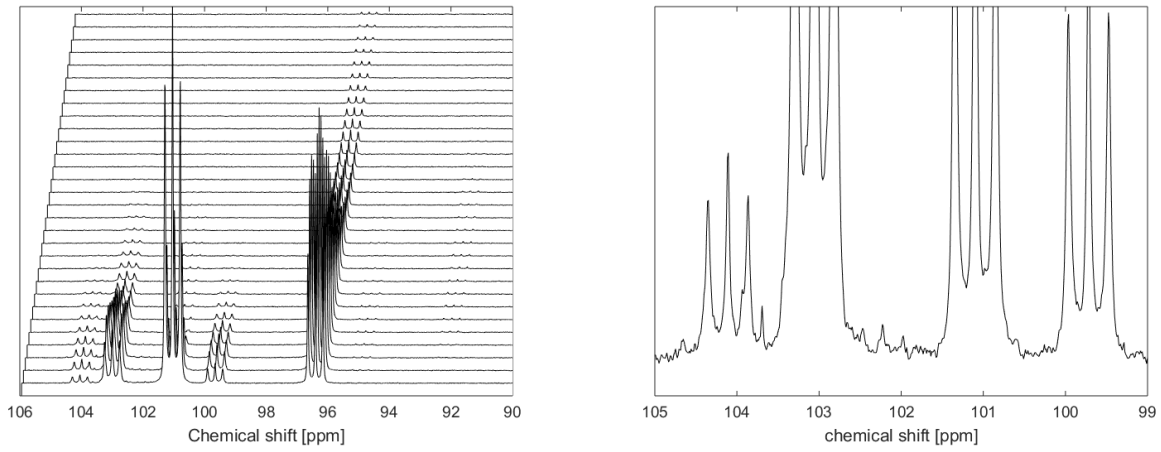


Figure 24 – Left: The first 30 spectra of a sample with higher enzyme to substrate ratio compared to the other described experiments. Right: Zoom of spectrum 3 to show smaller signals.

Lastly, as 5.0 mL buffer was used for dissolution, but only approximately 0.5 mL injected into the NMR tube for the ensuing acquisition, a relatively simple purification method was developed to recycle the unused doubly labelled substrate.

3.1.3 Conclusion

The natural abundance substrate was used to determine approximate conditions for the DNP buildup and subsequent kinetics, but as it turned out, partly because the T_1 of the anomeric position was faster than the quaternary positions of the nitrophenyl, more enzyme and larger flip angles than initially thought were optimal.

Using these optimised conditions, previously unknown transglycosylation activity of the enzyme was observed, and when compared to other analysis methods capable of observing transglycosylation, a far higher temporal resolution was achievable. In order to achieve even better results, several improvements could be made to the experimental setup. For example, an automated transfer system could increase the transfer speed and accuracy for starting the NMR aqcuisition[56]. Furthermore, achieving higher polarisation, either by improving the DNP conditions or by using a magnet of higher field[63], could increase the initial signal strength, and allow for more measurements before the signal decays. Another natural progression of the study could be to synthesise the doubly labelled lactose as substrate instead, and repeat the experiments to determine any transient glucose analogues of the reported species with the natural substrate.

However, even with the relatively low magnetic field of the HyperSense and the manual transfer of sample, which led to numerous variances across repetitions, it was possible to observe previously unknown short-lived and low-populated transglycosylation products of one of the best described glycosidases.

3.1.4 Experimental

D-[1-^{13}C;1-^{2}H]galactose was purchased from Omicron Biochemicals, DMSO from Merck, OX063 from GE Healthcare and all other chemicals were purchased from Sigma Aldrich. The synthesis of the *o*-nitrophenyl-β-D-[1-13C;1-2H]galactopyranoside was done in four steps adapted from literature[156–158] and is described in the supporting information for paper 4. The recycling was performed by first concentrating the sample on a rotary evaporator, redissolving it in acetonitrile and filtering it through a celite plug. Following this procedure, the product was concentrated again and recrystallised from methanol.

DNP buildup was performed using a 3.35 T Hypersense polariser (Oxford Instruments) operating at <1.4 K with a microwave source at 94 GHz. The samples were approximately 1.4M Galp-onp, with or without isotope labelling, dissolved in DMSO, using sonication, with approximately 18 mM OX063 trityl radical added. After polarization for at least 120 minutes, dissolution was performed with 5.0 mL of 0.05 M pH 7.4 phosphate buffer preheated to 175 °C, followed by rapid manual

transfer to the NMR magnet. In the experiments with added acceptor, approximately 1 mmol in 600 µL was mixed in after dissolution, but before transfer to the NMR magnet.

The subsequent NMR acquisition was carried out on a Varian Inova 400 MHz equipped with a 5 mm AUTO DB X probe and started immediately after the transfer was complete, by using 100 scans with 20° flip angles and a repetition time of 2 seconds. The NMR magnet was preshimmed on a phosphate buffer sample at 37 °C. The NMR acquisition was performed using VnmrJ 4.2 (Agilent) and processed with TopSpin 3.5 (Bruker). The data analysis was performed using Matlab R2015a (MathWorks), and the fitting model utilised the Matlab functions ode45 and fmincon.

3.2 Transglycosylation products of GH20 enzymes

This project was carried out in collaboration with Triinu Visnapuu and David Teze of the Birte Svensson group at DTU Bioengineering. The collaboration was part of the OliGram project, which focuses on the design and gram scale enzymatic synthesis of human milk oligosaccharides. My part of the project was to analyse the transglycosylation products. The Paraglaciecola GH20 studies can also be found in paper 5, and the remaining studies are to be published at a later time.

3.2.1 Introduction

Human milk oligosaccharides (HMO) are important prebiotics for boosting infants' immune systems, and numerous beneficial effects have been attributed to HMO[6]. Generally, HMO are constituted of five monosaccharides: Galactose, glucose, *N*-acetylglucosamine, L-fucose and *N*-acetylneuraminic acid, with lactose at the reducing end[159]. The galactose in the non-reducing end of lactose is substituted with β-GlcpNAc on the 3- or 6-position, or both, which is then further substituted predominantly with galactose[160]. As human milk has a much higher concentration and complexity of oligosaccharides compared to bovine milk[161], the large scale production of HMO is of great interest to improve infant formula[162]. For this reason, our collaborators at DTU Bioengineering used three different GH20 glycosidases to perform transglycosylation using different β-GlcpNAc species as donors and lactose as acceptor using the following enzymes:

The *Paraglaciecola hydrolytica*, a marine bacterium recently isolated from eelgrass at Denmark, has been found to contain an array of algal-polysaccharide degrading enzymes[163,164]. The genome of the bacterium was found to contain several genes encoding GH20 *N*-acetyl hexosaminidases (EC 3.2.1.52), which could prove useful for the enzymatic synthesis of the core structures of HMO. Using lactose as a transglycosylation acceptor for a β-*N*-acetylglucosaminidase could be used for the first step in the production of HMO.

Correspondingly, another GH20 β-acetylglucosaminidase isolated from *Bifidobacterium longum*[165], a bacterium present in the gastrointestinal tract, was used for similar experiments. The *Bifidobacterium* species are important regarding HMO degradation as they constitute up to 90% of the infant gastrointestinal tract bacteria population, while for adults it is only a few percent[166–168]. In addition, a third GH20 β-acetylglucosaminidase BbhI originated from *Bifidobacterium bifidum*[169].

3.2.2 Results and discussion

Initially, the transglycosylation reactions were performed using *N*-acetylated chitobiose as donor, but as it proved difficult to separate the products from remaining *N*-acetylated chitobiose, 2-methyl-(1,2-dideoxy-α-D-glucopyrano) [2,1-d]-oxazoline (NAG-oxazoline) was synthesised[170] and used as

a donor. Furthermore, NAG-oxazoline is the enzyme's reactions intermediate, and therefore it is considered an "activated donor" and allows for far superior transglycosylation yields than possible with chitobiose.

3.2.2.1 Paraglaciecola GH20-mediated transglycosylation

The samples of trisaccharidic products prepared with GH20 enzyme from *P. hydrolytica* obtained from our collaborators. As observed by thin layer chromatography (TLC), shown in figure 25, they contained numerous sugars, and three fractions from gel permeation chromatography separation containing trisaccharides were investigated. Each of the samples contained at least two detectable species, but with one of them having a higher concentration, and each section is focused on the major trisaccharide of the sample.

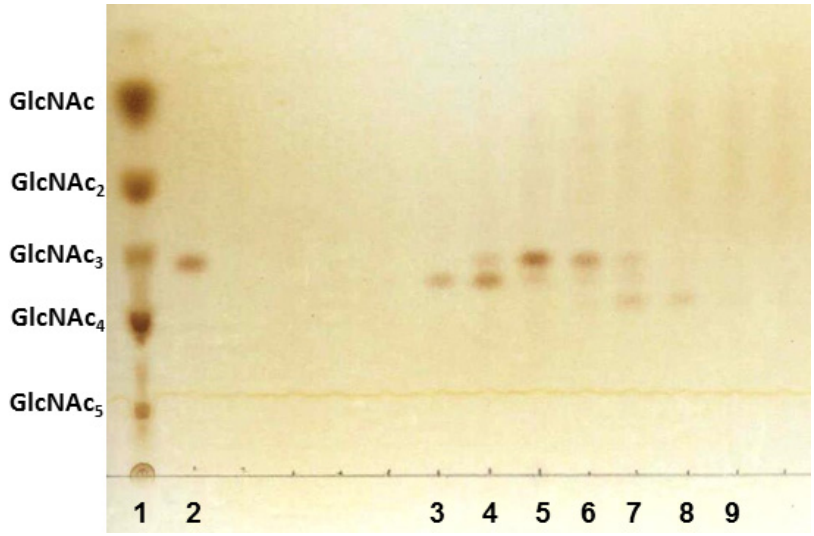


*Figure 25 – TLC analysis of the trisaccharide-containing fractions. COS (lane 1) and LNT2 (lane 2) were used as reference sugars. 3 - fraction 49; **4 - fraction 50**; **5 - fraction 51**; 6 - fraction 52; **7 - fraction 53**; 8 - fraction 54; 9 - fraction 55. Fraction 50, 51 and 53 were used for NMR characterisation of the products.*

3.2.2.1.1 Sample one (P1F50, non-reducing trisaccharide)

The first of the three analysed samples contained two trisaccharides, the same two as in the second sample, but in opposite concentration ratio. In this section the structural elucidation of the major species of this sample will be discussed, whereas the minor species will be discussed in the next section (0). The sample amount was approximately 0.4 mg dissolved in 500 µL D_2O.

From the HSQC spectrum, figure 26, it was evident that the major species in the sample was a non-reducing sugar, as it lacked a reducing end and had upfield β-anomeric ^{13}C chemical shifts[153]. Additionally, the sample contained another trisaccharide, and the ratio between the two was approximately 7:3.

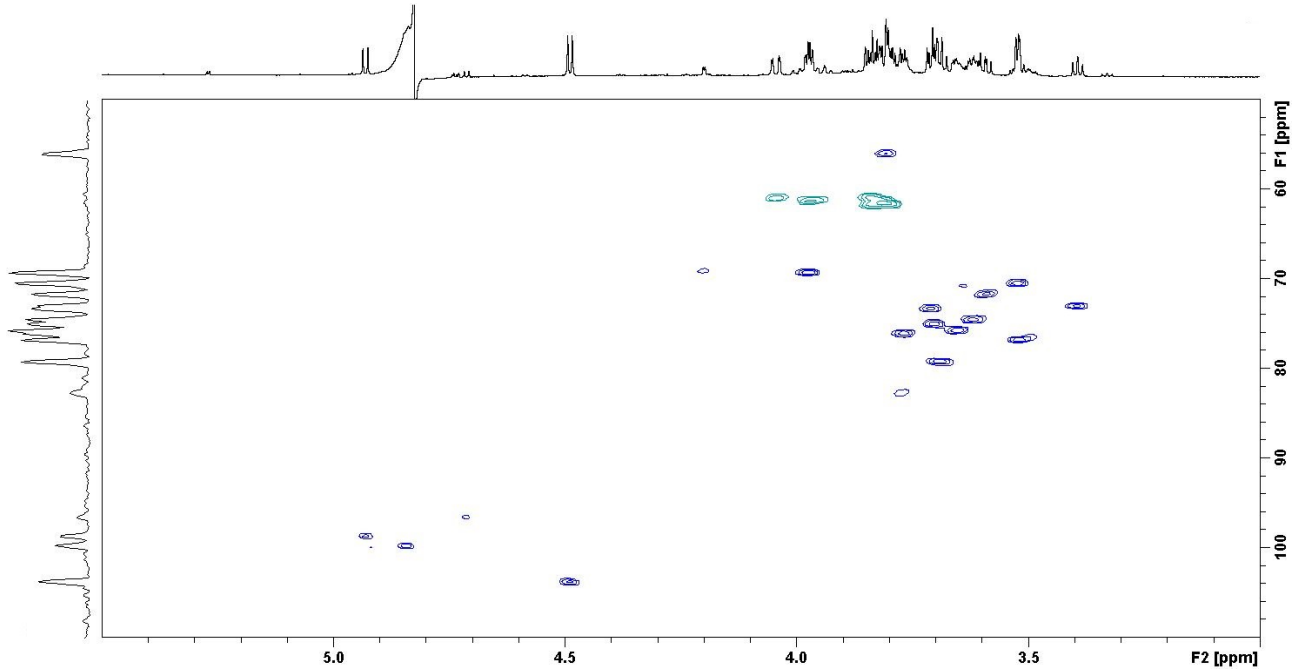
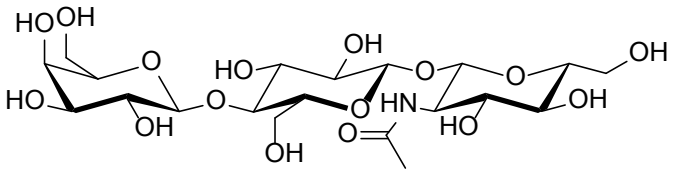


Figure 26 – HSQC spectrum of the P1F50 sample. The 1H projection was from a separate 1H experiment

This made the number of possible transglycosylation products fairly limited, as only the 1-position of the Glc in lactose could act as an acceptor to result in a non-reducing trisaccharide. The $^{3}J^{1}_{H1,}{}^{1}_{H2}$ coupling constant of the β-GlcpNAc could be determined from the ^{1}H spectrum, confirming the expected β-configuration, whereas the anomeric configuration of the glucose had to be verified using DQF-COSY, due to overlap with the water signal. From this experiment it was determined to have the structure shown in scheme 6.



Scheme 6 – The non-reducing trisaccharide isolated in fraction 50.

This structure had previously been synthesised using transglycosylation by GH20 family fungal enzyme, but only assigned as its peracetylated version[171]. A full assignment of the non-reducing trisaccharide can be found intable 10 and a summary of measurable the $^{3}J^{1}_{H,}{}^{1}_{H}$, coupling constants can be found in table 11.

Table 10 – NMR assignment of the non-reducing trisaccharide of the P1F50 samples. The methyl of the acetyl group of the GlcpNAc was at 2.092 ppm for ^{1}H and 22.81 ppm for ^{13}C and the carbonyl of the acetyl was at 176.06 ppm ^{13}C.

P1F50 major product	1	2	3	4	5	6
β-Galp-(1,4-	4.490 103.77	3.590 71.70	3.711 73.36	3.975 69.29	3.768 76.16	3.775/3.845 61.81
4)-β-Glcp-(1,1-	4.843 99.82	3.396 72.98	3.701 75.01	3.692 79.21	3.654 75.78	3.845/4.045 61.02
β-GlcpNAc-(1,1-	4.929 98.80	3.806 56.06	3.619 74.50	3.523 70.43	3.520 76.79	3.813/3.972 61.40

Table 11 – Summary of measurable $^{3}J_{^{1}H,^{1}H}$ coupling constants in the non-reducing trisaccharide from the P1F50 sample.

P1F50 major product	H1-H2	H2-H3	H3-H4	H4-H5	H5-H6a	H5-H6b	H6a-H6b
β-Galp-(1,4	8.05 Hz	10.40 Hz	3.50 Hz	nd	nd	nd	nd
4)-β-Glcp-(1,1	8.44 Hz	9.40 Hz	Nd	nd	7.33 Hz	2.15 Hz	12.42 Hz
β-GlcpNAc-(1,1-	8.53 Hz	9.92 Hz	9.84 Hz	nd	8.34 Hz	nd	12.45 Hz

3.2.2.1.2 Sample two (P1F51, LNT2)

As mentioned in 3.2.2.1.1, this sample contained the same two compounds, here the major species (reducing trisaccharide) was the minor species from before and vice versa. From the HSQC spectrum, shown in figure 27, four anomeric signals of similar intensity could be observed, two of them belonging to a reducing end. The minor species present in the sample was in approximately 1:4 ratio to the major species, and is assigned intable 10. The sample amount was approximately 0.6 mg dissolved in 500 µL D_2O.

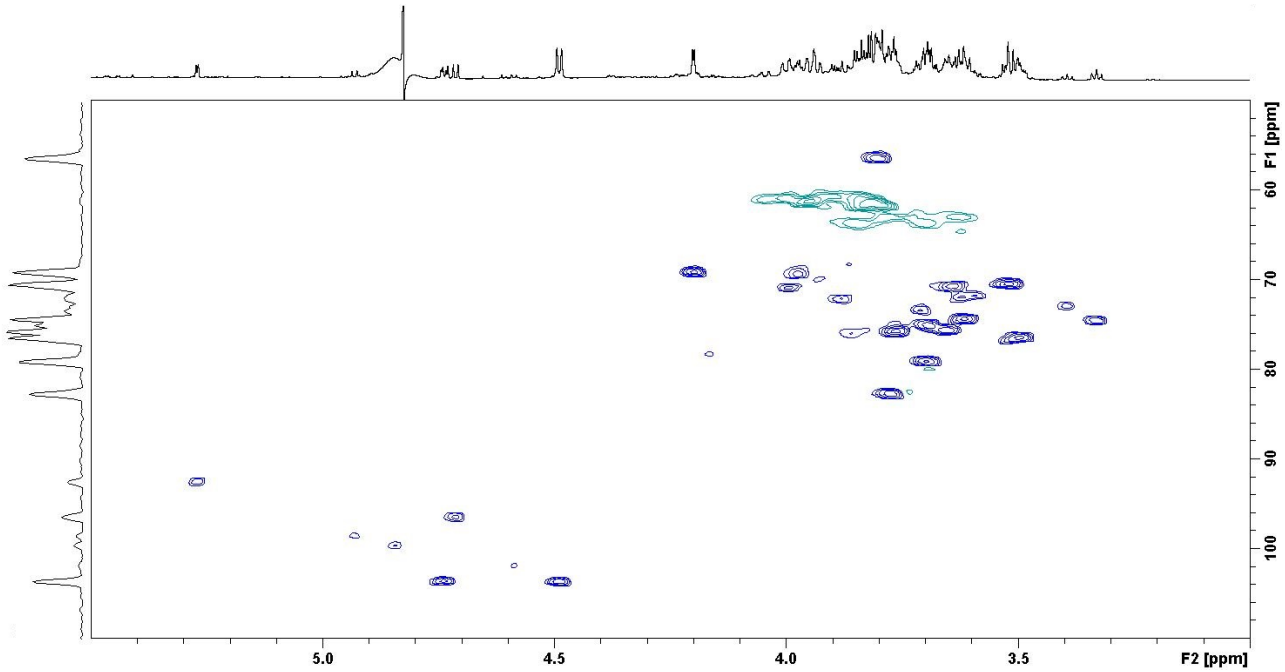


Figure 27 – HSQC spectrum of the P1F51 sample. The ^{1}H projection was from a separate 1H experiment.

Compared to the first product (described in section 3.2.2.1.1), this product had a lot more possible structures. The glucose residue was determined to only be 4-substituted, meaning one of the hydroxyl groups of the galactose was the acceptor. Based on both the ^{13}C chemical shift, in which the 3-position had shifted quite a bit downfield compared to unsubstituted β-galactopyranoside[40], and the HMBC correlation from the anomeric position of the β-GlcpNAc to the 3-position of the β-Galp, it was evident of the species to have the following structure: β-GlcpNAc-1,3-β-Galp-1,4-Glc.

This trisaccharide is commonly known as lacto-*N*-triose II (LNT2), and while commercially available, its synthesis constitute one of the aims of the OliGram project. The full assignment can be found in table 12.

Table 12 – NMR assignment of LNT2 from the P1F51 sample. The methyl of the acetyl group of the GlcpNAc was at 2.088 ppm for ^{1}H and 22.70 ppm for ^{13}C and the carbonyl of the acetyl was at 175.92 ppm ^{13}C

P1F51 (LNT2)	**1**	**2**	**3**	**4**	**5**	**6**
β-GlcNAc-(1,3-	4.736 103.71	3.801 56.23	3.614 74.26	3.518 70.39	3.493 76.35	3.799/3.867 61.60
-3)-β-Galp-(1,4-	4.488 103.71	3.642 70.72	3.773 82.79	4.195 69.11	3.760 75.70	3.778/3.819 62.08
-4)-β-Glcp (red)	4.711 96.46	3.330 74.42	3.690 75.06	3.693 79.08	3.652 75.54	3.848/3.936 62.00
-4)-α-Glcp (red)	5.267 92.60	3.610 71.84	3.877 72.16	3.693 79.19	3.973 69.27	3.848/4.005 62.13

3.2.2.1.3 Sample three (P1F53)

The third sample contained analyte in much smaller concentration, and gave poor SNR even with numerous scans. In figure 28, the anomeric region of the sample is shown, and it contains five anomeric signals of similar intensity, two of them from a reducing end and two of them with similar relative intensity as the two reducing ends. Additionally, this fraction also contains small amounts of LNT2. The sample amount was less than 0.2 mg dissolved in 500 µL D_2O, hence the low concentration and poor SNR.

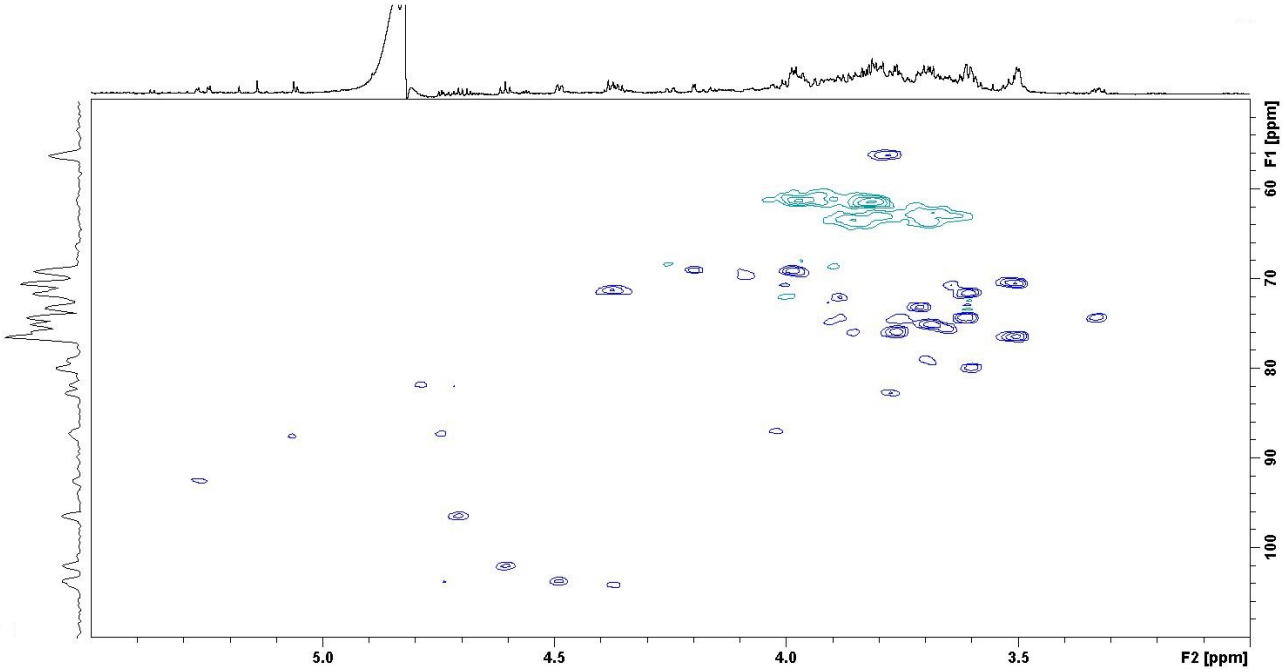


Figure 28 – HSQC spectrum of the P1F53 sample. The ^{1}H projection was from a separate experiment.

Due to the low concentration, and the overlap with signals from LNT2, it was not possible to fully assign the unknown sugar in the sample. However, from the spectra it would seem most likely that it is a trisaccharide, which was supported by the TLC shown in figure 25, where the β-GlcpNAc was transglycosylated to either the 2-, or 3-position of the glucose of the lactose acceptor. This suggestion was made based on the following observations: Firstly, the signals corresponding to the galactose overlapped nicely with the corresponding signals in lactose, which is not the case for LNT2 due to the 3-substitution. Secondly, the signals from the glucose do not overlap with the corresponding signals from lactose, while for LNT2 they do.

This would imply that the β-GlcpNAc is attached to the glucose residue, and as there is nothing indicating a 6- nor 1-substitution, only 2- or 3- substitution is plausible. However, as neither the anomeric signals from the reducing glucose nor from the galactose have shifted significantly compared to the corresponding signals in LNT2 or lactose, it was not evident from those which position was substituted.

Lastly, by comparing with the O-polysaccharide from *Proteus mirabilis* reported by Drzewiecka *et al*. (2010)[172], in which repeating unit of the dephosphorylated polysaccharide was reported as [-3)-β-Galp-1,4-[β-GlcpNAc-1,2-]β-Glcp-1,3-β-GlcpNAc-1-], the β-GlcpNAc would most likely be attached to the 2-position of the Glc. This would also be the best explanation for the two non-reducing signals that have intensities corresponding to the equilibrium between the α- and β-

reducing ends. But even then, the weak signal intensity in the spectra makes the structural model rather uncertain.

3.2.2.2 Bifidobacterium longum GH20 transglycosylation

This sample was prepared in a similar manner as those described above, except that it was purified in a easier manner to simply reduce the lactose concentration rather than separating each produced trisaccharide.

From the HSQC spectrum shown in figure 29, it was evident that a larger number of trisaccharides were present. By comparing the anomeric region with the three samples discussed above, all the compounds observed were present alongside one additional anomeric signal. From the bulk region it was evident that a 1,6-linkage was likely present, due to the downfield methylene signals.

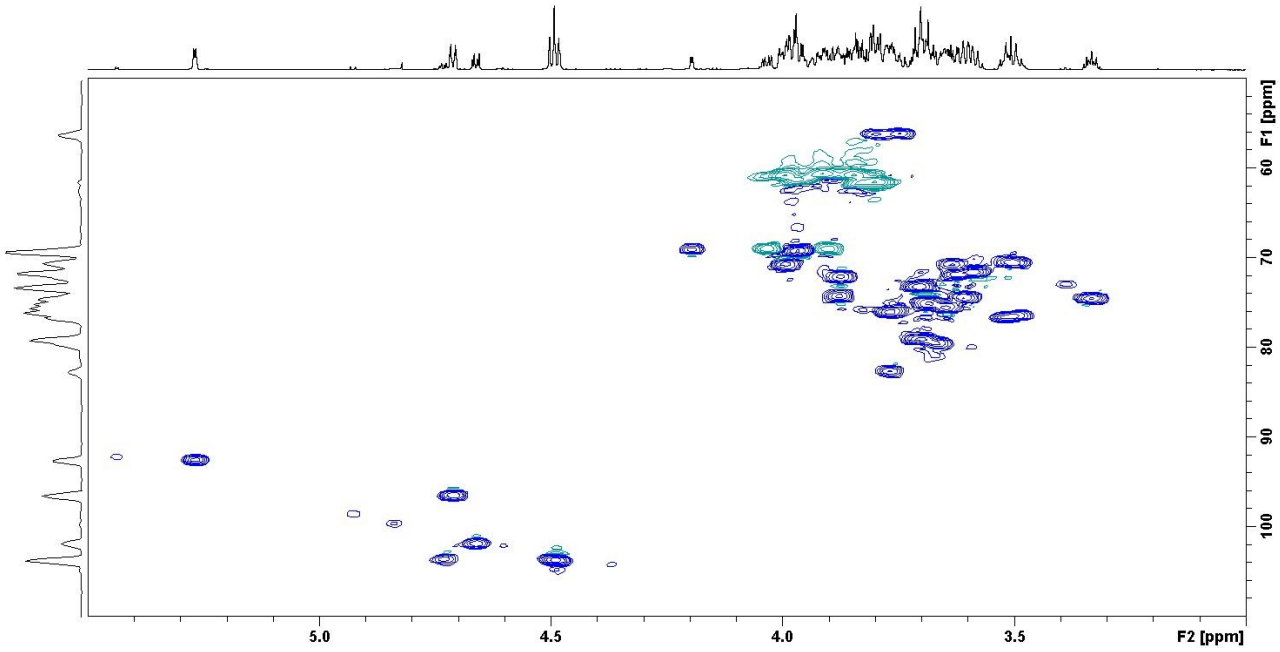


Figure 29 – HSQC spectrum of the Bifidobacterium longum sample. The ^{1}H projection was from a separate experiment.

The unique anomeric signal was determined to be from a β-GlcpNAc connected to the 6-position of the galactose of a lactose, and a full assignment can be found in table 13. The shifts for the β-GlcpNAc were similar to those observed in LNT2, and similarly the glucose and galactose were mostly similar to the shifts of lactose, with the exception of the 6-position of the galactose. The linkage was confirmed by HMBC and ROESY from the anomeric position of the β-GlcpNAc to the downfield 6-position. Whether the 6-position was from galactose or glucose was harder to determine, as the data suffered from severe spectral overlap, but by utilising DQF-COSY, ROESY, HMBC and HSQC-TOCSY it was possible to determine that it belonged to the galactose.

Table 13 – Table of the chemical shifts of the β-GlcpNAc-1,6-β-Galp-1,4-Glc observed in the bifidobacterium longum GH20 transglycosylation products sample. The methyl of the acetyl group of the GlcpNAc was at 2.107 ppm for ^{1}H and 22.84 ppm for ^{13}C and the carbonyl of the acetyl was at 175.37 ppm ^{13}C.

	1	2	3	4	5	6
β-GlcpNAc-(1,6	4.660 101.77	3.752 56.10	3.604 74.44	3.511 70.47	3.511 76.5	3.807/3.967 61.31
6)-β-Galp-(1,4-	4.497 103.69	3.586 71.61	3.703 73.21	3.966 69.16	3.881 74.25	3.898/4.036 68.99
4)-β-Glc (red)	4.711 96.46	3.336 74.54	3.681 75.04	3.698 79.07	3.657 75.46	3.852/3.852 60.76
4)-α-Glc (red)	5.267 92.53	3.624 71.85	3.875 72.09	3.701 79.03	3.991 70.68	3.848/3.987 60.76

Based on the integrals of the anomeric peaks obtained from the HSQC spectrum, the sample contained approximately 18% LNT2, 37% of the trisaccharide assigned in table 13, 3% of the trisaccharide described in section 0, 4% of the non-reducing trisaccharide shown in scheme 6 and 38% lactose. The sample might contain additional sugars in low concentration that were not detected, or were indistinguishable due to spectral overlap.

3.2.2.3 Bifidobacterium bifidum GH20 (BbhI) transglycosylation

This sample contained mostly LNT2 (>95% from NMR integrals), see figure 30 for HSQC, the assignment of which can be found in table 12. This translgycosylation product has been reported previously for the enzyme[173].

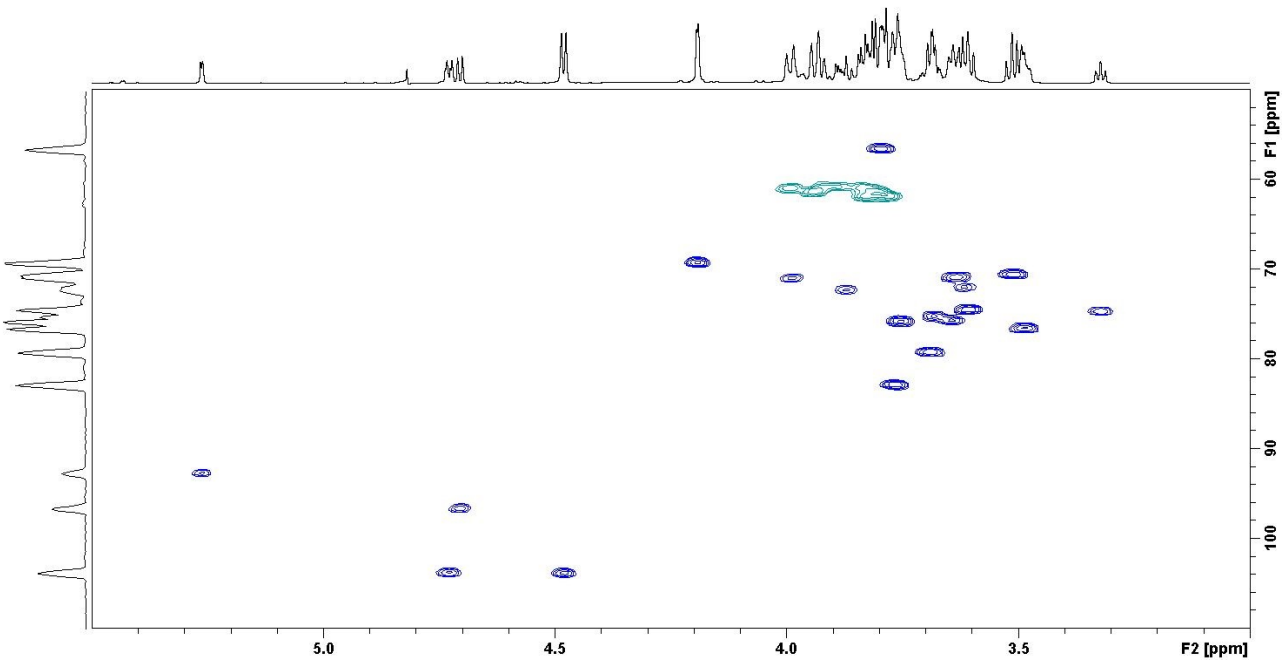


Figure 30 – HSQC spectrum of the Bifidobacterium bifidum sample. The ^{1}H projection was from a separate experiment.

3.2.3 Conclusion

It is currently impossible to infer from a GH20 sequence if it will catalyse transglycosylation reactions, let alone the regioselectivity of the possibly formed bonds. Unfortunately, no efficient screening procedures has been developed, and the only available method is the rather labor-intensive production and purification of each enzyme candidate, its use for transglycosylation reactions, followed by the purification and identification of each formed product. Usually, these last steps are performed by peracetylating all the products formed, purification of the obtained compounds, and the use of NMR to determine the regioselectivity - the stereoselectivity being generally assured by the enzyme.

In this work, the identification of products from a mixture and at milligram scale allowed a significantly faster pace. In the context of the OliGram project, it showed that the *B. longum* and *P. hydrolytica* enzymes are not suitable for the production of free LNT2. However, as the results show that with the *P. hydrolytica* enzyme, the only other bond than the desired 1,3 bond on the Gal residue are formed on the Glc moiety, it opens up the possibility to use this enzyme for transglycosylation in which the Glc is covalently bound to bulky groups. This is particularly relevant as the HMO are often multipresented on mucins or other glycoproteins, to which they are covalently bound by the Glc of their lactose part.

More importantly, this work proved that LNT2 is produced from BbhI at the required level of purity. The production of this compound had been recently reported[173], but without providing any characterisation besides HPLC retention times. The OliGram partners have thus started to work on this enzyme, particularly aiming at improving the existing low yields.

3.2.4 Experimental

3.2.4.1 NMR spectroscopy

All spectra were recorded on an 800 MHz Bruker Avance III (799.75 MHz for ^{1}H and 201.10 MHz for ^{13}C) equipped with a 5 mm TCI cryoprobe. Acetone was used as internal reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively. The following experiments were used for the elucidations: 1H with presaturation, DQF-COSY, ROESY, HSQC, HSQC-TOCSY and HMBC, and they were all carried out under similar conditions as those described in section 2.3.

3.2.4.2 Collaborators experimental work in short

The NAG-oxazoline synthesis was adapted from André-Miral et al. (2014)[170]. Lacto-*N*-triose II was obtained from Elicityl Oligotech (France) and *N*-acetylated chitooligosaccharides (COS) from Koyo Chemicals (Japan).

3.2.4.2.1 Cloning, expression and purification of *P. hydrolytica* Nah20A

β-*N*-acetyl hexosaminidase GH20 gene was isolated from the genomic DNA of *P. hydrolytica* using specific primers. The gene was cloned into pURI3TEV expression vector by PCR-based cloning. Overexpression of the gene was carried out in *E. coli* strain BL21(DE3) harbouring the expression plasmid. The transformant was grown in lysogeny broth (LB) medium containing ampicillin and Nah20A protein production was induced using isopropyl β-D-1-thiogalactopyranoside (IPTG). *N*-terminally His-tagged recombinant Nah20A was purified from the cell lysate with Ni-affinity chromatography on HisTrapHP column and size-exclusion chromatography on HiLoad 16/60 Superdex 200 pg column coupled with ÄKTA Avant chromatography system (GE Healthcare). The fractions containing enzymatically active and sufficiently pure Nah20A were pooled, concentrated, filtered and stored in Na-phosphate buffer (50 mM, pH 7) containing 0.3 M NaCl at 4°C.

3.2.4.2.2 Expression and purification of BbhI and GH20 from *B. longum*

Expression plasmids with β-acetylglucosaminidases from *B. longum* and *B. bifidum* were obtained from Prof. Takane Katayama (Kyoto, Japan). The plasmids contained kanamycin and ampicillin resistance genes, respectively, as selection markers. Overexpression and purification of the proteins was carried out similarly as shown above (previous section).

3.2.4.2.3 Enzymatic synthesis and purification of transglycosylation products

Purified proteins were used as catalysts to synthesise transglycosylation products. The reaction mixture contained 100 mM NAG-oxazoline and 200 mM lactose in Na-phosphate buffer (50 mM, pH 8). The reaction of Nah20A contained additionally 0.5% bovine serum albumin (BSA) to enhance the stability of the protein. Amount of the protein in the mixtures was 10 activity units per mL (116 µg of Nah20A, 263 µg of BbhI, 93 µg of *B. longum* GH20). Proteins were reacted 2 h at 37°C and reaction was stopped with heating of the mixture 5 min at 90°C. The samples were diluted in milliQ water and protein was removed from the sample.

Separation of transglycosylation products was conducted using gel permeation chromatography on Bio-Gel P-2 (Bio-Rad) column coupled with ÄKTAprime plus chromatography system (GE Healthcare). Approximately 10 mg of sugar was applied and milliQ water was used as an eluent at flow rate of 0.1 ml/min at room temperature. The presence of reducing sugars in the fractions was detected. Fractions were dried, dissolved in milliQ water and analysed on TLC. For NMR analysis, the trisaccharide-containing fractions were dried and dissolved in 0.5 mL D_2O.

TLC was performed on Silica Gel 60 F_{254} plates (Merck). 1 µL of the samples were spotted onto the plates and sugars were separated with two runs in *n*-butanol: ethanol: water (5:3:2; vol/vol/vol). Sugars were visualised by spraying the plates with orcinol dye followed by heating.

3.3 Blood group B galactosyltransferase

This work was part of my external stay with Thomas Peters and his group at the University of Lübeck in the autumn of 2015.

3.3.1 Introduction

α-1,3-Galactosyltransferase (GTB) is one of the two enzymes responsible for the final step in the enzymatic synthesis of the ABO(H) blood group A and B antigens[174], the other enzyme being α-1,3-*N*-acetylgalactosaminyltransferase (GTA). The two enzymes are members of the GT6 family, and only differ in four of their otherwise identical 354 amino acid residues, namely Arg or Gly at position 176, Gly or Ser at position 235, Leu or Met at position 266 and Gly or Ala at position 268 for GTA or GTB, respectively, and these differences dictate whether the enzyme is specific for UDP-GalpNAc or UDP-Galp as donor[3,175–178].

The enzyme mechanisms have been the objects of much attention, including NMR studies[179–181], trapping of covalent intermediates[182] and computational studies[183,184]. The mechanism for the enzyme class is still undetermined, and originally it was thought to have a similar mechanism to the double displacement mechanism of many retaining glycoside hydrolases. However, in many retaining glycosyltransferases, including GTA and GTB, the base catalyst is distant from the substrates, making a double displacement mechanism implausible. This has led to the theory that an S_Ni mechanism might be more realistic[3], and in order to gain more insight to the mechanism of GTA and GTB, and about retaining glycosyltransferases of this type in general, mutations of specific residues as well as specific isotope labelling can be used in conjunction with NMR spectroscopy. In order to study the enzymology of GTA and GTB, experiments were carried out as discussed in the following sections.

Unlike in carbohydrates, nitrogen is a major part of proteins, and as such ^{15}N NMR spectroscopy and ^{15}N labelling is commonly used in protein NMR[185,186].

3.3.2 Results and discussion

Expression of different enzyme constructs were performed for multiple purposes. The first construct, the E303M mutant, was expressed in an attempt to trap a bound intermediate, while the remaining expressions, with different isotopic labelling schemes during expression, was made with the intent of being used for NMR titration studies.

3.3.2.1 GTB-E303M expression in BL21

E303 is believed to be crucial for the mechanism, and when mutated to E303C Soya *et al.*[182] trapped a covalent intermediate. Based on these findings it would be interesting to use other residues for similar experiments, hence the attempt to express GTB-E303M in BL21 was pursued. However, the expression gave very low yields and using an activity assay, little to no activity was observed. A hypothesis as to why the construct had poor expression yield and very low activity, was that it possibly could not fold correctly with the methionine in the active center.

3.3.2.2 Auxotrophic strains for specific labelling and unlabelling

Auxotrophic strains can be useful to obtain specific labelling, and for this purpose, the following two *E. coli* strains were used:

- ML8, which is resistant to kanamycin and auxotrophic for arginine.
- ML40K1, which is resistant to chloramphenicol and auxotrophic for isoleucine, valine, histidine, arginine, methionine and lysine.

During the initial growth of the bacteria in LB media for glycerol stocks, slow growth was observed for ML40K1, but in TB media both of the auxotrophic strains were observed to overexpress GTB-wt, although ML40K1 still grew much at a much slower rate.

3.3.2.2.1 ML8 expressions

ML8 was used for two separate expressions, the first of which was for the expression of GTB-wt with ^{15}N-Arg labelling. The cells were grown in minimal media, and L-arginine-$^{15}N_4$ hydrochloride was added to the mixture 50 minutes before the IPTG. The activity assay showed that it had similar activity to natural abundance GTB, but the NMR results, see figure 31, showed that the targeted labelling had not worked as intended. Rather than having arginine as the only labelled species, the labelling had been incorporated into the entire enzyme, but non-uniformly, making it rather useless for titration.

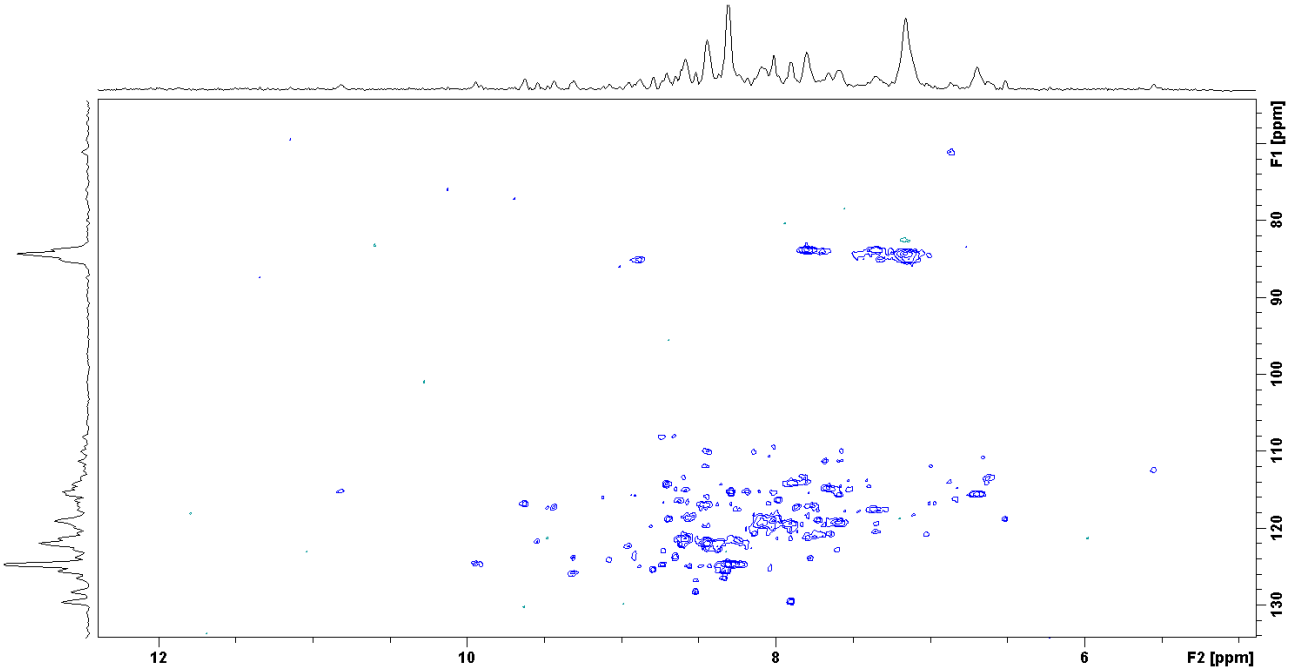
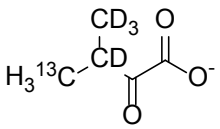


Figure 31 – Transverse relaxation optimized spectroscopy (TROSY) spectrum of the ML8 expressed GTB grown on minimal media with ^{15}N-Arg added.

The second of the two ML8 expressions was done with labelled minimal media and natural abundance arginine was added an hour before IPTG. Furthermore, 2-Keto-3-(methyl-d3)-butyric acid-4-^{13}C,3-d (labelled α-ketoisovalerate, see scheme 7) was added one hour before induction as an isotopically labelled precursor for the methyl groups in valine and leucine[187]. At induction the cells were cold shocked in an ice/water bath for 10 minutes and then further shaken at 17 °C for several days. Based on internally shared results, it had previously been observed that expressions at lower temperature would incorporate the labelling into the valine only.



Scheme 7 – 2-Keto-3-(methyl-d3)-butyric acid-4-^{13}C,3-d (racemic)

However, the cells grew very slowly at 17 °C, hence the need to conduct expression over several days, but even then little to no GTB could be detected by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The exact cause of the lack of expression was never determined, but it is likely due to a low expression at the reduced temperature, which had not previously been attempted with ML8 within the group.

3.3.2.2.2 ML40K1 expressions

As ML40K1 is auxotrophic for six amino acids, it should be possible to express enzymes with selective labelling on either of the amino acids. However, initial results when growing the cells in non-deuterated minimal media were rather unsatisfactory, as the cells only grew to a low cell density, shown in the top row of table 14. It was decided to try additional seven attempts at growing the cells, to determine if any of the six amino acids were the bottleneck: One, where all six of the amino acids that the bacteria was auxotrophic for were added, and six with five each, the results of which are summarised in table 14.

Table 14 – Summary of the 8 expressions with ML40K1 in non-deuterated minimal media.

	OD_{600} at induction	OD_{600} after 1 hour	OD_{600} after 15 hours
None added	0.50	0.57	0.58
All 6 added	0.60	0.68	0.75
No Ile added	0.57	0.69	0.79
No Val added	0.63	0.73	0.78
No His added	0.57	0.71	0.74
No Arg added	0.61	0.69	0.50
No Met added	0.64	0.77	0.59
No Lys added	0.64	0.75	0.62

Compared to the ML8 grown in minimal media with ^{15}N-Arg added, which reached an OD600 of 2.3 15 hours after incubation, all these results were unsatisfactory for labelled expression. From SDS-PAGE analysis, it was evident that all the cultures expressed the protein, but due to the low growth, even with added amino acids, the expression of isotopically labelled proteins with this method would not seem viable.

3.3.3 Conclusion

None of the expressions worked as intended, presumably because of the problems with

- Folding in the case of GTB-E303M expressed in BL21
- Backbone scrambling with the ^{15}N-Arg labelling
- Little to no expression at lower temperatures when expressing in ML8
- Poor growth of ML40K1 in minimal media

3.3.4 Experimental

3.3.4.1 GTB expresion

GTB-E303M expression in BL21 was performed by growing the cells in terrific broth (TB) medium with M9 salts and supplements overnight at 37 °C. This culture was used to inoculate a 1 L culture in TB medium to an OD_{600} of 0.1 and grown to a OD_{600} of 0.6, at which point 1 mL of 1 M IPTG was added and the culture was shaken overnight before the cells were harvested.

The ML8 GTB-^{15}N-Arg culture were grown in TB medium with M9 salts and supplements at 37 °C overnight. The culture was used to inoculate 20 mL culture of deuterated minimal media to an OD_{600} of 0.1, which was shaken at 37 °C until it reached an OD_{600} of 0.35. This culture was then used to inoculate 50 mL culture of deuterated minimal medium to an OD_{600} of 0.1, which was then gradually enlarged to a volume of 280 mL while growing at 37 °C. After reaching an OD_{600} of 0.28, 100 mg of L-arginine-$^{15}N_4$ hydrochloride was added, and after an additional 50 minutes, 280 µL of 1 M IPTG in D_2O were added and the culture shaken overnight at 37 °C, after which the cells had reached an OD_{600} of 2.3 and were harvested.

The other ML8 and ML40K1 cultures grown on minimal media, deuterated or not, were grown in a similar manner, with the following exceptions:

The other ML8 culture had 32 mg of labelled ketoisovalerate added alongside 100 mg arginine instead of the labelled arginine, and afterwards it was shaken for an hour at 37 °C before it was cold shocked in an ice/water bath for 10 minutes and 280 µL of 1M IPTG was added. Subsequently, the culture was shaken at 17 °C for 67 hours after which the cells had reached an OD_{600} of 1.9 and the cells were harvested.

The ML40K1 expressions were grown in non-deuterated minimal medium with added amino acids to give a doubling in concentration compared to the Silantes OD2 media, shown in table 15, in accordance with the eight experiments described in table 14. After induction, the samples were shaken overnight at 37 °C to the OD_{600} densities described in table 14 before being harvested.

Table 15 – Summary of the concentration of the six amino acids of interest in the Silantes OD2 media

	Ile	Val	His	Arg	Met	Lys
µmol/L	10.3	14.58	47.18	11.06	8.7	16.52

Enzyme purification, protein concentration determination as well as enzymatic activity were determined as described previously[188,189].

3.3.4.2 NMR spectroscopy

The TROSY[190] spectrum were acquired at 25 °C using 2048x256 with 200 scans, and a sweep width of 7500.0 and 3544.6 Hz in F2 and F1, respectively. It was measured on a Bruker 500 MHz spectrometer (499.67 in ^{1}H and 50.63 in ^{15}N) equipped with a 5 mm TCI cryogenic probe in a 3 mm NMR tube. All spectra were recorded using TopSpin 3.1 and processed using 3.5.

4 Overall conclusions and perspectives

This thesis describes several smaller projects divided into two chapters, one concerning polysaccharide structural assignments and one concerning enzymatic carbohydrate synthesis. Challenges like sample concentration, purity and carbohydrate complexity and heterogeneity were overcome, and serves as an example that using high-field NMR spectroscopy it is possible to identify and elucidate various types of complex oligo- and polysaccharide structures.

The *Streptococcus* project aimed to elucidate the structure of the novel serotype 7D capsular polysaccharide, and as part of the work, the 7C cps was also elucidated. The cps of the novel serotype turned out to be a mixture of the cps of 7C and 7B, and while such hybrid structures have been observed previously, the implications of this in terms of virulence have not been determined. In parallel with this, a similar project concerning serogroup 24 was also started, but it turned out all the cps had poor NMR properties, presumably because they were too large, resulting in very short T_2 relaxation times, and was never completed. To continue this project, the acid hydrolysis methods used for the LY03 and 1187 polysaccharides in section 2.3 could be used to try to produce oligosaccharides or smaller polysaccharides containing the repeating unit. Additionally, there are still plenty of other serotypes where the cps structure have not been assigned[4], and similar projects could be carried out concerning those.

The chaga project elucidated many of the structural properties of the extracted polysaccharides. Further analysis of the fungal polysaccharides immunomodulating or other beneficial effects could be performed. Another improvement could perhaps have been additional analyses of the sample obtained from extraction under basic conditions, which could have been subjected to acid hydrolysis to reduce the molecular size, but due to its heterogeneous nature, much structural information could have been lost.

In the β-lactoglubolin project, the identified oligosaccharide structures could be used alongside the protein NMR data to perform *in silico* docking studies. Additionally, as the binding has only been studied on the protein to identify binding patches on the surface, additional information could be gained by studying the oligosaccharides interactions with the protein, for example by using saturation transfer difference (STD) NMR[191,192], but this would likely require higher amounts of oligosaccharide to be produced, and a major challenge in this project was obtaining enough oligosaccharide for the NMR studies.

In the OliGram project any future work seems rather obvious. The goal of the project was gram scale synthesis of HMO, and as this part of the project was only concerned with the first step in the synthesis, much additional work would be required, including characterisation of future products. As the oligosaccharides grow in size, they would only become more challenging to analyse, and with more hydroxyl groups present to act as acceptor, more potential products could be formed, increasing the need for NMR spectroscopy.

The dDNP project showed that even though carbohydrates have some of the worst possible properties for dDNP, novel reaction intermediates of one of the best described carbohydrate converting enzymes could be observed. The project could likely have been improved in several ways, some of which are discussed in section 3.1. In short, improving the polarisation and transfer system could greatly increase SNR, and especially in the first few seconds where most of the transient species were detected. As this was meant more as a proof of concept to show that detection of transient species during enzymatic carbohydrate conversion, additional dDNP NMR studies of similar enzymes could be used to further elucidate the mechanisms of the conversion.

5 References

(1) Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* **1996**, *96*, 683–720.

(2) Juers, D. H.; Matthews, B. W.; Huber, R. E. LacZ β-Galactosidase: Structure and Function of an Enzyme of Historical and Molecular Biological Importance. *Protein Sci.* **2012**, *21*, 1792–1807.

(3) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555.

(4) Geno, K. A.; Gilbert, G. L.; Song, J. Y.; Skovsted, I. C.; Klugman, K. P.; Jones, C.; Konradsen, H. B.; Nahm, M. H. Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin. Microbiol. Rev.* **2015**, *28*, 871–899.

(5) Khatun, F.; Stephenson, R. J.; Toth, I. An Overview of Structural Features of Antibacterial Glycoconjugate Vaccines That Influence Their Immunogenicity. *Chem. - A Eur. J.* **2017**, *23*, 4233–4254.

(6) Bode, L. Human Milk Oligosaccharides: Every Baby Needs a Sugar Mama. *Glycobiology* **2012**, *22*, 1147–1162.

(7) Bode, L. Human Milk Oligosaccharides: Prebiotics and beyond. *Nutr. Rev.* **2009**, *67*, 183–191.

(8) Bode, L.; Contractor, N.; Barile, D.; Pohl, N.; Prudden, A. R.; Boons, G. J.; Jin, Y. S.; Jennewein, S. Overcoming the Limited Availability of Human Milk Oligosaccharides: Challenges and Opportunities for Research and Application. *Nutr. Rev.* **2016**, *74*, 635–644.

(9) Zhang, M.; Cui, S. W.; Cheung, P. C. K.; Wang, Q. Antitumor Polysaccharides from Mushrooms: A Review on Their Isolation Process, Structural Characteristics and Antitumor Activity. *Trends Food Sci. Technol.* **2007**, *18*, 4–19.

(10) Borchers, A. T.; Stern, J. S.; Hackman, R. M.; Keen, C. L.; Gershwin, M. E. Mushrooms, Tumors, and Immunity. *Proc. Soc. Exp. Biol. Med.* **1999**, *4*, 281–293.

(11) Saar, M. Fungi in Khanty Folk Medicine. *J. Ethnopharmacol.* **1991**, *31*, 175–179.

(12) Kim, Y. Immunomodulatory Activity of the Water Extract from Medicinal Mushroom Inonotus Obliquus. *Mycobiology* **2005**, *33*, 158–162.

(13) Duus, J. Ø.; Gotfredsen, C. H.; Bock, K. Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations. *Chem. Rev.* **2000**, *100*, 4589–4614.

(14) Dwek, R. A.; Edge, C. J.; Harvey, D. J.; Wormald, M. R. Analysis of Glycoprotein-Associated Oligosaccharides. *Annu. Rev. Biochem.* **1993**, *62*, 65–100.

(15) Yang, L.; Zhang, L. M. Chemical Structural and Chain Conformational Characterization of Some Bioactive Polysaccharides Isolated from Natural Sources. *Carbohydr. Polym.* **2009**, *76*, 349–361.

(16) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **1997**, *62*, 7512–7515.

(17) Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29*, 2176–2179.

(18) Jeener, J. Lecture at Ampère International Summer School, Basko Polje, (Proposal). In; 1971.

(19) Aue, W. P.; Bartholdi, E.; Ernst, R. R. Two-Dimensional Spectroscopy. Application to Nuclear Magnetic Resonance. *J. Chem. Phys.* **1976**, *64*, 2229–2246.

(20) Piantini, U.; Sørensen, O. W.; Ernst, R. R. Multiple Quantum Filters for Elucidating NMR Coupling Networks. *J. Am. Chem. Soc.* **1982**, *104*, 6800–6801.

(21) Braunschweiler, L.; Ernst, R. R. Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy. *J. Magn. Reson.* **1983**, *53*, 521–528.

(22) Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **1953**, *92*, 411–415.

(23) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by Two-Dimensional NMR Spectroscopy. *J. Chem. Phys.* **1979**, *71*, 4546–4553.
(24) Bothner-By, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **1984**, *106*, 811–813.
(25) Bodenhausen, G.; Ruben, D. J. Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett.* **1980**, *69*, 185–189.
(26) Doddrell, D. M.; Pegg, D. T.; Bendall, M. R. Distortionless Enhancement of NMR Signals by Polarization Transfer. *J. Magn. Reson.* **1982**, *48*, 323–327.
(27) Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. Gradient Selection in Inverse Heteronuclear Correlation Spectroscopy. *Magn. Reson. Chem.* **1993**, *31*, 287–292.
(28) Enthart, A.; Freudenberger, J. C.; Furrer, J.; Kessler, H.; Luy, B. The CLIP/CLAP-HSQC: Pure Absorptive Spectra for the Measurement of One-Bond Couplings. *J. Magn. Reson.* **2008**, *192*, 314–322.
(29) Davis, D. G. Enhanced Correlations in 1H-Detected Heteronuclear Two-Dimensional Spectroscopy via Spin-Locked H Magnetization Transfer. *J. Magn. Reson.* **1989**, *84*, 417–424.
(30) Bax, A.; Subramanian, S. Sensitivity-Enhanced Two-Dimensional Heteronuclear Shift Correlation NMR Spectroscopy. *J. Magn. Reson.* **1986**, *67*, 565–569.
(31) Bax, A.; Summers, M. F. 1H and 13C Assignments from Sensitivity-Enhanced Detection of Heteronuclear Multiple-Bond Connectivity by 2D Multiple Quantum NMR. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
(32) Nyberg, N. T.; Duus, J. Ø.; 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.* **2005**, *127*, 6154–6155.
(33) Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. Editing of H2BC NMR Spectra. *Magn. Reson. Chem.* **2005**, *43*, 971–974.
(34) CASPER - Main Page URL: http:/www.casper.organ.su.se/casper/ (accessed Nov 12, 2017).
(35) 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.* **2006**, *341*, 1003–1010.
(36) Lundborg, M.; Widmalm, G. Structural Analysis of Glycans by NMR Chemical Shift Prediction. *Anal. Chem.* **2011**, *83*, 1514–1517.
(37) Bock, K.; Thøgersen, H. Nuclear Magnetic Resonance Spectroscopy in the Study of Mono- and Oligosaccharides. *Annu. reports NMR Spectrosc.* **1982**, *13*, 1–57.
(38) Bock, K.; Pedersen, C. A Study of 13CH Coupling Constants in Hexopyranoses. *J. Chem. Soc., Perkin Trans. 2* **1974**, 293–297.
(39) Bock, K.; Pedersen, C. Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 27–66.
(40) Agrawal, P. K. NMR Spectroscopy in the Structural Elucidation of Oligosaccharides and Glycosides. *Phytochemistry* **1992**, *31*, 3307–3330.
(41) Schwalbe, H. New 1.2 GHz NMR Spectrometers— New Horizons? *Angew. Chemie Int. Ed.* **2017**, *56*, 10252–10253.
(42) Carver, T. R.; Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **1953**, *92*, 212–213.
(43) Ardenkjær-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K. Increase in Signal-to-Noise Ratio of > 10,000 Times in Liquid-State NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 10158–10163.
(44) Hall, D. A.; Maus, D. C.; Gerfen, G. J.; Inati, S. J.; Becerra, L. R.; Dahlquist, F. W.; Griffin, R. G. Polarization-Enhanced NMR Spectroscopy of Biomolecules in Frozen Solution. *Science* **1997**, *276*, 930–932.
(45) Green, R. A.; Adams, R. W.; Duckett, S. B.; Mewis, R. E.; Williamson, D. C.; Green, G. G. R.

The Theory and Practice of Hyperpolarization in Magnetic Resonance Using Parahydrogen. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *67*, 1–48.
(46) Dumez, J. N. Perspectives on Hyperpolarised Solution-State Magnetic Resonance in Chemistry. *Magn. Reson. Chem.* **2017**, *55*, 38–46.
(47) Golman, K.; In't Zandt, R.; Lerche, M. H.; Pehrson, R.; Ardenkjær-Larsen, J. H. Metabolic Imaging by Hyperpolarized 13C Magnetic Resonance Imaging for in Vivo Tumor Diagnosis. *Cancer Res.* **2006**, *66*, 10855–10860.
(48) Keshari, K. R.; Wilson, D. M.; Chen, A. P.; Bok, R.; Larson, P. E. Z.; Hu, S.; Van Criekinge, M.; Macdonald, J. M.; Vigneron, D. B.; Kurhanewicz, J. Hyperpolarized [2-13C]-Fructose: A Hemiketal DNP Substrate for in Vivo Metabolic Imaging. *J. Am. Chem. Soc.* **2009**, *131*, 17591–17596.
(49) Warburg, O. On the Origin of Cancer Cells. *Science* **1956**, *123*, 309–314.
(50) Vander Heiden, M. G.; Cantley, L. C.; Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **2009**, *324*, 1029–1033.
(51) Hurd, R. E.; Yen, Y.-F.; Chen, A.; Ardenkjaer-Larsen, J. H. Hyperpolarized 13C Metabolic Imaging Using Dissolution Dynamic Nuclear Polarization. *J. Magn. Reson. Imaging* **2012**, *36*, 1314–1328.
(52) Nelson, S. J.; Kurhanewicz, J.; Vigneron, D. B.; Larson, P. E. Z.; Harzstark, A. L.; Ferrone, M.; van Criekinge, M.; Chang, J. W.; Bok, R.; Park, I.; et al. Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-^{13}C]pyruvate. *Sci. Transl. Med.* **2013**, *5*, 198ra108.
(53) Keshari, K. R.; Wilson, D. M. Chemistry and Biochemistry of 13C Hyperpolarized Magnetic Resonance Using Dynamic Nuclear Polarization. *Chem. Soc. Rev.* **2014**, *43*, 1627–1659.
(54) Jensen, P. R.; Meier, S.; Ardenkjær-Larsen, J. H.; Duus, J. Ø.; Karlsson, M.; Lerche, M. H. Detection of Low-Populated Reaction Intermediates with Hyperpolarized NMR. *Chem. Commun. (Camb).* **2009**, 5168–5170.
(55) Bohndiek, S. E.; Kettunen, M. I.; Hu, D. E.; Kennedy, B. W. C.; Boren, J.; Gallagher, F. A.; Brindle, K. M. Hyperpolarized [1-13C]-Ascorbic and Dehydroascorbic Acid: Vitamin C as a Probe for Imaging Redox Status in Vivo. *J. Am. Chem. Soc.* **2011**, *133*, 11795–11801.
(56) Bowen, S.; Hilty, C. Time-Resolved Dynamic Nuclear Polarization Enhanced NMR Spectroscopy. *Angew. Chemie Int. Ed.* **2008**, *47*, 5235–5237.
(57) Bowen, S.; Hilty, C. Temporal Chemical Shift Correlations in Reactions Studied by Hyperpolarized Nuclear Magnetic Resonance. *Anal. Chem.* **2009**, *81*, 4543–4547.
(58) Hilty, C.; Bowen, S. Applications of Dynamic Nuclear Polarization to the Study of Reactions and Reagents in Organic and Biomolecular Chemistry. *Org. Biomol. Chem.* **2010**, *8*, 3361–3365.
(59) Zeng, H.; Lee, Y.; Hilty, C. Quantitative Rate Determination by Dynamic Nuclear Polarization Enhanced NMR of a Diels-Alder Reaction. *Anal. Chem.* **2010**, *82*, 8897–8902.
(60) Lee, Y.; Heo, G. S.; Zeng, H.; Wooley, K. L.; Hilty, C. Detection of Living Anionic Species in Polymerization Reactions Using Hyperpolarized NMR. *J. Am. Chem. Soc.* **2013**, *135*, 4636–4639.
(61) Lee, Y. Hyperpolarized NMR Analysis of Enzymatic Reaction: Extension of Observable Reaction Time by Deuterium Isotope Labeling. *Bull. Korean Chem. Soc.* **2016**, *37*, 810–814.
(62) Ardenkjaer-Larsen, J. H.; MacHoll, S.; Jóhannesson, H. Dynamic Nuclear Polarization with Trityls at 1.2 K. *Appl. Magn. Reson.* **2008**, *34*, 509–522.
(63) Jóhannesson, H.; Macholl, S.; Ardenkjaer-Larsen, J. H. Dynamic Nuclear Polarization of [1-13C]pyruvic Acid at 4.6 Tesla. *J. Magn. Reson.* **2009**, *197*, 167–175.
(64) WHO. Pneumococcal Vaccines WHO Position Paper – 2012. *Wkly. Epidemiol. Rec.* **2012**, *87*, 129–144.
(65) O'Brien, K. L.; Wolfson, L. J.; Watt, J. P.; Henkle, E.; Deloria-Knoll, M.; McCall, N.; Lee, E.; Mulholland, K.; Levine, O. S.; Cherian, T. Burden of Disease Caused by Streptococcus Pneumoniae in Children Younger than 5 Years: Global Estimates. *Lancet* **2009**, *374*, 893–902.

(66) WHO. Pneumococcal Conjugate Vaccine for Childhood Immunization – WHO Position Paper. *Wkly. Epidemiol. Rec.* **2007**, *82*, 93–104.
(67) Morens, D. M.; Taubenberger, J. K.; Fauci, A. S. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. *J. Infect. Dis.* **2008**, *198*, 962–970.
(68) AlonsoDeVelasco, E.; Verheul, A. F.; Verhoef, J.; Snippe, H. Streptococcus Pneumoniae: Virulence Factors, Pathogenesis, and Vaccines. *Microbiol. Rev.* **1995**, *59*, 591–603.
(69) Henrichsen, J. Six Newly Recognized Types of Streptoccus Penumoniae . *J. Clin. Microbiol* **1995**, *33*, 2759–2762.
(70) van Tonder, A. L.; Bray, J. E.; Quirk, S. J.; Haraldsson, G.; Jolley, K. A.; Maiden, M. C.; Hoffmann, S.; Bentley, S. D.; Ásgeir, H.; Erlendsdóttir, H.; et al. Putatively Novel Serotypes and the Potential for Reduced Vaccine Effectiveness: Capsular Locus Diversity Revealed among 5,405 Pneumococcal Genomes. *Microb. Genomics* **2016**, *2*, 90.
(71) Geno, K. A.; Saad, J. S.; Nahm, H. Discovery of Novel Pneumococcal Serotype 35D, a Natural WciG-Deficient Variant of Serotype 35B. *J. Clin. Microbiol.* **2017**, *55*, 1416–1425.
(72) Ko, K. S.; Baek, J. Y.; Song, J. H. Capsular Gene Sequences and Genotypes Of “serotype 6E” streptococcus Pneumoniae Isolates. *J. Clin. Microbiol.* **2013**, *51*, 3395–3399.
(73) Burton, R. L.; Geno, K. A.; Saad, J. S.; Nahm, M. H. Pneumococcus with the “6E” Cps Locus Produces Serotype 6B Capsular Polysaccharide. *J. Clin. Microbiol.* **2016**, *54*, 967–971.
(74) Kapatai, G.; Sheppard, C. L.; Troxler, L. J.; Litt, D. J.; Furrer, J.; Hilty, M.; Fry, N. K. Pneumococcal 23B Molecular Subtype Identified Using Whole Genome Sequencing. *Genome Biol. Evol.* **2017**, *9*, 2145–2158.
(75) Paradiso, P. R. Advances in Pneumococcal Disease Prevention: 13-Valent Pneumococcal Conjugate Vaccine for Infants and Children. *Clin. Infect. Dis.* **2011**, *52*, 1241–1247.
(76) WHO. Pneumococcal Vaccines WHO Position Paper - 2012 - Recommendations. *Vaccine* **2012**, *30*, 4717–4718.
(77) Steenhoff, A. P.; Shah, S. S.; Ratner, A. J.; Patil, S. M.; McGowan, K. L. Emergence of Vaccine-Related Pneumococcal Serotypes as a Cause of Bacteremia. *Clin. Infect. Dis.* **2006**, *42*, 907–914.
(78) Singleton, R. J.; Hennessy, T. W.; Bulkow, L. R.; Hammitt, L. L.; Zulz, T.; Hurlburt, D. a; Butler, J. C.; Rudolph, K.; Parkinson, A. Invasive Pneumococcal Disease Caused by Nonvaccine Serotypes Among Alaska Native Pneumococcal Conjugate Vaccine Coverage. *J. Am. Med. Assoc.* **2007**, *297*, 1784–1792.
(79) Millard, C. M.; Baiano, J. C. F.; Chan, C.; Yuen, B.; Aviles, F.; Landos, M.; Chong, R. S. M.; Benedict, S.; Barnes, A. C. Evolution of the Capsular Operon of Streptococcus Iniae in Response to Vaccination. *Appl. Environ. Microbiol.* **2012**, *78*, 8219–8226.
(80) Waight, P. A.; Andrews, N. J.; Ladhani, S. N.; Sheppard, C. L.; Slack, M. P. E.; Miller, E. Effect of the 13-Valent Pneumococcal Conjugate Vaccine on Invasive Pneumococcal Disease in England and Wales 4 Years after Its Introduction: An Observational Cohort Study. *Lancet Infect. Dis.* **2015**, *15*, 535–543.
(81) Kapatai, G.; Sheppard, C. L.; Al-Shahib, A.; Litt, D. J.; Underwood, A. P.; Harrison, T. G.; Fry, N. K. Whole Genome Sequencing of Streptococcus Pneumoniae: Development, Evaluation and Verification of Targets for Serogroup and Serotype Prediction Using an Automated Pipeline. *PeerJ* **2016**, *4*, e2477.
(82) Moreau, M.; Richards, J. C.; Perry, M. B. Application of High-resolution NMR Spectroscopy to the Elucidation of the Structure of the Specific Capsular Polysaccharide of Streptococcus Pneumoniae Type 7F. *Carbohydr. Res.* **1988**, *182*, 79–99.
(83) Backman-Marklund, I.; Jansson, P.-E.; Lindberg, B.; Henrichsen, J. Structural Studies of the Capsular Polysaccharide from Streptococcus Pneumoniae Type 7A. *Carbohydr. Res.* **1990**, *198*, 67–77.
(84) Jansson, P. E.; Lindberg, J.; Swarna Wimalasiri, K. M.; Henrichsen, J. The Structure of the Capsular Polysaccharide from Streptococcus Pneumoniae Type 7B. *Carbohydr. Res.* **1991**,

217, 171–180.
(85) Sørensen, U. B. S.; Henrichsen, J.; Chen, H. C.; Szu, S. C. Covalent Linkage between the Capsular Polysaccharide and the Cell Wall Peptidoglycan of Streptococcus Pneumoniae Revealed by Immunochemical Methods. *Microb. Pathog.* **1990**, *8*, 325–334.
(86) Skovsted, I. C.; Kerrn, M. B.; Sonne-Hansen, J.; Sauer, L. E.; Nielsen, A. K.; Konradsen, H. B.; Petersen, B. O.; Nyberg, N. T.; Duus, J. Purification and Structure Characterization of the Active Component in the Pneumococcal 22F Polysaccharide Capsule Used for Adsorption in Pneumococcal Enzyme-Linked Immunosorbent Assays. *Vaccine* **2007**, *25*, 6490–6500.
(87) Vialle, S.; Sepulcri, P.; Dubayle, J.; Talaga, P. The Teichoic Acid (C-Polysaccharide) Synthesized by Streptococcus Pneumoniae Serotype 5 Has a Specific Structure. *Carbohydr. Res.* **2005**, *340*, 91–96.
(88) Mavroidi, A.; Aanensen, D. M.; Godoy, D.; Skovsted, I. C.; Kaltoft, M. S.; Reeves, P. R.; Bentley, S. D.; Spratt, B. G. Genetic Relatedness of the Streptococcus Pneumoniae Capsular Biosynthetic Loci. *J. Bacteriol.* **2007**, *189*, 7841–7855.
(89) Aanensen, D. M.; Mavroidi, A.; Bentley, S. D.; Reeves, P. R.; Spratt, B. G. Predicted Functions and Linkage Specificities of the Products of the Streptococcus Pneumoniae Capsular Biosynthetic Loci. *J. Bacteriol.* **2007**, *189*, 7856–7876.
(90) Sheppard, C. L.; Pichon, B.; George, R. C.; Hall, L. M. C. Streptococcus Pneumoniae Isolates Expressing a Capsule with Epitopes of Both Serotypes 6A and 6B. *Clin. Vaccine Immunol.* **2010**, *17*, 1820–1822.
(91) Oliver, M. B.; Jones, C.; Larson, T. R.; Calix, J. J.; Zartler, E. R.; Yother, J.; Nahm, M. H. Streptococcus Pneumoniae Serotype 11D Has a Bispecific Glycosyltransferase and Expresses Two Different Capsular Polysaccharide Repeating Units. *J. Biol. Chem.* **2013**, *288*, 21945–21954.
(92) Oliver, M. B.; Van Der Linden, M. P. G.; Küntzel, S. A.; Saad, J. S.; Nahm, M. H. Discovery of Streptococcus Pneumoniae Serotype 6 Variants with Glycosyltransferases Synthesizing Two Differing Repeating Units. *J. Biol. Chem.* **2013**, *288*, 25976–25985.
(93) Park, I. H.; Geno, K. A.; Yu, J.; Oliver, M. B.; Kim, K. H.; Nahm, M. H. Genetic, Biochemical, and Serological Characterization of a New Pneumococcal Serotype, 6H, and Generation of a Pneumococcal Strain Producing Three Different Capsular Repeat Units. *Clin. Vaccine Immunol.* **2015**, *22*, 313–318.
(94) Derome, A. E.; Williamson, M. P. Rapid-Pulsing Artifacts in Double-Quantum-Filtered COSY. *J. Magn. Reson.* **1990**, *88*, 177–185.
(95) Bax, A.; Davis, D. G. MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.* **1985**, *65*, 355–360.
(96) Hwang, T.; Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Waveforms and Pulsed Field Gradients. *J.Magn. Reson. Ser. A* **1995**, *112*, 275.
(97) Lee, M.-W.; Hur, H.; Chang, K.-C.; Lee, T.-S.; Ka, K.-H.; Jankovsky, L. Introduction to Distribution and Ecology of Sterile Conks of Inonotus Obliquus. *Mycobiology* **2008**, *36*, 199–202.
(98) Cha, J. Y.; Lee, S. Y.; Lee, S. Y.; Chun, K. W. Basidiocarp Formation by Inonotus Obliquus on a Living Paper Birch Tree. *For. Pathol.* **2011**, *41*, 163–164.
(99) Shashkina, M. Y.; Shashkin, P. N.; Sergeev, A. V. Chemical and Medicobiological Properties of Chaga (Review). *Pharm. Chem. J.* **2006**, *40*, 560–568.
(100) Shikov, A. N.; Pozharitskaya, O. N.; Makarov, V. G.; Wagner, H.; Verpoorte, R.; Heinrich, M. Medicinal Plants of the Russian Pharmacopoeia; Their History and Applications. *J. Ethnopharmacol.* **2014**, *154*, 481–536.
(101) Zheng, W.; Miao, K.; Liu, Y.; Zhao, Y.; Zhang, M.; Pan, S.; Dai, Y. Chemical Diversity of Biologically Active Metabolites in the Sclerotia of Inonotus Obliquus and Submerged Culture Strategies for up-Regulating Their Production. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 1237–1254.
(102) Kim, Y. O.; Han, S. B.; Lee, H. W.; Ahnb, H. J.; Yoonb, Y. D.; Jungb, J. K.; Kimb, H. M.;

Shin, C. S. Immuno-Stimulating Effect of the Endo-Polysaccharide Produced by Submerged Culture of Inonotus Obliquus. *Life Sci.* **2005**, *77*, 2438–2456.
(103) Kim, Y. O.; Park, H. W.; Kim, J. H.; Lee, J. Y.; Moon, S. H.; Shin, C. S. Anti-Cancer Effect and Structural Characterization of Endo-Polysaccharide from Cultivated Mycelia of Inonotus Obliquus. *Life Sci.* **2006**, *79*, 72–80.
(104) Rhee, S. J.; Cho, S. Y.; Kim, K. M.; Cha, D. S.; Park, H. J. A Comparative Study of Analytical Methods for Alkali-Soluble β-Glucan in Medicinal Mushroom, Chaga (Inonotus Obliquus). *LWT - Food Sci. Technol.* **2008**, *41*, 545–549.
(105) Fan, L.; Ding, S.; Ai, L.; Deng, K. Antitumor and Immunomodulatory Activity of Water-Soluble Polysaccharide from Inonotus Obliquus. *Carbohydrate Polymers*, 2012, *90*, 870–874.
(106) Du, X.; Mu, H.; Zhou, S.; Zhang, Y.; Zhu, X. Chemical Analysis and Antioxidant Activity of Polysaccharides Extracted from Inonotus Obliquus Sclerotia. *Int. J. Biol. Macromol.* **2013**, *62*, 691–696.
(107) Chen, Y.; Huang, Y.; Cui, Z.; Liu, J. Purification, Characterization and Biological Activity of a Novel Polysaccharide from Inonotus Obliquus. *Int. J. Biol. Macromol.* **2015**, *79*, 587–594.
(108) Wasser, S. Medicinal Mushrooms as a Source of Antitumor and Immunomodulating Polysaccharides. *Appl. Microbiol. Biotechnol.* **2002**, *60*, 258–274.
(109) Ina, K.; Kataoka, T.; Ando, T. The Use of Lentinan for Treating Gastric Cancer. *Anticancer. Agents Med. Chem.* **2013**, *13*, 681–688.
(110) Erwig, L. P.; Gow, N. A. R. Interactions of Fungal Pathogens with Phagocytes. *Nat. Rev. Microbiol.* **2016**, *14*, 163–176.
(111) Fraser, I. P.; Stuart, L.; Ezekowitz, R. A. B. TLR-Independent Pattern Recognition Receptors and Anti-Inflammatory Mechanisms. *J. Endotoxin Res.* **2004**, *10*, 120–124.
(112) Bock, K.; Pedersen, C.; Pedersen, H. Carbon-13 Nuclear Magnetic Resonance Data for Oligosaccharides. *Advances in carbohydrate chemistry and biochemistry*, 1984, *42*, 193–225.
(113) Shashkov, A. S.; Nifantev, N. E.; Amochaeva, V. E.; Kochetkov, N. K. 1H and 13C NMR Data for 2-O-, 3-O-, and 2,3-Di-O-Glycosylated Methyl α- and β-D-Glucopyranosides and β-D-Galactopyranosides. *Magn. Reson. Chem.* **1993**, *31*, 599–613.
(114) Flower, D. R. The Lipocalin Protein Family: Structure and Function. *Biochem. J.* **1996**, *318*, 1–14.
(115) Bordin, G.; Raposo, F. C.; de la Calle, B.; Rodriguez, A. R. Identification and Quantification of Major Bovine Milk Proteins by Liquid Chromatography. *J. Chromatogr. A* **2001**, *928*, 63–76.
(116) Girard, M.; Turgeon, S. L.; Gauthier, S. F. Quantification of the Interactions between β-Lactoglobulin and Pectin through Capillary Electrophoresis Analysis. *J. Agric. Food Chem.* **2003**, *51*, 6043–6049.
(117) Weinbreck, F.; de Vries, R.; Schrooyen, P.; de Kruif, C. G. Complex Coacervation of Whey Proteins and Gum Arabic. *Biomacromolecules* **2003**, *4*, 293–303.
(118) Hosseini, S. M. H.; Emam-Djomeh, Z.; Razavi, S. H.; Moosavi-Movahedi, A. A.; Saboury, A. A.; Atri, M. S.; Van der Meeren, P. β-Lactoglobulin-Sodium Alginate Interaction as Affected by Polysaccharide Depolymerization Using High Intensity Ultrasound. *Food Hydrocoll.* **2013**, *32*, 235–244.
(119) Stender, E. G. P.; Khan, S.; Ipsen, R.; Madsen, F.; Hägglund, P.; Hachem, M. A.; Almdal, K.; Westh, P.; Svensson, B. Effect of Alginate Size, Mannuronic/guluronic Acid Content and pH on Particle Size, Thermodynamics and Composition of Complexes with β-Lactoglobulin. *Food Hydrocoll.* **2018**, *75*, 157–163.
(120) Degeest, B.; Vaningelgem, F.; Laws, A. P.; De Vuyst, L. UDP-N-Acetylglucosamine 4-Epimerase Activity Indicates the Presence of N-Acetylgalactosamine in Exopolysaccharides of Streptococcus Thermophilus Strains. *Appl. Environ. Microbiol.* **2001**, *67*, 3976–3984.
(121) Faber, E. J.; Kamerling, J. P.; Vliegenthart, J. F. Structure of the Extracellular Polysaccharide Produced by *Lactobacillus Delbrueckii* Subsp. *Bulgaricus* 291. *Carbohydr.*

Res. **2001**, *331*, 183–194.
(122) Lee, K. Y.; Mooney, D. J. Alginate: Properties and Biomedical Applications. *Prog. Polym. Sci.* **2012**, *37*, 106–126.
(123) Wong, T. Y.; Preston, L. A.; Schiller, N. L. ALGINATE LYASE: Review of Major Sources and Enzyme Characteristics, Structure-Function Analysis, Biological Roles, and Applications. *Annu. Rev. Microbiol.* **2000**, *54*, 289–340.
(124) Campa, C.; Holtan, S.; Nilsen, N.; Bjerkan, T. M.; Stokke, B. T.; Skjåk-Braek, G. Biochemical Analysis of the Processive Mechanism for Epimerization of Alginate by Mannuronan C-5 Epimerase AlgE4. *Biochem. J.* **2004**, *381*, 155–164.
(125) Li, L.; Jiang, X.; Guan, H.; Wang, P. Preparation, Purification and Characterization of Alginate Oligosaccharides Degraded by Alginate Lyase from Pseudomonas Sp. HZJ 216. *Carbohydr. Res.* **2011**, *346*, 794–800.
(126) Jacobson, R. H.; Zhang, X. J.; DuBose, R. F.; Matthews, B. W. Three-Dimensional Structure of Beta-Galactosidase from E. Coli. *Nature* **1994**, *369*, 761–766.
(127) Juers, D. H.; Heightman, T. D.; Vasella, A.; McCarter, J. D.; Mackenzie, L.; Withers, S. G.; Matthews, B. W. A Structural View of the Action of Escherichia Coli (lacZ) β-Galactosidase. *Biochemistry* **2001**, *40*, 14781–14794.
(128) Ardèvol, A.; Rovira, C. The Molecular Mechanism of Enzymatic Glycosyl Transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-like Species. *Angew. Chemie Int. Ed.* **2011**, *50*, 10897–10901.
(129) Wheatley, R. W.; Huber, R. E. An Allolactose Trapped at the lacZ β-Galactosidase Active Site with Its Galactosyl Moiety in a 4H3 Conformation Provides Insights into the Formation, Conformation, and Stabilization of the Transition State. *Biochem. Cell Biol.* **2015**, *93*, 531–540.
(130) Egel, R. The Lac-Operon for Lactose Degradation, or Rather for the Utilization of Galactosylglycerols from Galactolipids? *J. Theor. Biol.* **1979**, *79*, 117–119.
(131) Huber, R. E.; Kurz, G.; Wallenfels, K. A Quantitation of the Factors Which Affect the Hydrolase and Transgalactosylase Activities of β-Galactosidase (E. Coli) on Lactose. *Biochemistry* **1976**, *15*, 1994–2001.
(132) Jacob, F.; Monod, J. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J. Mol. Biol.* **1961**, *3*, 318–356.
(133) Viratelle, O. M.; Yon, J. M. Nucleophilic Competition in Some β-Galactosidase-Catalyzed Reactions. *Eur. J. Biochem.* **1973**, *116*, 110–116.
(134) Deschavanne, P. J.; Viratelle, O. M.; Yon, J. M. Conformational of the Active Site of β-Galactosidase. *J. Biol. Chem.* **1979**, *253*, 833–838.
(135) Chen, C. W.; Ou-Yang, C. C.; Yeh, C. W. Synthesis of Galactooligosaccharides and Transgalactosylation Modeling in Reverse Micelles. *Enzyme Microb. Technol.* **2003**, *33*, 497–507.
(136) Brás, N. F.; Fernandes, P. A.; Ramos, M. J. QM/MM Studies on the β-Galactosidase Catalytic Mechanism: Hydrolysis and Transglycosylation Reactions. *J. Chem. Theory Comput.* **2010**, *6*, 421–433.
(137) Matella, N. J.; Dolan, K. D.; Lee, Y. S. Comparison of Galactooligosaccharide Production in Free-Enzyme Ultrafiltration and in Immobilized-Enzyme Systems. *J. Food Sci.* **2006**, *71*, 363–368.
(138) Hsu, C. a; Lee, S. L.; Chou, C. C. Enzymatic Production of Galactooligosaccharides by Beta-Galactosidase from Bifidobacterium Longum BCRC 15708. *J. Agric. Food Chem.* **2007**, *55*, 2225–2230.
(139) Rodriguez-Colinas, B.; De Abreu, M. A.; Fernandez-Arrojo, L.; De Beer, R.; Poveda, A.; Jimenez-Barbero, J.; Haltrich, D.; Ballesteros Olmo, A. O.; Fernandez-Lobato, M.; Plou, F. J. Production of Galacto-Oligosaccharides by the β-Galactosidase from Kluyveromyces Lactis: Comparative Analysis of Permeabilized Cells versus Soluble Enzyme. *J. Agric. Food Chem.* **2011**, *59*, 10477–10484.
(140) Adam, A. C.; Rubio-Texeira, M.; Polaina, J. Lactose: The Milk Sugar from a Biotechnological

Perspective. *Crit. Rev. Food Sci. Nutr.* **2004**, *44*, 553–557.
(141) Husain, Q. β Galactosidases and Their Potential Applications: A Review. *Crit. Rev. Biotechnol.* **2010**, *30*, 41–62.
(142) Robillard, G.; Shaw, E.; Shulman, R. G. 13C High-Resolution Nuclear Magnetic Resonance Studies of Enzyme-Substrate Reactions at Equilibrium. Substrate Strain Studies of Chymotrypsin-N-Acetyltyrosine Semicarbazide Complexes. *Proc Natl Acad Sci USA* **1974**, *71*, 2623–2626.
(143) Withers, S. G.; Dombroski, D.; Berven, L. A.; Kilburn, D. G.; Miller, R. C.; J. Warren, R. A.; Gilkes, N. R. Direct 1H N.M.R. Determination of the Stereochemical Course of Hydrolyses Catalysed by Glucanase Components of the Cellulase Complex. *Biochem. Biophys. Res. Commun.* **1986**, *139*, 487–494.
(144) Beeren, S. R.; Petersen, B. O.; Bøjstrup, M.; Hindsgaul, O.; Meier, S. Time-Resolved in-Situ Observation of Starch Polysaccharide Degradation Pathways. *ChemBioChem* **2013**, *14*, 2506–2511.
(145) Meier, S.; Jensen, P. R.; Duus, J. Ø. Real-Time Detection of Central Carbon Metabolism in Living Escherichia Coli and Its Response to Perturbations. *FEBS Lett.* **2011**, *585*, 3133–3138.
(146) Meier, S.; Karlsson, M.; Jensen, P. R.; Lerche, M. H.; Duus, J. Ø. Metabolic Pathway Visualization in Living Yeast by DNP-NMR. *Mol. Biosyst.* **2011**, *7*, 2834–2836.
(147) Meier, S.; Jensen, P. R.; Duus, J. Ø. Direct Observation of Metabolic Differences in Living Escherichia Coli Strains K-12 and BL21. *ChemBioChem* **2012**, *13*, 308–310.
(148) Rodrigues, T. B.; Serrao, E. M.; Kennedy, B. W. C.; Hu, D.-E.; Kettunen, M. I.; Brindle, K. M. Magnetic Resonance Imaging of Tumor Glycolysis Using Hyperpolarized 13C-Labeled Glucose. *Nat. Med.* **2013**, *20*, 93–97.
(149) Miclet, E.; Abergel, D.; Bornet, A.; Milani, J.; Jannin, S.; Bodenhausen, G. Toward Quantitative Measurements of Enzyme Kinetics by Dissolution Dynamic Nuclear Polarization. *J. Phys. Chem. Lett.* **2014**, *5*, 3290–3295.
(150) Meier, S.; Karlsson, M.; Jensen, P. R. Detecting Elusive Intermediates in Carbohydrate Conversion: A Dynamic Ensemble of Acyclic Glucose–Catalyst Complexes. *ACS Sustain. Chem. Eng.* **2017**, *5*, 5571–5577.
(151) Defaye, J.; Gadelle, A.; Angyal, S. J. An Efficient Synthesis of L-Fucose and L-(4-2H)fucose. *Carbohydr. Res.* **1984**, *126*, 165–169.
(152) Lumata, L.; Merritt, M. E.; Malloy, C. R.; Sherry, A. D.; Kovacs, Z. Impact of Gd 3+ on DNP of [1- 13C]pyruvate Doped with Trityl OX063, BDPA, or 4-Oxo-TEMPO. *J. Phys. Chem. A* **2012**, *116*, 5129–5138.
(153) Koto, S.; Inada, S.; Zen, S. Upfield Shift of CMR of Anomeric Carbons of 1,1'-linked Glycopyranosyl Glycopyranosides. *Chem. Lett.* **1980**, 403–406.
(154) Hinz, S. W. A.; Verhoef, R.; Schols, H. A.; Vincken, J. P.; Voragen, A. G. J. Type I Arabinogalactan Contains β-D-Galp-(1→3)-β-D-Galp Structural Elements. *Carbohydr. Res.* **2005**, *340*, 2135–2143.
(155) Huber, R. E.; Wallenfels, K.; Kurz, G. The Action of β-Galactosidase (Escherichia Coli) on Allolactose. *Can. J. Biochem.* **1975**, *53*, 1035–1038.
(156) Thomas, A.; Shukla, A.; Sivakumar, S.; Verma, S. Assembly, Postsynthetic Modification and Hepatocyte Targeting by Multiantennary, Galactosylated Soft Structures. *Chem. Commun.* **2014**, *50*, 15752–15755.
(157) Percec, V.; Leowanawat, P.; Sun, H. J.; Kulikov, O.; Nusbaum, C. D.; Tran, T. M.; Bertin, A.; Wilson, D. A.; Peterca, M.; Zhang, S.; et al. Modular Synthesis of Amphiphilic Janus Glycodendrimers and Their Self-Assembly into Glycodendrimersomes and Other Complex Architectures with Bioactivity to Biomedically Relevant Lectins. *J. Am. Chem. Soc.* **2013**, *135*, 9055–9077.
(158) Green, D. E.; Ferreira, C. L.; Stick, R. V.; Patrick, B. O.; Adam, M. J.; Orvig, C. Carbohydrate-Bearing 3-Hydroxy-4-Pyridinonato Complexes of Gallium (III) and Indium (III). *Bionconjugate Chem.* **2005**, *16*, 1597–1609.

(159) Kobata, A. Structures and Application of Oligosaccharides in Human Milk. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **2010**, *86*, 731–747.
(160) Wu, S.; Tao, N.; German, J. B.; Grimm, R.; Lebrilla, C. B. Development of an Annotated Library of Neutral Human Milk Oligosaccharides. *J. Proteome Res.* **2010**, *9*, 4138–4151.
(161) Zhao, C.; Wu, Y.; Liu, X.; Liu, B.; Cao, H.; Yu, H.; Sarker, S. D.; Nahar, L.; Xiao, J. Functional Properties, Structural Studies and Chemo-Enzymatic Synthesis of Oligosaccharides. *Trends Food Sci. Technol.* **2017**, *66*, 135–145.
(162) Zeuner, B.; Jers, C.; Mikkelsen, J. D.; Meyer, A. S. Methods for Improving Enzymatic Trans-Glycosylation for Synthesis of Human Milk Oligosaccharide Biomimetics. *J. Agric. Food Chem.* **2014**, *62*, 9615–9631.
(163) Schultz-johansen, M.; Glaring, M. A.; Bech, P. K.; Stougaard, P. Draft Genome Sequence of a Novel Marine Bacterium , Paraglaciecola Sp . Strain S66 , with Hydrolytic Activity against Seaweed. *Genome Announc.* **2016**, *4*, 4–5.
(164) Bech, P. K.; Schultz-Johansen, M.; Glaring, M. A.; Barbeyron, T.; Czjzek, M.; Stougaard, P. Paraglaciecola Hydrolytica Sp. Nov., a Bacterium with Hydrolytic Activity against Multiple Seaweed-Derived Polysaccharides. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 2242–2247.
(165) Honda, Y.; Nishimoto, M.; Katayama, T.; Kitaoka, M. Characterization of the Cytosolic β-N-Acetylglucosaminidase from Bifidobacterium Longum Subsp. Longum. *J. Appl. Glycosci.* **2013**, *60*, 141–146.
(166) Yuan, J.; Zhu, L.; Liu, X.; Li, T.; Zhang, Y.; Ying, T.; Wang, B.; Wang, J.; Dong, H.; Feng, E.; et al. A Proteome Reference Map and Proteomic Analysis of Bifidobacterium Longum NCC2705. *Mol. Cell. Proteomics* **2006**, *5*, 1105–1118.
(167) Boesten, R.; Schuren, F.; Ben Amor, K.; Haarman, M.; Knol, J.; de Vos, W. M. Bifidobacterium Population Analysis in the Infant Gut by Direct Mapping of Genomic Hybridization Patterns: Potential for Monitoring Temporal Development and Effects of Dietary Regimens. *Microb. Biotechnol.* **2011**, *4*, 417–427.
(168) Garrido, D.; Ruiz-Moyano, S.; Jimenez-Espinoza, R.; Eom, H. J.; Block, D. E.; Mills, D. A. Utilization of Galactooligosaccharides by Bifidobacterium Longum Subsp. Infantis Isolates. *Food Microbiol.* **2013**, *33*, 262–270.
(169) Miwa, M.; Horimoto, T.; Kiyohara, M.; Katayama, T.; Kitaoka, M.; Ashida, H.; Yamamoto, K. Cooperation of β-Galactosidase and β-N-Acetylhexosaminidase from Bifidobacteria in Assimilation of Human Milk Oligosaccharides with Type 2 Structure. *Glycobiology* **2010**, *20*, 1402–1409.
(170) André-Miral, C.; Koné, F. M.; Solleux, C.; Grandjean, C.; Dion, M.; Tran, V.; Tellier, C. De Novo Design of a Trans-β-N-Acetylglucosaminidase Activity from a GH1 β-Glycosidase by Mechanism Engineering. *Glycobiology* **2015**, *25*, 394–402.
(171) Rauvolfová, J.; Kuzma, M.; Weignerová, L.; Fialová, P.; Přikrylová, V.; Pišvejcová, A.; Macková, M.; Křen, V. β-N-Acetylhexosaminidase-Catalysed Synthesis of Non-Reducing Oligosaccharides. *J. Mol. Catal. B Enzym.* **2004**, *29*, 233–239.
(172) Drzewiecka, D.; Arbatsky, N. P.; Staczek, P.; Shashkov, A. S.; Knirel, Y. A.; Sidorczyk, Z. Structural and Serological Studies of the O-Polysaccharide of Strains from a Newly Created Proteus O78 Serogroup Prevalent in Polish Patients. *FEMS Immunol. Med. Microbiol.* **2010**, *58*, 269–276.
(173) Chen, X.; Xu, L.; Jin, L.; Sun, B.; Gu, G.; Lu, L.; Xiao, M. Efficient and Regioselective Synthesis of β-GalNAc/GlcNAc-Lactose by a Bifunctional Transglycosylating β-N-Acetylhexosaminidase from Bifidobacterium Bifidum. *Appl. Environ. Microbiol.* **2016**, *82*, 5642–5652.
(174) Yamamoto, F.; Clausen, H.; White, T.; Marken, J.; Hakomori, S. Molecular Genetic Basis of the Histo-Blood Group ABO System. *Nature* **1990**, *345*, 229–233.
(175) Patenaude, S. I.; Seto, N. O. L.; Borisova, S. N.; Szpacenko, A.; Marcus, S. L.; Palcic, M. M.; Evans, S. V. The Structural Basis for Specificity in Human ABO(H) Blood Group Biosynthesis. *Nat. Struct. Biol.* **2002**, *9*, 685–690.
(176) Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. Structures and Mechanisms of

Glycosyltransferases. *Glycobiology* **2006**, *16*, 29R–37R.
(177) Alfaro, J. A.; Zheng, R. B.; Persson, M.; Letts, J. A.; Polakowski, R.; Bai, Y.; Borisova, S. N.; Seto, N. O. L.; Lowary, T. L.; Palcic, M. M.; et al. ABO(H) Blood Group A and B Glycosyltransferases Recognize Substrate via Specific Conformational Changes. *J. Biol. Chem.* **2008**, *283*, 10097–10108.
(178) Breton, C.; Fournel-Gigleux, S.; Palcic, M. M. Recent Structures, Evolution and Mechanisms of Glycosyltransferases. *Curr. Opin. Struct. Biol.* **2012**, *22*, 540–549.
(179) Angulo, J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Rama Krishna, N.; Peters, H.; Palcic, M. M.; Peters, T. Blood Group B Galactosyltransferase: Insights into Substrate Binding from NMR Experiments. *J. Am. Chem. Soc.* **2006**, *128*, 13529–13538.
(180) Blume, A.; Angulo, J.; Biet, T.; Peters, H.; Benie, A. J.; Palcic, M. m.; Peters, T. Fragment-Based Screening of the Donor Substrate Specificity of Human Blood Group B Galactosyltransferase Using Saturation Transfer Difference NMR. *J. Biol. Chem.* **2006**, *281*, 32728–32740.
(181) Rademacher, C.; Landström, J.; Sindhuwinata, N.; Palcic, M. M.; Widmalm, G.; Peters, T. NMR-Based Exploration of the Acceptor Binding Site of Human Blood Group B Galactosyltransferase with Molecular Fragments. *Glycoconj. J.* **2010**, *27*, 349–358.
(182) Soya, N.; Fang, Y.; Palcic, M. M.; Klassen, J. S. Trapping and Characterization of Covalent Intermediates of Mutant Retaining Glycosyltransferases. *Glycobiology* **2011**, *21*, 547–552.
(183) Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *Ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 7528–7547.
(184) Hurtado-Guerrero, R.; Davies, G. J. Recent Structural and Mechanistic Insights into Post-Translational Enzymatic Glycosylation. *Curr. Opin. Chem. Biol.* **2012**, *16*, 479–487.
(185) Gardner, K. H.; Kay, L. E. The Use of 2 H, 13 C, 15 N Multidimensional Nmr Gto Study the Structure and Dynamics of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357–406.
(186) Dyson, H. J.; Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **2004**, *104*, 3607–3622.
(187) Goto, N.; Gardner, K.; Mueller, G.; Willis, R.; Kay, L. A Robust and Cost-Effective Method for the Production of Val, Leu, Ile (δ1) Methyl-Protonated 15N-, 13C-, 2H-Labeled Proteins. *J. Biomol. NMR* **1999**, *13*, 369–374.
(188) Palcic, M. M.; Heerze, L. D.; Pierce, M.; Hindsgaul, O. The Use of Hydrophobic Synthetic Glycosides as Acceptors in Glycosyltransferase Assays. *Glycoconj. J.* **1988**, *5*, 49–63.
(189) Sindhuwinata, N.; Munoz, E.; Munoz, F. J.; Palcic, M. M.; Peters, H.; Peters, T. Binding of an Acceptor Substrate Analog Enhances the Enzymatic Activity of Human Blood Group B Galactosyltransferase. *Glycobiology* **2010**, *20*, 718–723.
(190) Pervushin, K. V; Wider, G.; Wüthrich, K. Single Transition-to-Single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR* **1998**, *12*, 345–348.
(191) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **1999**, *38*, 1784–1788.
(192) Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chemie Int. Ed.* **2003**, *42*, 864–890.

Paper 1

Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D

Kjeldsen, C.; Slott, S; Elverdal, P. E.; Sheppard, C. L.; Kapatai, G.; Fry, N. K.; Skovsted, I. C.; Duus, J. Ø.

Manuscript in preparation

Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D

Christian Kjeldsen[a], Sofie Slott[a], Pernille Landsbo Elverdal[b], Carmen L. Sheppard[c], Georgia Kapatai[c], Norman K. Fry[c], Ian C. Skovsted[b], Jens Ø. Duus[a,*]

[a] Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark

[b] SSI Diagnostica A/S, Hilleroed, Denmark

[c] Respiratory and Vaccine Preventable Bacterial Reference Unit, Public Health England – National Infection Service, Colindale, London, UK

*Corresponding author. jduus@kemi.dtu.dk, +45 45252451

Abstract

Streptococcus pneumoniae is characterised into 92 serotypes based on antigenic reactions of commercial rabbit sera to the capsular polysaccharides. During development of a bioifromatical serotyping tool (PneumoCaT), an isolate exhibited a novel codon at residue 385 of the glycosyltransferase gene *wcwK* encoding a distinct amino acid, which differentiates genogroup 7. Investigation by repeat serotyping and Quellung reaction revealed a novel pattern of factor sera with the isolate reacting very strongly with 7f, but also with 7e factor sera. The structure of the capsular polysaccharide was determined by NMR spectroscopy to be an approximately 5:1 mixture between the structures of 7C and 7B, respectively, and the structure of 7C was also elucidated. All data from whole genome sequencing, NMR spectroscopy, production of antisera and serotyping of the novel 7 strain shows that it is a new serotype, which will be named in the Danish nomenclature as 7D.

Keywords: NMR spectroscopy, structure elucidation, *Streptococcus pneumoniae*, whole genome sequence, hybrid serotype

1 Introduction

Streptococcus pneumoniae (pneumococcus) is a major cause of morbidity and mortality worldwide. The pneumococcal capsular polysaccharide (cps) which surrounds the organism is a major virulence factor, enabling the cell to evade phagocytosis (AlonsoDeVelasco, Verheul, Verhoef, & Snippe, 1995). The currently licensed 23–serotype valent plain polysaccharide (PPV) and 7, 10 and 13-valent protein-conjugate vaccines (PCV) cover the most important circulating serotypes.

To date, 92 serotypes of pneumococcus have been described based on reaction of their capsule with the commercial rabbit antisera (Henrichsen, 1995; McEllistrem & Nahm, 2012) available from SSI Diagnostica A/S, in this scheme immunogenically cross-reactive serotypes are assigned to serogroups in the Danish serotype classification system. As the capsule is the target of all currently licensed vaccines, accurate surveillance of pneumococcal serotypes is essential for the evaluation of the efficacy of the vaccines and to inform national vaccine policies. Vaccine pressure has resulted in an increase in the prevalence of non-vaccine serotypes and could also result in emergence of novel serotypes as the organism evolves to escape the vaccine.

However, many additional novel serotypes have been proposed based on genetic analysis (Geno, Saad, & Nahm, 2017; van Tonder et al., 2016) or reaction with monoclonal antibodies (Yu, Carvalho, Beall, & Nahm, 2008). Some of the described novel variants based on genetic analysis actually produce no difference in the structure of the polysaccharide, for example the putative serotype 6E which in fact produces a polysaccharide identical to serotype 6B (Burton, Geno, Saad, & Nahm, 2016), and the genetic variant of 23B which, although it has low sequence homology in the capsular operon and is from an entirely different genetic lineage to the originally described 23B, produces an identical polysaccharide and is therefore the same serotype (Kapatai et al., 2017). In other cases, the capsular polysaccharide produced appears to be either a hybrid or variable expression of different polysaccharide structures, this has been seen in the case of rare serogroup 6 variant isolates (Oliver, Van Der Linden, Küntzel, Saad, & Nahm, 2013; Sheppard, Pichon, George, & Hall, 2010) and 11A variants (Calix & Nahm, 2010), though in neither of these cases is it possible to determine a difference to standard 6A, 6B or 11A, using the widely available commercial typing sera.

During development and evaluation of a new bioinformatical serotyping tool (PneumoCaT, pneumococcal capsule typing) (Kapatai et al., 2016) to aid in surveillance of pneumococcal serotypes using whole genome data, a novel variant within serogroup 7 was discovered. The source of the pneumococcal isolate demonstrating a novel sequence was from an ear swab from an outpatient attending a London hospital in 2012, submitted for antibiotic susceptibility testing.

The WGS data from this strain have been submitted to the European National Archive (designated PHESPD0846 - https://www.ebi.ac.uk/ena/data/view/ERS1194144).

Serogroup 7 is an important serogroup as it contains the vaccine type 7F which is included in the currently licensed PPV23, PCV10 and PCV13 vaccines. Serogroup 7 contains four recognised serotypes, 7F, 7A, 7B and 7C, of which 7F is by far the most common worldwide, representing 98.8% of serogroup 7 isolates in UK serotype surveillance for epidemiological years 2006-2014 with serotypes 7A, 7B, and 7C, representing 0.2%, 0.2% 0.8% respectively (Public Health England surveillance data).

The polysaccharide repeat unit structures of serotypes 7F (Moreau, Richards, & Perry, 1988), 7A (Backman-Marklund, Jansson, Lindberg, & Henrichsen, 1990) and 7B (Jansson, Lindberg, Swarna Wimalasiri, & Henrichsen, 1991) are known, shown in Figure 1. All of the serotypes contain cell wall polysaccharide (CWPS) as well as their respective cps, and the structure of the CWPS has been elucidated (Skovsted et al., 2007). The antigenic formula of serotype 7F and 7A is similar and the only structural difference between them is the presence of a terminal side-chain β-D-galactose residue in 7F. Serotype 7B only has one antigenic component in common with serotype 7F and 7A and does not have O-acetyl groups (Kamerling, 2000). There was no published structure for serotype 7C. In the present study, it was realised that in order to describe structure of the newly isolated serotype, it was necessary to first make a full structure identification of 7C.

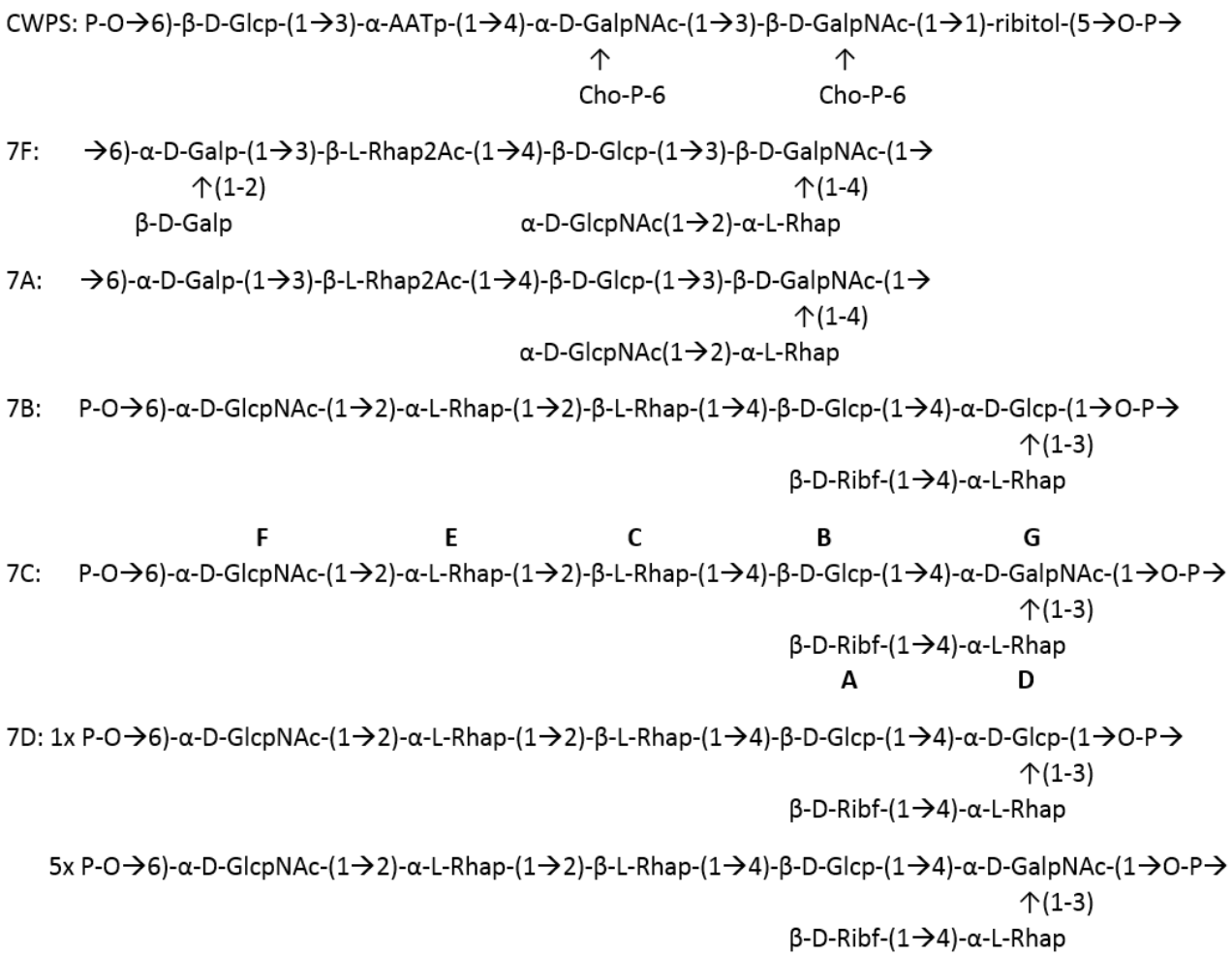


Figure 1: Structures of the serogroup 7 capsular polysaccharides as well as CWPS. The **A**-**G** labels for 7C have been highlighted, and are used for reference in the structure elucidation

The genetic arrangement of the cps operon in serogroup 7 was described by Mavroidi *et al.* in 2007. The genetic arrangement of the *cps* loci form two syntenic pairs, 7F/7A and 7B/7C and the two pairs cluster apart from each other on amino acid sequence comparisons (Mavroidi et al., 2007). The lack of the side chain in 7A compared to 7F can be explained by a frame shift mutation in *wcwD* which codes for a glycosyltransferase (Aanensen, Mavroidi, Bentley, Reeves, & Spratt, 2007).

During development of PneumoCaT, nine SNP positions in *wcwK*, a glycosyltransferase, which can be used to differentiate serotype 7B from serotype 7C and also serotype 40 which is closely genetically related to 7B and 7C (Kapatai et al., 2016), were described. Nucleotide differences at residues 46 and 385 of *wcwK* are considered as more significant, as both result in distinct amino acids in the three serotypes differentiated in this way.

When the PneumoCaT was evaluated on panels of previously serotyped isolates, one 7C isolate displayed an atypical SNP pattern which did not fit the expected patterns in the Capsular Type Variant database (CTV) used by the tool to define serotypes (Kapatai et al., 2016).

This paper describes the serological, genetic and structural properties of this novel variant, which is to be named in the Danish nomenclature as serotype 7D. We also describe the structure of the capsular polysaccharide of serotype 7C which has not been published previously. Furthermore, as the NMR assignment of the different capsular polysaccharides have been obtained at different conditions, and in some cases have incomplete assignments, the full assignment of 7F, 7A (supporting information) and 7B (table 2) can also be found in the supporting information.

2 Methods

2.1 DNA extraction and sequencing

Isolates were grown overnight on horse blood agar (PHE Media Services) with 5% CO_2. DNA was extracted from an entire agar plate of growth for each isolate using the QIAsymphony SP automated instrument (Qiagen) and QIAsymphony DSP DNA Mini Kit, using the manufacturer's recommended tissue extraction protocol for Gram negative bacteria (including a 1 hour pre-incubation with proteinase K in ATL buffer and RNAse A treatment). DNA concentrations were measured using the Quant-iT dsDNA Broad-Range Assay Kit (Life Technologies, Paisley, UK) and GloMax® 96 Microplate Luminometer (Promega, Southampton, UK). DNA was sent for whole genome sequencing by Illumina sequencing using the PHE Genomic Services and Development Unit (Colindale, UK) 29. Illumina Nextera DNA libraries were constructed and sequenced using the Illumina HiSeq 2500.

2.2 Bioinformatic methods

Raw Illumina reads were pre-processed as described in Kapatai et al. 2016. Kmer ID (https://github.com/phe-bioinformatics/kmerid) was used to confirm species identification as part of the PHE NGS workflow. Multi-locus Sequence Typing (MLST) on short-read data was performed using Metric Oriented Sequence Typing (Tewolde et al. 2016; https://github.com/phe-bioinformatics/MOST).

PneumoCaT (https://github.com/phe-bioinformatics/PneumoCaT) was used to predict capsular type from whole genome sequencing data as described in Kapatai *et al.* (2016). In brief, processed reads are initially mapped to capsular locus sequences for all 94 serotypes (92 serotypes plus 2 molecular subtypes) which predicts serotype or genogroup based on mapping coverage (> 90%). If

genogroup is predicted the analysis uses the CTV database and the reads are mapped to genogroup relevant genes allowing for variant analysis. Serotype can be predicted if 100% match with an available variant profile is achieved.

2.3 Assembly-based sequence analysis

Genomic reads were assembled using SPAdes (version 2.5.1) *de novo* assembly software (Bankevich et al., 2012) with the following parameters 'spades.py --careful -1 strain.1.fastq.gz -2 strain.2.fastq -t 4 -k 33,55,77,85,93'. The resulting contigs.fasta file was annotated using Prokka (Seemann, 2014) and the capsular locus was extracted using the Artemis Comparison Tool (Carver et al., 2005).

2.4 Serotyping

Isolates were initially serotyped at Public Health England (PHE) using slide agglutination, in which cultures were grown in Todd Hewitt Broth (PHE media services) for four hours, the majority of the supernatant was removed and drops of culture were mixed with either latex absorbed pool or standard group and factor antisera (all sera from SSI Diagnostica A/S, Hilleroed, Denmark) on a glass slide and observed for agglutination reactions.

Isolates were serotyped at SSI Diagnostica using the gold-standard method for Pneumococcus serotyping - Quellung reaction. The cultures were grown in serum broth (SSI Diagnostica A/S, Denmark) overnight. Two to four µL of culture were added on to a glass slide and mixed with the same amount of either pool, group or factor antisera (SSI Diagnostica A/S, Denmark). The mixture was then observed for capsule swelling reaction in a phase contrast microscope, and typed using the Key to pneumococcal factor serum from SSI Diagnostica A/S ("Key to pneumoccoccal factor serum," 2013).

2.5 Production of antiserum at SSI Diagnostica A/S

A vaccine was developed from the novel strain, and rabbits were immunised. Bleeds from the rabbits were taken throughout the immunisation period, and the titer of the reaction between antiserum and the novel strain was monitored using the Quellung reaction.

The titer of the antiserum was found by making a twofold dilution of the antisera from each bleed and mixing it with either serotype 7F, 7A, 7B, 7C and the novel strain.

After finding the reaction pattern of the antiserum, the antiserum was absorbed with serotype 7B, 7C and the novel strain, respectively.

2.6 Capsular polysaccharide purification

The polysaccharides were isolated from cultivated pneumococcus bacteria of the relevant serotypes. The 7D pneumococcus was grown in serum bouillon. The cells were harvested by centrifugation and lysed using deoxycholate. The solution was adjusted to pH 5 and the precipitate, deoxycholate and cell debris, was removed by centrifugation. Proteins were precipitated by adding $CaCl_2$ to reach 0.1% concentration followed by ethanol to reach 25% concentration and removed by filtration. Polysaccharides were precipitated by increasing the concentration of ethanol to 80% and recovered by centrifugation. Subsequently, the precipitate were dissolved in pH 5 0.05 M acetate buffer and DNAse and RNAse were added to the solution. The pH was adjusted and trypsin was added to decompose any remaining protein residues. Finally, the polysaccharides were freeze-dried after diafiltration.

2.7 NMR spectroscopy

Samples were dissolved in 99.9% D_2O (Sigma Aldrich), and spectra were recorded at 40 °C on a Bruker Avance (798.80 MHz for ^{1}H and 200.88 MHz for ^{13}C) equipped with a 5 mm TCI $^{1}H/(^{13}C, ^{15}N)$ cryoprobe using acetone as reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively). ^{1}H-NMR spectra were acquired with presaturation, sweep widths of 6393.9 Hz and 32768 points zero filled to 65536 points. ^{13}C-NMR spectra were acquired with power-gated decoupling of ^{1}H, sweep widths of 48076.9 Hz and 32768 points zero filled to 65536 points. The DQF-COSY spectra were acquired with sweep widths of 6393.9 and 6390.4 Hz in F2 and F1, respectively and 4096x512 points zero filled to 8192x1024 points. The TOCSY spectra was acquired width the same sweep widths and number of points as the DQF-COSY using MLEV17 spinlock sequence with a mixing time of 60 ms and 31887.8 Hz spinlock power. The NOESY spectra was acquired with presaturation and the same sweep widths and number of points as the DQF-COSY spectra with a mixing time of 200 ms. The multiplicity edited ^{1}H-^{13}C HSQC was recorded with 2048x512 points zero filled to 4096x1024 points with sweep widths of 6393.9 and 35354.2 Hz in F2 and F1, respectively. The ^{1}H-^{13}C HSQC-TOCSY spectra were acquired with 1024x512 points zero filled to 4096x1024 points and the same sweep widths as the HSQC spectra using the DIPSI2 spinlock sequence with a mixing time of 60 ms and 31887.8 Hz spinlock power. The long range optimized ^{1}H-^{13}C HMBC spectra, optimized for 8 Hz couplings, was recorded with 4096x512 points zero filled to 8192x1024 points with sweep widths of 6393.9 and 44192.7 Hz in F2 and F1, respectively.

The 31P NMR was recorded on a Bruker 400 AvanceIII with a 5 mm Smartprobe. The 1D 31P NMR spectra was recorded with sweep widths of 12931.0 Hz and 65536 points zero filled to

131072 points. The 1H-31P HMBC was optimized for 8 Hz coupling constants with 2406.9 and 1944.2 Hz in the F2 and F1, respectively, and 4096x512 points zero filled to 8192x1024.

The CLIP-HQSC (Enthart, Freudenberger, Furrer, Kessler, & Luy, 2008) spectra was measured on a Bruker AvanceIIIHD with a 5 mm prodigy cryoprobe using 4795.4 and 16668.5 Hz in F2 and F1, respectively, and 2048x512 points zero filled to 4096x1024 points.

With the exception of the CLIP-HSQC, all two-dimensional spectra were recorded using standard Bruker pulse sequences and were acquired using TopSpin 2.1. The CLIP-HSQC was recorded using TopSpin 3.2. All NMR data were processed using TopSpin 3.5.

3 RESULTS

When run through the PneumoCaT pipeline, the isolate demonstrated a novel codon at residue 385 which results in a novel amino acid change (CTT (Leu) compared to ACT (Thr) for serotype 40, TTT (Phe) for 7B and TGT (Cys) for 7C), the rest of the capsular variant pattern was otherwise the same as expected for 7B in the CTV database (Kapatai et al., 2016). When tested by slide agglutination at PHE a non-recognised pattern of reaction was seen with the serogroup 7 factor serum, giving reactions with both factor 7e and 7f.

The novel 7 isolate was sent to SSI Diagnostica A/S for confirmation of serotype. The isolate was confirmed to have an unusual pattern of factor sera reactions by Quellung test.

The novel 7 isolate was grown in serum broth and then tested with ImmuLex Pneumotest (SSI Diagnostica A/S, Denmark), which is a latex agglutination test. The novel 7 isolate agglutinated with Pool C and P, which shows that the isolate belonged to serogroup 7. The result was then confirmed with Group 7 antiserum, and further serotyped using factor 7b, 7c, 7e and 7f antisera. The novel 7 strain showed capsule swelling with factor 7e and 7f, the capsule swelling was very strong with factor 7f. The Key to pneumococcal factor serum from SSI Diagnostica A/S (“Key to pneumoccoccal factor serum,” 2013), adapted in table 1, show that serotype 7B only reacts with factor 7e and serotype 7C only reacts with factor 7f, so when the novel 7 isolate show a reaction with both factor 7e and 7f, this indicates that it is a new serotype within serogroup 7.

Rabbits were immunised with a vaccine made from the novel 7 strain, and when examining the antiserum from the rabbits it showed reactions with both serotype 7B, 7C and the novel strain. The antiserum showed a stronger reaction with the novel strain then with the serotype 7B and 7C strains. This also indicated that the strain was a novel strain, which share antigenic epitopes with both serotype 7B and 7C.

Table 1: Serogroup 7 agglutination results

Type	Reaction in factor serum			
	7b	7c	7e	7f
7F	+++			
7A	(+)	+++		
7B			+++	
7C				+++
7D			+	+++

Further genomic analysis at PHE showed that the novel 7 isolate was multilocus sequence type (ST) 1533, which was the same as four other serotype 7B isolates in the PneumoCaT evaluation analysis, and a double locus variant of the most common serotype 7F sequence type in PHE data, ST191. Comparison of the 7B capsular locus (CR931641.1) and the extracted capsular locus for the novel 7 isolate revealed that the two sequences differ by a single non-synonymous SNP at position 385 of *wcwK.*

In order to elucidate the structural background for the serological results and all strains were grown and polysaccharides were isolated. The yield of the 7D cps purification was 83.6 mg from 5 L cell broth, and using NanoDrop the protein impurities was determined to be 1.4%. The remaining serogroup 7 strains are commercially available (SSI Diagnostica A/S).

The NMR investigation started by inspection of standard 1D ^{1}H spectra, and all samples (7B, 7C and the new serotype) resulted in good NMR spectra as seen in figure 2, indicating that the polysaccharides in the concentrations used are well soluble. Initial NMR data of the cps of the novel 7 strain revealed that it had many similarities to the spectra of 7B and 7C. As the structure for 7C was not previously described, it was necessary to fully assign this structure before focusing on the novel 7D strain. The spectra for 7C contained signals corresponding to seven different anomeric signals, see figure 3, one of which had a phosphodiester linkage, and these were labelled **A** to **G** in order of descending ^{13}C chemical shift. Furthermore, a total of five methyl groups were observed, three of which had almost identical chemical shifts corresponding to the methyl of 6-deoxy sugars, and the other two were more downfield with chemical shifts corresponding to the methyl of an acetyl group. The assignment of the different carbohydrate units were performed using DQF-COSY, NOESY, TOCSY, HSQC, HSQC-TOCSY and HMBC.

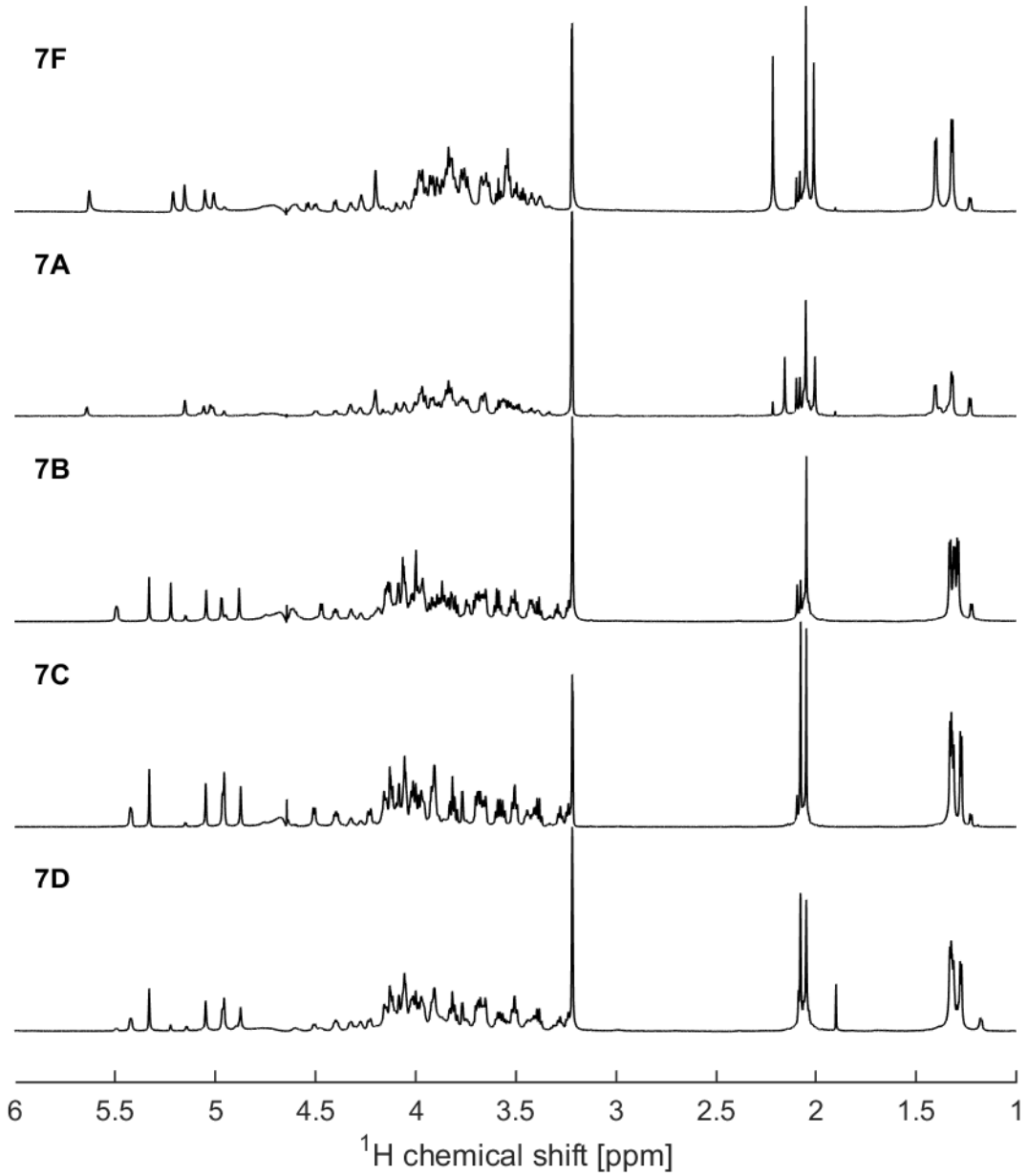


Figure 2: ^{1}H NMR spectra of serogroup 7 serotypes. From top to bottom: 7F, 7A, 7B, 7C, 7D

One of these anomeric signals, **A**, was identified as a furanoside, while the remaining six were identified as pyranosides. The anomeric configuration of the six pyranosides were determined using the $^1J_{H,C}$ coupling constants (Bock & Pedersen, 1974), shown in table A.1 of the supplementary data, and four of them, **D**, **E**, **F** and **G**, were identified as α-configuration, $^1J_{H1,C1}$ of ~170 Hz, while the remaining two, **B** and **C**, were identified as having β-configuration, $^1J_{H1,C1}$ of ~160 Hz.

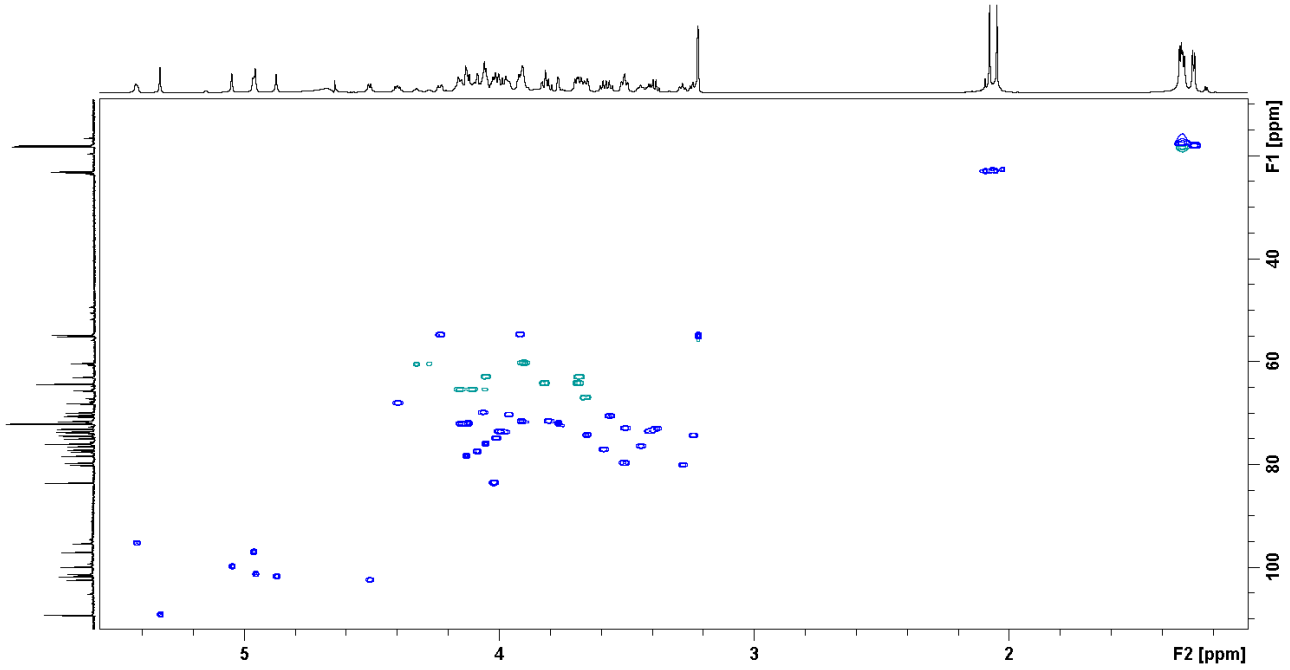


Figure 3: HSQC spectrum of the 7C sample. The ^{1}H and ^{13}C projections are from separate experiments

The three pyranosides **C**, **D** and **E** had very similar couplings constants (table A.1 of the supplementary data), including the $^{3}J_{H1,H2}$ coupling constant, even though only one of them were of β-configuration, and were determined to be rhamnoses, and all three monosubstituted. The presence of three rhamnoses would also explain the presence of the three upfield methyl groups described earlier. The β-rhamnose **C** was determined to be 2-substituted, as evident by its downfield chemical shift, as was the α-rhamnose **E** while the other α-rhamnose, **D**, was 4-substituted. The rhamnoside configurations were determined by the $^{3}J_{H,H}$ coupling constants, which corresponded to 6-deoxy mannose configurations with the 2-position proton being equatorial, and the 3, 4 and 5-position protons being axial.

The pyranoside **B** was determined to be a β-glucose, as it only has large $^{3}J_{H,H}$ coupling constants, and was found to be 4-substituted, based on the downfield chemical shift of the respective position. The pyranoside **F** was also of glucose configuration, but contained a 2-*N*-acetyl, and had α-configuration. **F** was found to be 6-substituted, and similar to the anomeric position of the final pyranoside, **G**, the 6-position was connected to a phosphodiester. The last pyranoside, **G**, was determined to be an *N*-acetyl α-galactosamine and was 3,4-substituted as well as being anomerically bound to a phosphodiester. This galactoside configuration was determined mainly from the $^{3}J_{H,H}$ coupling constants, as seen in table A.1 of the supplementary data, as the chemical shifts were quite heavily shifted due to the disubstitution compared to unsubstituted galactose (Agrawal, 1992). Finally, the furanoside **A** was determined to be a terminal β-ribose as the

chemical shifts matched those of methylated β-ribofuranose (Agrawal, 1992), as well as those of the corresponding β-ribofuranose in 7B (Jansson et al., 1991).

The only thing remaining was connecting the different monosaccharide units, which was done using HMBC and NOESY (All assigned correlations is seen in table Y in supplementary data). Starting from the 6-substituted α-GlcpNAc **F**, the anomeric proton had HMBC and NOESY correlations to the 2-position of the α-rhamnose **E**, as well as NOESY correlations to the anomeric proton of **E**, further confirming the 2-substitution. From the anomeric position of **E** there were HMBC and NOESY correlations to the 2-position of the β-rhamnose **C**, the anomeric of which had HMBC and NOESY correlations with the 4-position of the β-glucose **B**. This in turn had correlations to the 4-position of the disubstituted α-GalpNAc, which were the last part of the backbone of the repeating unit as the anomeric position of **G** were connected to the same phosphodiester as the 6-position of **F**, which was confirmed by ^{1}H, ^{31}P HMBC (figure A.17 in supplementary data). The terminal β-ribose **A** had HMBC and NOESY correlations to the 4-position of the 4-position of the α-rhamnose **D**, which in turn had correlations with the 3-position of **G**. This means the capsular polysaccharide structure of 7C was almost identical to that of 7B, with the only difference being the α-GalpNAc **G** instead of an α-glucose, and all of the signals from **A**, **C**, **E**, and **F** where practically identical to their counterparts in 7B. The signals from **B** and **D**, which were the two carbohydrates directly connected to the changed unit, had shifted to some extent, although mostly on the anomeric position. The full assignment of the 7C capsular polysaccharide is given in table 2.

286 Table 2: Table containing the assignments of 7C, 7B and 7D as it contains both repeating units in a 5:1 ratio.

7C table, intensity 5 in 7D	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Ribf-(1- **A**	5.330 108.84	4.054 75.63	4.124 71.64	4.024 83.17	3.691/3.824 63.77	N/A	N/A	N/A
-4)-α-Rhap-(1- **D**	4.956 100.9	3.771 71.39	3.916 71.24	3.510 79.31	4.401 67.71	1.276 17.60	N/A	N/A
O-P-6)-α-GlcpNAc-(1- **F**	4.964 96.63	3.921 54.39	3.808 71.18	3.571 70.25	4.154 71.72	4.113/4.154 65.17	2.049 22.60	N/A 174.99
-2)-α-Rhap-(1- **E**	5.049 99.48	4.089 77.10	3.967 70.01	3.508 72.62	4.066 69.56	1.324 17.54	N/A	N/A
-2)-β-Rhap-(1- **C**	4.874 101.33	4.131 77.95	3.659 73.93	3.39 72.69	3.414 73.19	1.317 17.77	N/A	N/A
-4)-β-Glcp-(1- **B**	4.509 102.01	3.240 74.02	3.594 76.71	3.283 79.69	3.447 76.06	3.688/4.054 62.63	N/A	N/A
-3,4)-α-GalpNAc-(1-O-P- **G**	5.422 94.94	4.234 54.45	4.014 74.56	3.976 73.37	4.002 73.25	3.909 59.94	2.076 22.78	N/A 174.87
7B table, intensity 1 in 7D	**1**	**2**	**3**	**4**	**5**	**6**	**NAc-Me**	**C=O**
β-Ribf-(1-	5.330 108.85	4.062 75.61	4.146 71.61	4.022 83.10	3.699/3.837 63.89	N/A	N/A	N/A
-4)-α-Rhap-(1-	5.222 100.89	3.999 71.17	3.989 71.28	3.524 79.64	4.401 67.50	1.290 17.61	N/A	N/A
O-P-6)-α-GlcpNAc-(1-	4.968 96.56	3.933 54.35	3.804 71.11	3.596 70.13	4.145 71.67	4.119/4.188 64.99	2.048 22.56	N/A 174.96
-2)-α-Rhap-(1-	5.046 99.42	4.087 76.98	3.967 69.96	3.506 72.53	4.062 69.53	1.331 17.47	N/A	N/A
-2)-β-Rhap-(1-	4.881 101.24	4.130 77.87	3.657 73.90	3.385 72.64	3.421 73.14	1.330 17.54	N/A	N/A
-4)-β-Glcp-(1-	4.473 102.00	3.237 74.05	3.586 76.59	3.292 79.54	3.433 76.01	3.687/4.066 62.61	N/A	N/A
-3,4)-α-Glcp-(1-O-P-	5.493 95.95	3.746 73.26	4.003 75.29	3.869 72.81	3.967 72.86	3.893 59.90	N/A	N/A

287 (N/A for non applicable)

After having elucidated the structure of 7C it was evident from the 7D spectra, as seen in the 1D spectra in figure 2 and overlay of the anomeric region of the 2D HSQC spectra, figure 4, that it was a combination of the structures of 7B and 7C. The relative ratio was approximate 1:5 between 7B and 7C, as measured by integrals from ^{1}H NMR, ^{31}P NMR (figure A.8 in supplementary data) and HSQC. It cannot be determined based on the current data if the polysaccharide for 7D is one long chain with both structural elements present, or if it is a mixture of the two structures in separate chains. The peaks corresponding to the CWPS (Skovsted et al., 2007), which is present in all the samples, was in higher concentration in the 7D sample, and the sample also had a higher salt concentration compared to the 7B and 7C samples. As CWPS is highly charged, containing both phosphocholine groups and backbone phosphodiesters, some of the signals have shifted quite a lot compared to the CWPS present in the 7B and 7C samples. This point to the fact that sample preparation and concentration should be kept in mind, when comparing chemical shifts to previous studies.

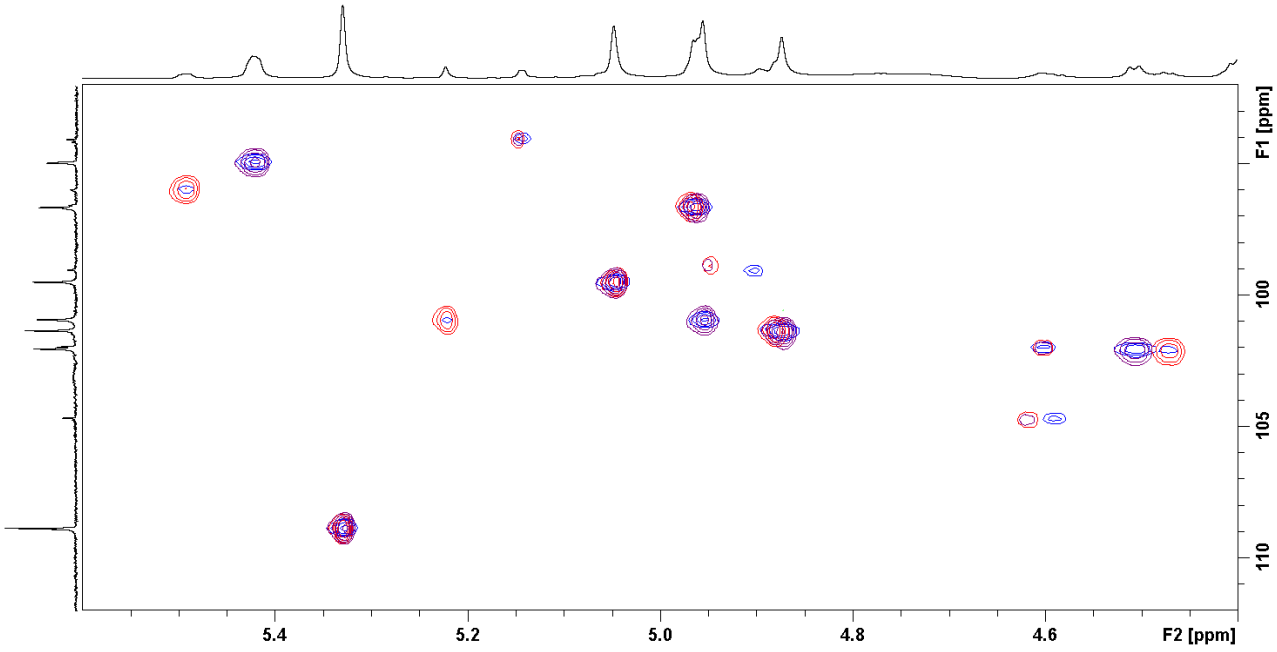


Figure 4: HSQC of 7D (blue) with 7B (red) and 7C (purple) overlayed. The projections are from 7D 1H and 13C NMR, note that the signals from 7C are more intense in 7D compared to signals originating form 7B

4 Discussion

With modern developments in both serological and genomic techniques the sensitivity for detection of novel variants of previously described serotypes has improved and many potential novel pneumococcal serotypes have been reported. However the question remains, when is a novel variant a real “new serotype”? In the case of pneumococcus, the phenotypical serotype is the most

important determinant for surveillance purposes as vaccine-driven immunity against the pneumococcus is based on serotype-specific immune responses to polysaccharide and conjugated-polysaccharide vaccines. Therefore the determination of a “serotype” should be aligned with the immunological properties of the organism in order to accurately reflect the efficacy of a vaccination program.

As observed in the agglutination test, the novel serotype described here was determined to be part of serogroup 7 and acted as a hybrid between serotype 7C and 7B, although more like 7C than 7B. After full structure elucidation of the new cps, as well as the previously isolated 7C, it was evident that this serological observation was is in good agreement with the structure elucidation where the 7D cps was identified as an approximately 5:1 mixture of the 7C and 7B cps, figure 4, respectively. The full assignment for 7D, as well as 7C and 7B, can be found in table 2. However, whether the 7C and 7B cps repeating units are both present as separate cps, or if it is mixed into a hybrid cps is still unknown. This would also serve to explain why 7D reacts with antiserum of rabbits immunised with 7B or 7C, as both repeating units are present, and why it had a stronger reaction with factor 7f than 7e in the agglutination test, as it has higher amounts of the repeating unit from 7C than from 7B.

As the structure of the capsular polysaccharide of 7C was previously unknown, it was fully assigned, see table 2, and as it turned out the only difference between 7B and 7C is the presence of a branched α-GalpNAc in 7C instead of the branched α-Glcp in 7B, see scheme 1. As result of this, the NMR spectra of 7B and 7C are quite similar, with the biggest differences, aside from the changed monosaccharide, being from the positions directly attached to said monosaccharide. The only part of the serogroup 7 capsular polysaccharides that is maintained in all five serotypes is the presence of the α-GlcpNAc-1,2-α-Rhap, although in 7F and 7A these are on a branch off of the backbone and in 7B and 7C, and thus also 7D, it is part of the backbone. Another part of the cps that is only partly maintained across the serogroup is the β-L-Rhap-1,4- β-D-Glcp, which is present in 7B, 7C and 7D, whereas in 7F and 7A it the β-L-Rhap is 2-*O*-acetylated.

The new serotype, 7D in the Danish serotype system, was both serologically, genetically and structurally different from other serotypes, although the cps was a hybrid of 7C/7B. Its hybrid nature was also evident from the rabbit sera results, in which 7D reacted like both 7C and 7B, but mostly like 7C. Compared to the other reported hybrids in serogroup 6 (Oliver et al., 2013; Park et al., 2015; Sheppard et al., 2010), in which 6F, 6G and 7H are hybrids of 6A/6C, 6B/6D and 6A/6B, respectively. In the case of 6F, which genetically it is almost like 6A, but in the enzyme that differs between production of 6A or 6C there is a single point mutation, resulting in the ability to produce

both. This is similar to the case for 7D, which is mostly like 7B, but the enzyme responsible for the α-Glcp unique to 7B has been mutated to be capable of transferring both α-Glcp and α-GalpNAc, making the cps a mixture of 7C and 7B. While there is no direct evidence of the two repeating units being covalently bound, the overall results indicates that they are, as the expressed glycosyltransferase unique to 7D can transfer both the α-Glcp and the α-GalpNAc, and antiserum from 7D immunised rabbits also reacted with 7B and 7C.

In the future, in order to avoid confusion on the naming of pneumococcal“serotypes”, it may be best if there was an official overarching body with rules for accepting new serotype designations. For example that the isolate should demonstrate both a genetic difference and a serological difference as demonstrated by rabbit sera (enabling possible commercial supply of typing sera to identify it) plus elucidation of a distinct structural difference to other serotypes. Other variants noted by genetic analysis or only demonstrated by use of specific monoclonal antibodies, should be named in a different way to avoid confusion. In order for this to happen there needs to be a worldwide agreement on typing methods and nomenclature for both serological and genotypic capsular types. With serotypes designated only if there are widely available serologically based methods to determine them and enable global comparison between laboratories..

5 Acknowledgements

C.K. acknowledges the Novo Nordisk foundation for funding: Biotechnology-based synthesis and production research, program grant no 5371. The 800 MHz NMR spectra were recorded on the 800 MHz spectrometer of the Danish National Instrument Center for NMR Spectroscopy of Biological Macromolecules at the Carlsberg Laboratory and the Technical University of Denmark. We thank Katrine Helander Pedersen (SSI Diagnostica, Denmark) for doing all the serotyping and immunization of rabbits during this study.

6 References

Agrawal, P. K. (1992). NMR Spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry*, *31*(10), 3307–3330. http://doi.org/10.1016/0031-9422(92)83678-R

AlonsoDeVelasco, E., Verheul, A. F., Verhoef, J., & Snippe, H. (1995). Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. *Microbiological Reviews*, *59*(4), 591–603. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239389&tool=pmcentrez&rendertype=abstract

Backman-Marklund, I., Jansson, P.-E., Lindberg, B., & Henrichsen, J. (1990). STUDIES OF THE CAPSULAR POLYSACCHARIDE FROM Streptococcus pneumoniae TYPE 7A. *Carbohydrate Research*, *198*, 67–77. http://doi.org/10.1016/0008-6215(90)84277-2

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *19*(5), 455–77. http://doi.org/10.1089/cmb.2012.0021

Bock, K., & Pedersen, C. (1974). A study of 13CH coupling constants in hexopyranoses. *J. Chem. Soc., Perkin Trans. 2*, (3), 293–297. http://doi.org/10.1039/p29740000293

Burton, R. L., Geno, K. A., Saad, J. S., & Nahm, M. H. (2016). Pneumococcus with the “6E” cps Locus Produces Serotype 6B Capsular Polysaccharide. *Journal of Clinical Microbiology*, *54*(4), 967–971. http://doi.org/10.1128/JCM.03194-15

Calix, J. J., & Nahm, M. H. (2010). A New Pneumococcal Serotype, 11E, Has a Variably Inactivated *wcjE* Gene. *The Journal of Infectious Diseases*, *202*(1), 29–38. http://doi.org/10.1086/653123

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, *21*(16), 3422–3423. http://doi.org/10.1093/bioinformatics/bti553

Enthart, A., Freudenberger, J. C., Furrer, J., Kessler, H., & Luy, B. (2008). The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings. *Journal of Magnetic Resonance*, *192*(2), 314–322. http://doi.org/10.1016/j.jmr.2008.03.009

396 Geno, K. A., Saad, J. S., & Nahm, H. (2017). Discovery of Novel Pneumococcal Serotype 35D, a
397 Natural WciG-Deficient Variant of Serotype 35B. *Journal of Clinical Microbiology*, *55*(5),
398 1416–1425. http://doi.org/10.1128/JCM.00054-17

399 Henrichsen, J. (1995). Six newly recognized types of Streptoccus penumoniae . *J. Clin. Microbiol*,
400 *33*(10), 2759–2762.

401 Jansson, P. E., Lindberg, J., Swarna Wimalasiri, K. M., & Henrichsen, J. (1991). The structure of
402 the capsular polysaccharide from Streptococcus pneumoniae type 7B. *Carbohydrate*
403 *Research*, *217*, 171–180. http://doi.org/10.1016/0008-6215(91)84127-Z

404 Kamerling, J. P. (2000). Pneumococcal polysaccharides: A chemical view. In A. Tomasz (Ed.),
405 *Streptococcus pneumoniae: Molecular biology & mechanisms of Disease.* (pp. 81–114). New
406 York, NY: Mary Ann Liebert.

407 Kapatai, G., Sheppard, C. L., Troxler, L. J., Litt, D. J., Furrer, J., Hilty, M., & Fry, N. K. (2017).
408 Pneumococcal 23B molecular subtype identified using whole genome sequencing. *Genome*
409 *Biology and Evolution*, *9*(8), 2145–2158. http://doi.org/10.1093/gbe/evx092

410 Kapatai, G., Sheppard, C. L., Al-Shahib, A., Litt, D. J., Underwood, A. P., Harrison, T. G., & Fry, N.
411 K. (2016). Whole genome sequencing of Streptococcus pneumoniae: development,
412 evaluation and verification of targets for serogroup and serotype prediction using an
413 automated pipeline. *PeerJ*, *4*, e2477. http://doi.org/10.7717/peerj.2477

414 Key to pneumoccoccal factor serum. (2013). Retrieved October 26, 2017, from
415 http://www.ssidiagnostica.com/-/media/Admin/Diagnostica-Downloads/Downloads-
416 UK/Brochures/BrochurePneumococcal-factor-antisera-key-18058.ashx

417 Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kaltoft, M. S., Reeves, P. R., … Spratt,
418 B. G. (2007). Genetic Relatedness of the Streptococcus pneumoniae Capsular Biosynthetic
419 Loci. *Journal of Bacteriology*, *189*(21), 7841–7855. http://doi.org/10.1128/JB.00836-07

420 McEllistrem, M. C., & Nahm, M. H. (2012). Novel pneumococcal serotypes 6C and 6D: Anomaly or
421 harbinger. *Clinical Infectious Diseases*, *55*(10), 1379–1386. http://doi.org/10.1093/cid/cis691

422 Moreau, M., Richards, J. C., & Perry, M. B. (1988). APPLICATION OF HIGH-RESOLUTION
423 N.M.R. SPECTROSCOPY TO THE ELUCIDATION OF THE STRUCTURE OF THE
424 SPECIFIC CAPSULAR POLY-SACCHARIDE OF Streptococcus pneumoniae TYPE 7F*.
425 *Carbohydrate Research*, *182*, 79–99. http://doi.org/10.1016/0008-6215(88)84093-X

Oliver, M. B., Van Der Linden, M. P. G., Küntzel, S. A., Saad, J. S., & Nahm, M. H. (2013). Discovery of streptococcus pneumoniae serotype 6 variants with glycosyltransferases synthesizing two differing repeating units. *Journal of Biological Chemistry*, *288*(36), 25976–25985. http://doi.org/10.1074/jbc.M113.480152

Park, I. H., Geno, K. A., Yu, J., Oliver, M. B., Kim, K. H., & Nahm, M. H. (2015). Genetic, biochemical, and serological characterization of a new pneumococcal serotype, 6H, and generation of a pneumococcal strain producing three different capsular repeat units. *Clinical and Vaccine Immunology*, *22*(3), 313–318. http://doi.org/10.1128/CVI.00647-14

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. http://doi.org/10.1093/bioinformatics/btu153

Sheppard, C. L., Pichon, B., George, R. C., & Hall, L. M. C. (2010). Streptococcus pneumoniae isolates expressing a capsule with epitopes of both serotypes 6A and 6B. *Clinical and Vaccine Immunology*, *17*(11), 1820–1822. http://doi.org/10.1128/CVI.00335-10

Skovsted, I. C., Kerrn, M. B., Sonne-Hansen, J., Sauer, L. E., Nielsen, A. K., Konradsen, H. B., … Duus, J. (2007). Purification and structure characterization of the active component in the pneumococcal 22F polysaccharide capsule used for adsorption in pneumococcal enzyme-linked immunosorbent assays. *Vaccine*, *25*(35), 6490–6500. http://doi.org/10.1016/j.vaccine.2007.06.034

Tewolde, R., Dallman, T., Schaefer, U., Sheppard, C. L., Ashton, P., Pichon, B., … Underwood, A. (2016). MOST: a modified MLST typing tool based on short read sequencing. *PeerJ*, *4*, e2308. http://doi.org/10.7717/peerj.2308

van Tonder, A. L., Bray, J. E., Quirk, S. J., Haraldsson, G., Jolley, K. A., Maiden, M. C., … Brueggemann, A. B. (2016). Putatively novel serotypes and the potential for reduced vaccine effectiveness: capsular locus diversity revealed among 5,405 pneumococcal genomes. *Microbial Genomics*, *2*(10), 90. http://doi.org/10.1099/mgen.0.000090

Yu, J., Carvalho, M. D. G. S., Beall, B., & Nahm, M. H. (2008). A rapid pneumococcal serotyping system based on monoclonal antibodies and PCR. *Journal of Medical Microbiology*, *57*(2), 171–178. http://doi.org/10.1099/jmm.0.47549-0

Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R., & Spratt, B. G. (2007). Predicted functions and linkage specificities of the products of the Streptococcus pneumoniae capsular

456 biosynthetic loci. *Journal of Bacteriology*, *189*(21), 7856–7876.
457 http://doi.org/10.1128/JB.00837-07

458

Supplementary material

Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D

Christian Kjeldsen[a], Sofie Slott[a], Pernille Landsbo Elverdal[b], Carmen L. Sheppard[c], Georgia Kapatai[c], Norman K. Fry[c], Ian C. Skovsted[b], Jens Ø. Duus[a,*]

[a] Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark

[b] SSI Diagnostica A/S, Hilleroed, Denmark

[c] Respiratory and Vaccine Preventable Bacterial Reference Unit, Public Health England – National Infection Service, Colindale, London, UK

*Corresponding author. jduus@kemi.dtu.dk, +45 45252451

Abstract

Streptococcus pneumoniae is characterised into 92 serotypes based on antigenic reactions of commercial rabbit sera to the capsular polysaccharides. During development of a bioifromatical serotyping tool (PneumoCaT), an isolate exhibited a novel codon at residue 385 of the glycosyltransferase gene *wcwK* encoding a distinct amino acid, which differentiates genogroup 7. Investigation by repeat serotyping and Quellung reaction revealed a novel pattern of factor sera with the isolate reacting very strongly with 7f, but also with 7e factor sera. The structure of the capsular polysaccharide was determined by NMR spectroscopy to be an approximately 5:1 mixture between the structures of 7C and 7B, respectively, and the structure of 7C was also elucidated. All data from whole genome sequencing, NMR spectroscopy, production of antisera and serotyping of the novel 7 strain shows that it is a new serotype, which will be named in the Danish nomenclature as 7D.

Keywords: NMR spectroscopy, structure elucidation, *Streptococcus pneumoniae*, whole genome sequence, hybrid serotype

Table of Contents

Overview of supplementary material

For the two serogroup 7 cps structures described, a full set of NMR spectra is included (DQF-COSY, TOCSY, NOESY, HSQC, HSQC-TOCSY and HMBC as well as ^{31}P spectra). Additionally, for the 7C cps a table containing of list of measurable $^{1}J_{H,C}$ and $^{3}J_{H,H}$ coupling constants as well as a table containing anomeric HMBC and NOE correlations.

For the three previously described serogroup 7 serotypes, a single HSQC of each as well as a table with chemical shift assignment is included, aside for the 7B cps assignment, which is part of table 2 (7D assignment).

Serogroup 7 ^{13}C spectra

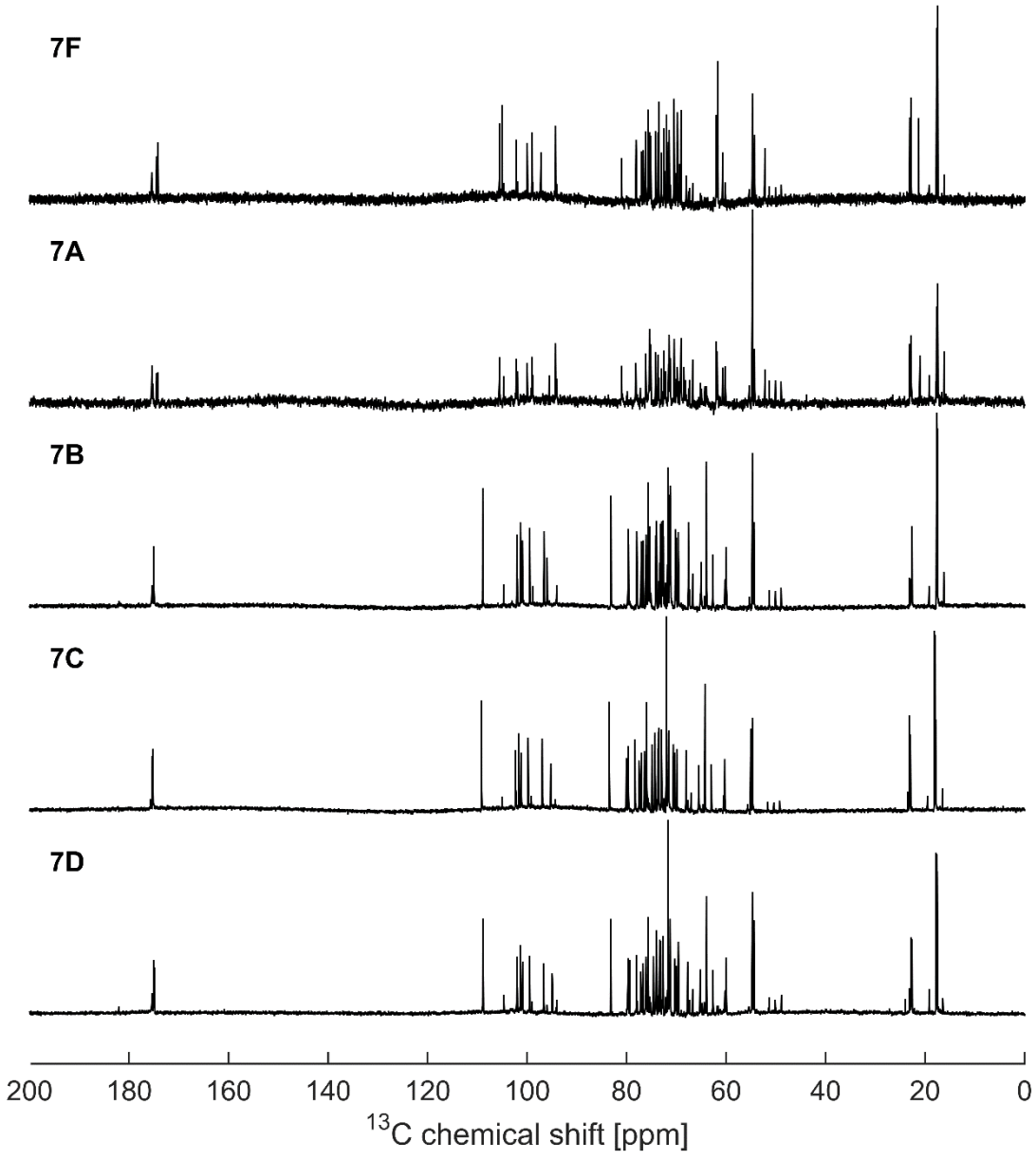


Figure A.1: ^{13}C NMR spectra of serogroup 7 serotypes. From top to bottom: 7F, 7A, 7B, 7C, 7D

7D supplementary material

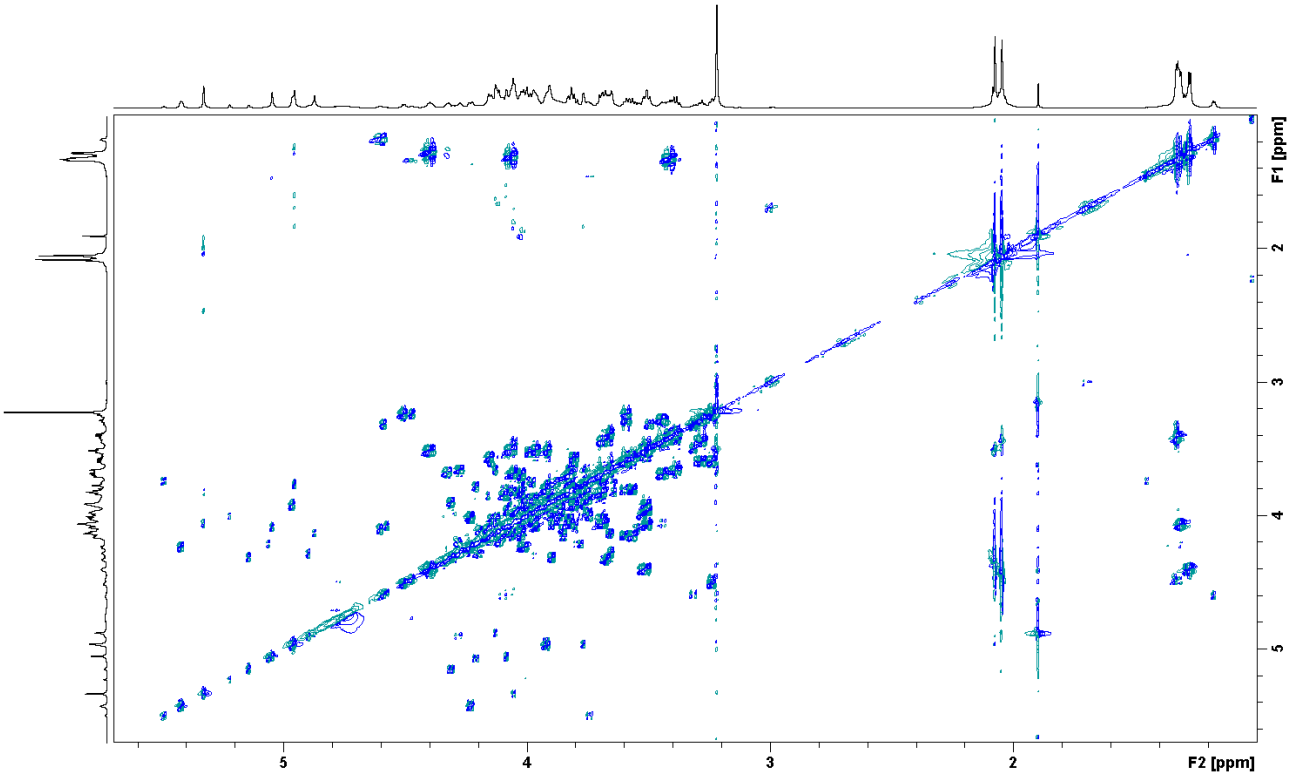


Figure A.2: DQF-COSY spectrum of the 7D sample. The projections are from separate experiments.

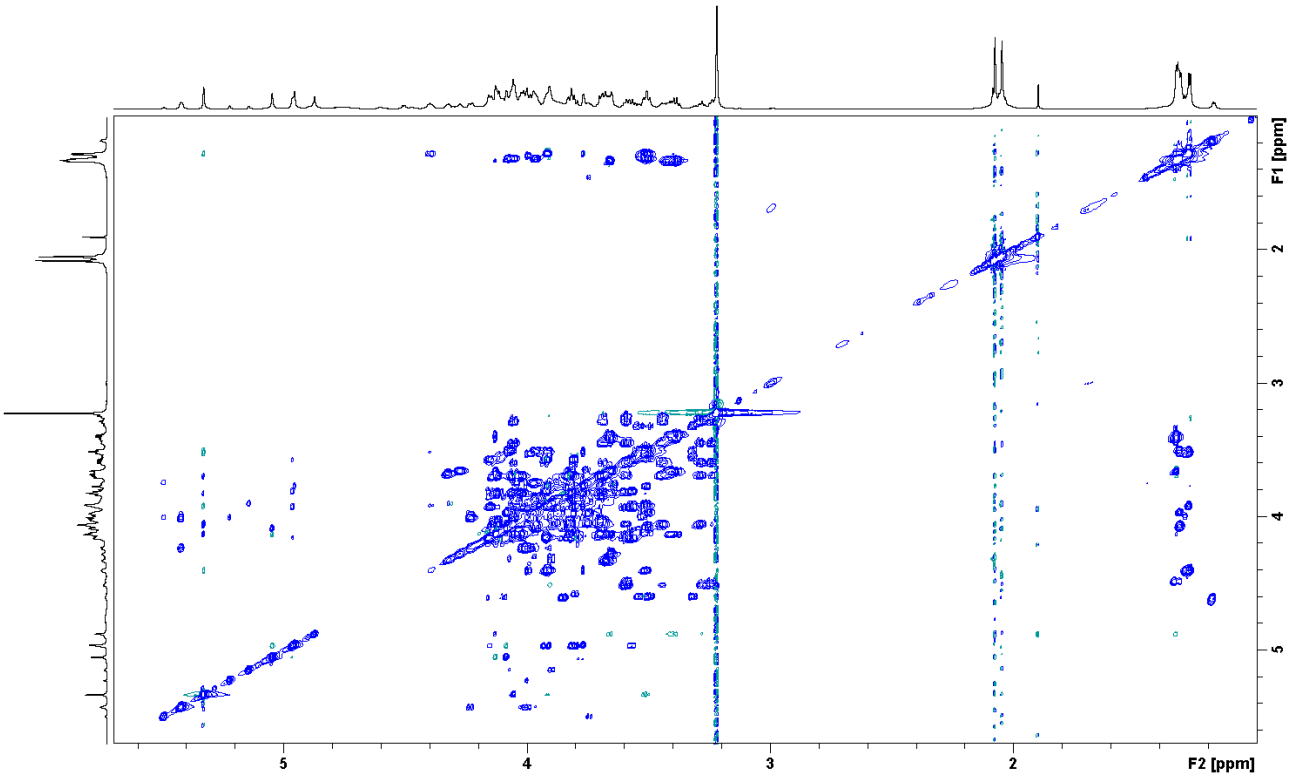


Figure A.3: TOCSY spectrum of the 7D sample. The projections are from separate experiments.

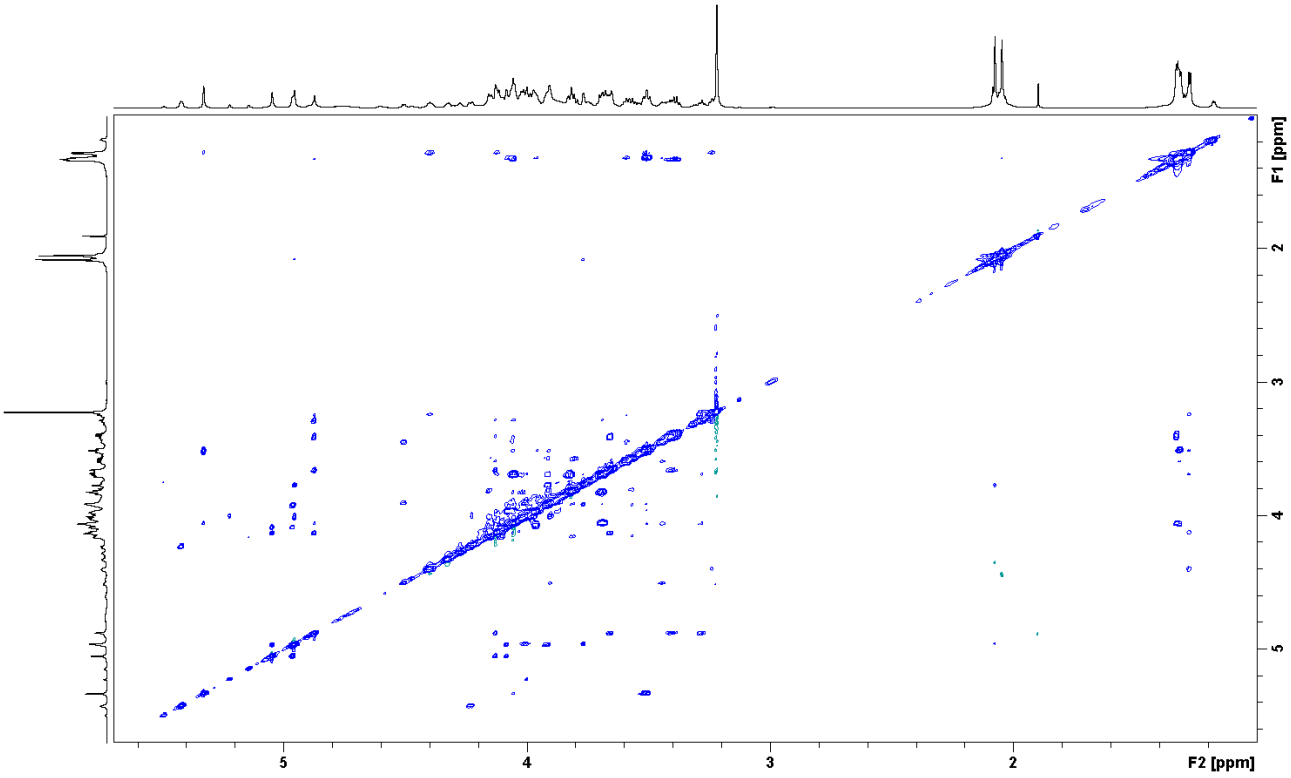


Figure A.4: NOESY spectrum of the 7D sample. The projections are from separate experiments.

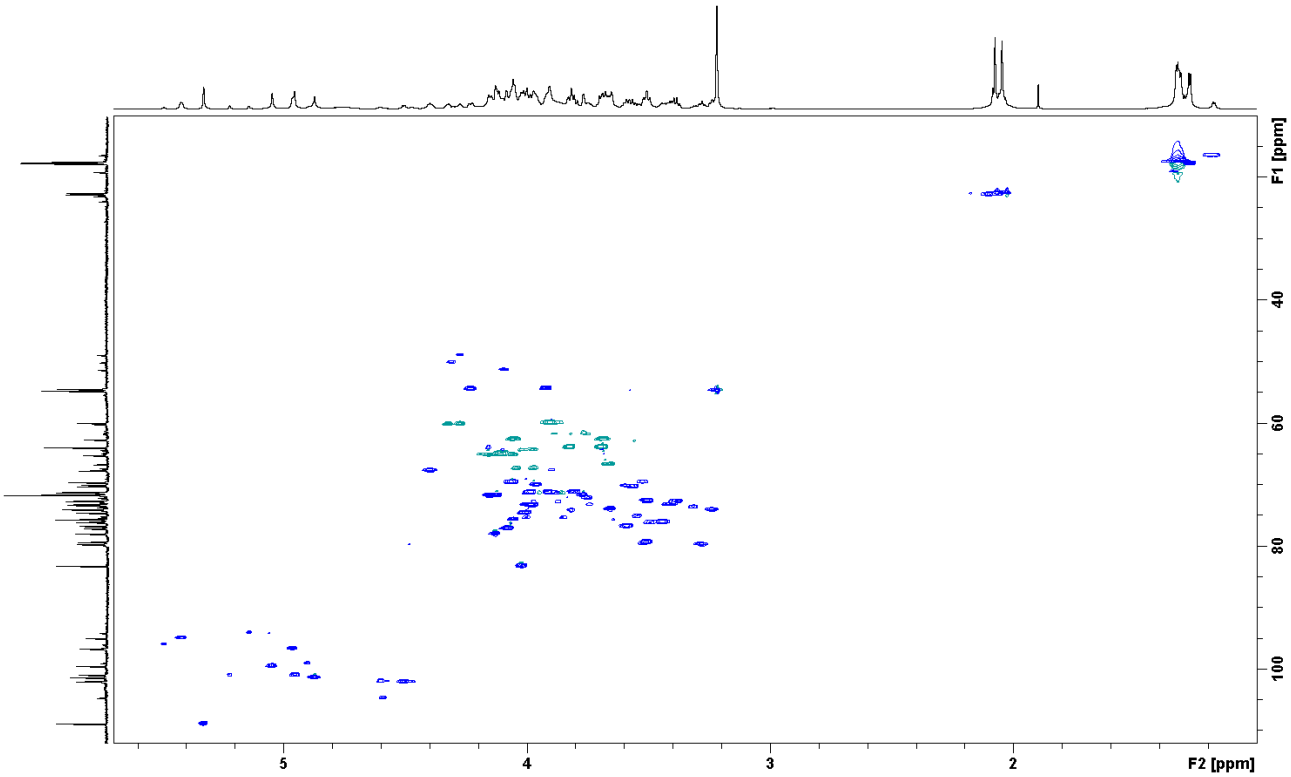


Figure A.5: HSQC spectrum of the 7D sample. The projections are from separate experiments.

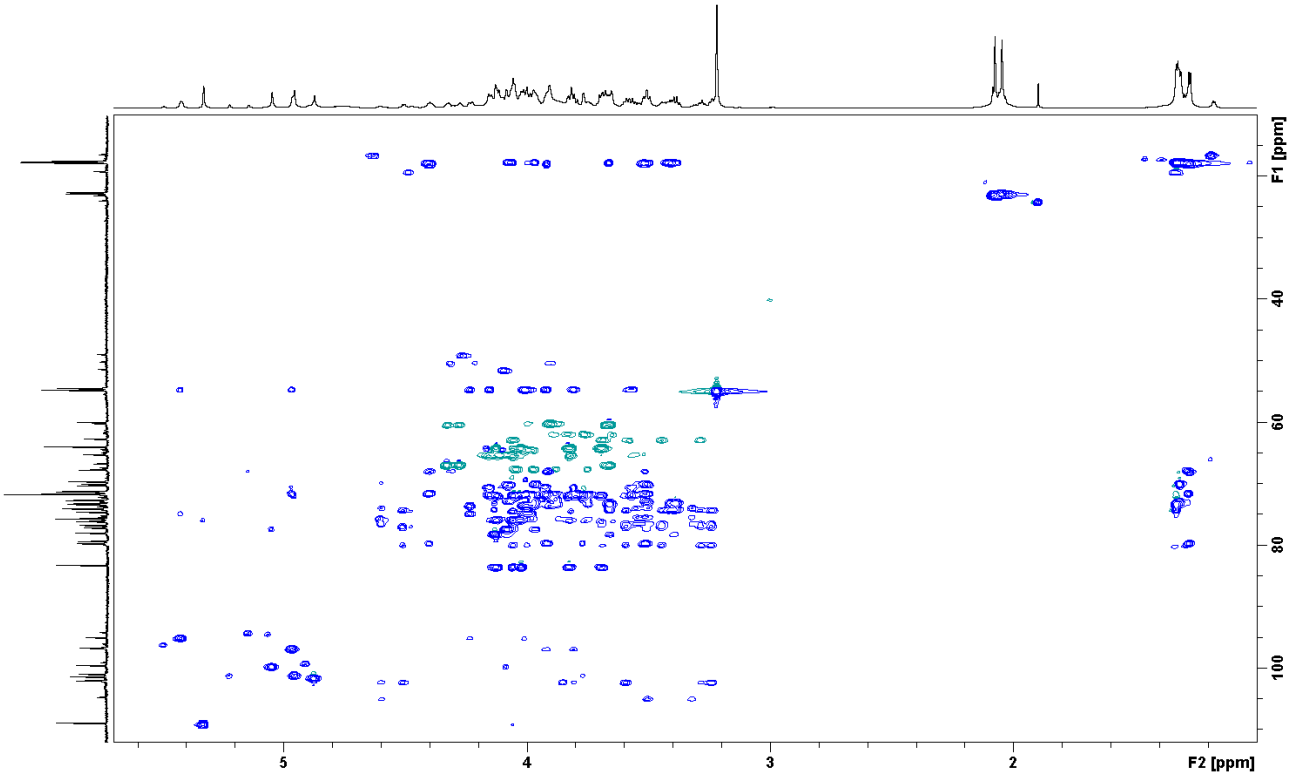


Figure A.6: HSQC-TOCSY spectrum of the 7D sample. The projections are from separate experiments.

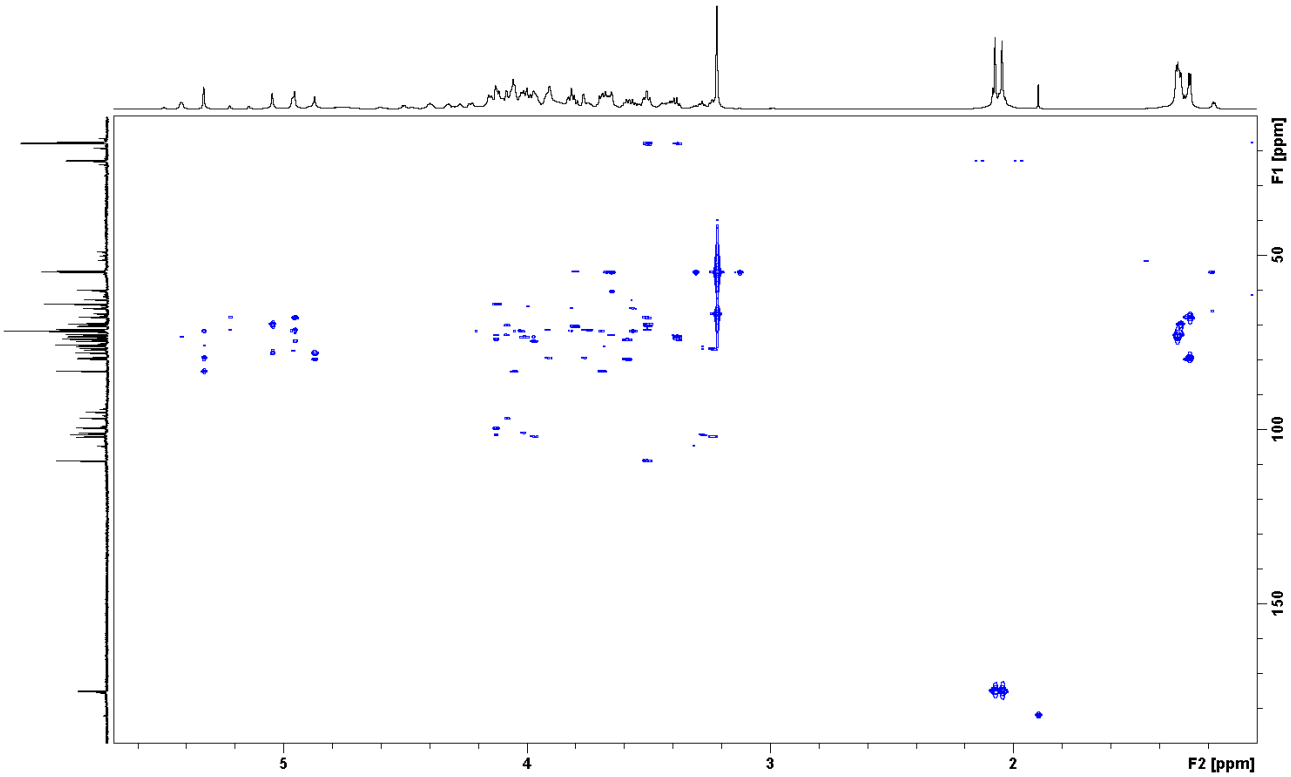


Figure A.7: HMBC spectrum of the 7D sample. The projections are from separate experiments.

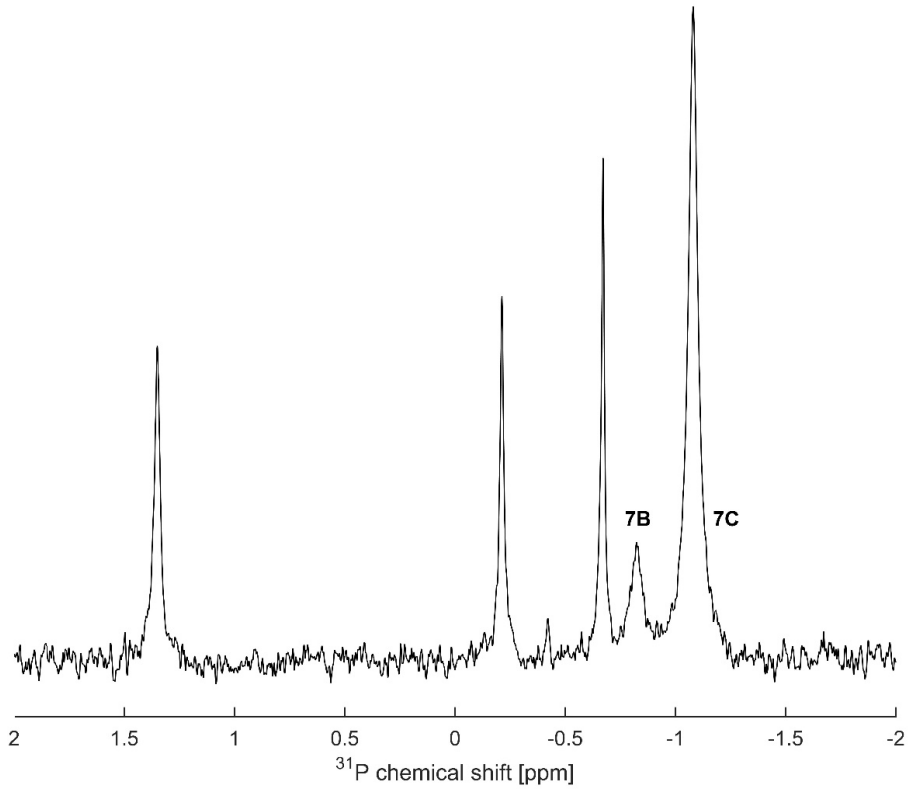


Figure A.8: ^{31}P NMR spectrum of sample 7D. Not calibrated. The peaks originating from 7C and 7B are labelled, and the remaining are from CWPS.

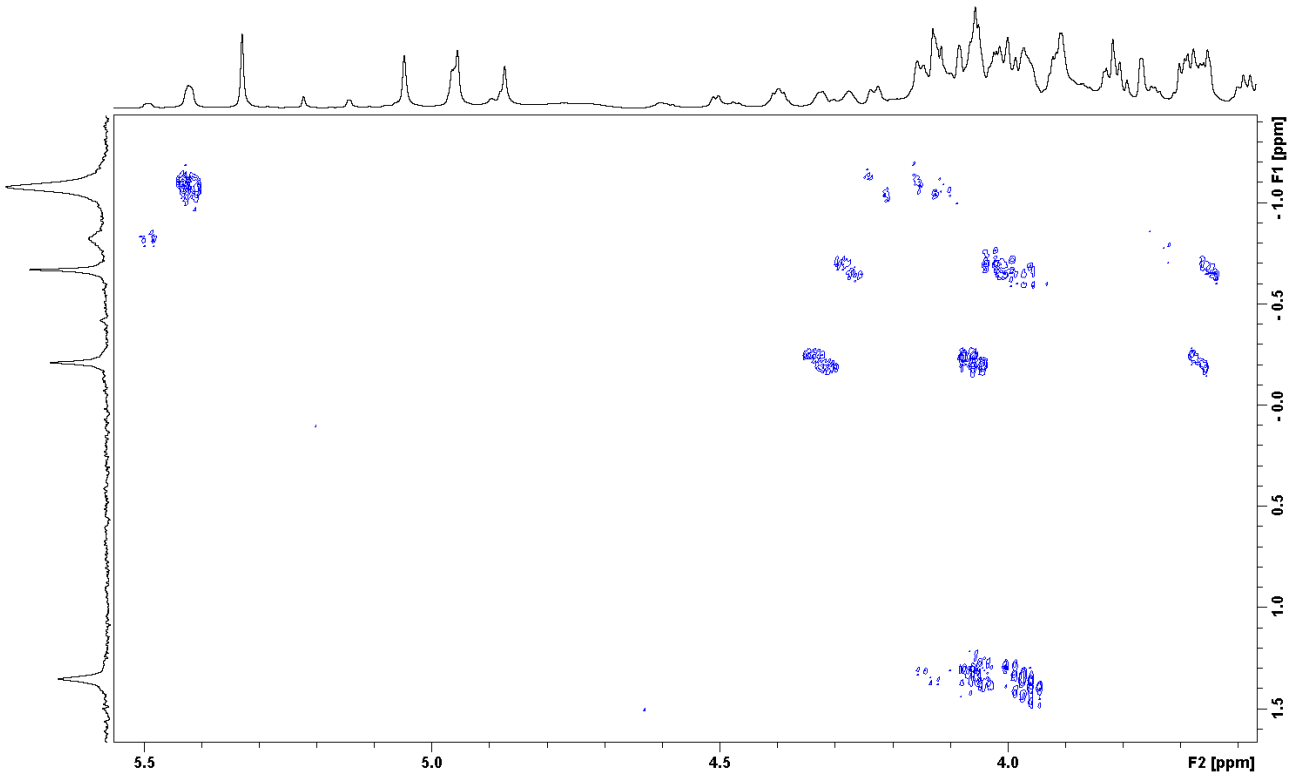


Figure A.9: ^{1}H,^{31}P HMBC spectrum of sample 7D. ^{31}P is not calibrated. The projections are from separate experiments.

7C NMR spectra

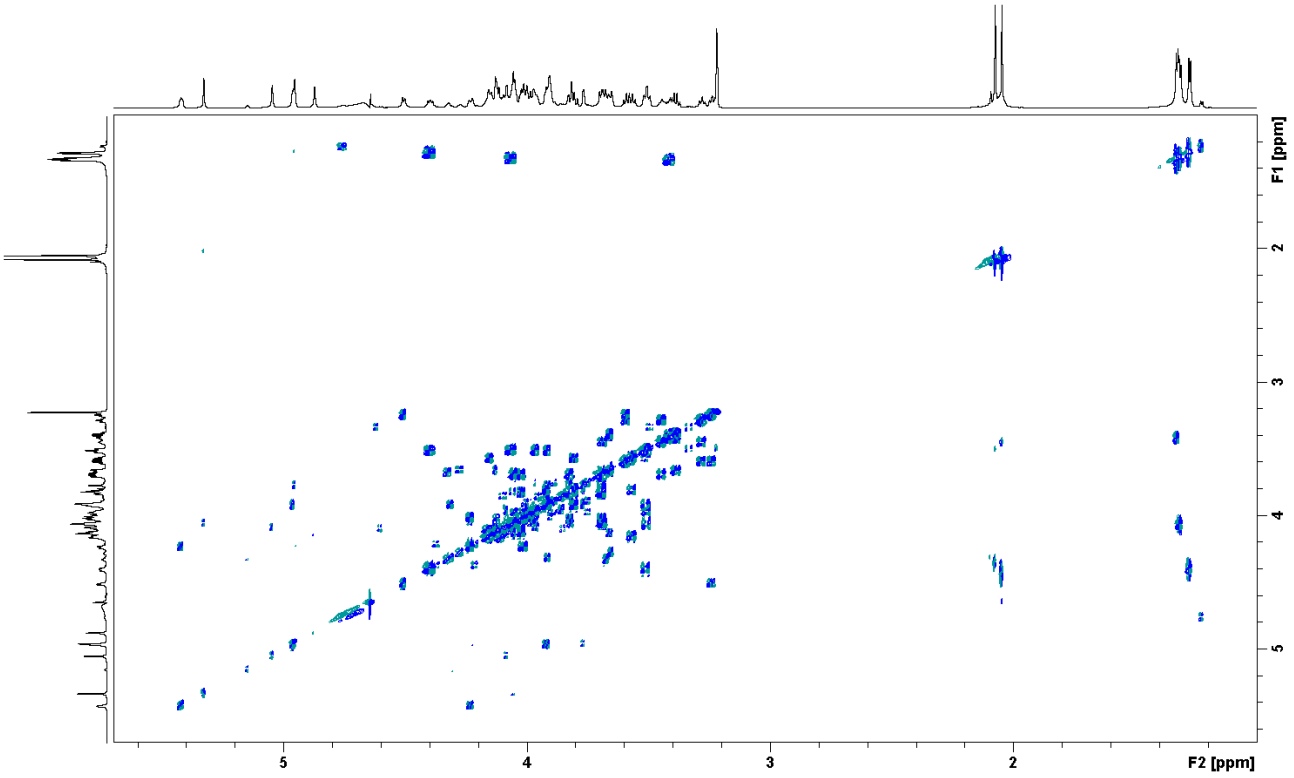


Figure A.10: DQF-COSY spectrum of the 7C sample. The projections are from separate experiments.

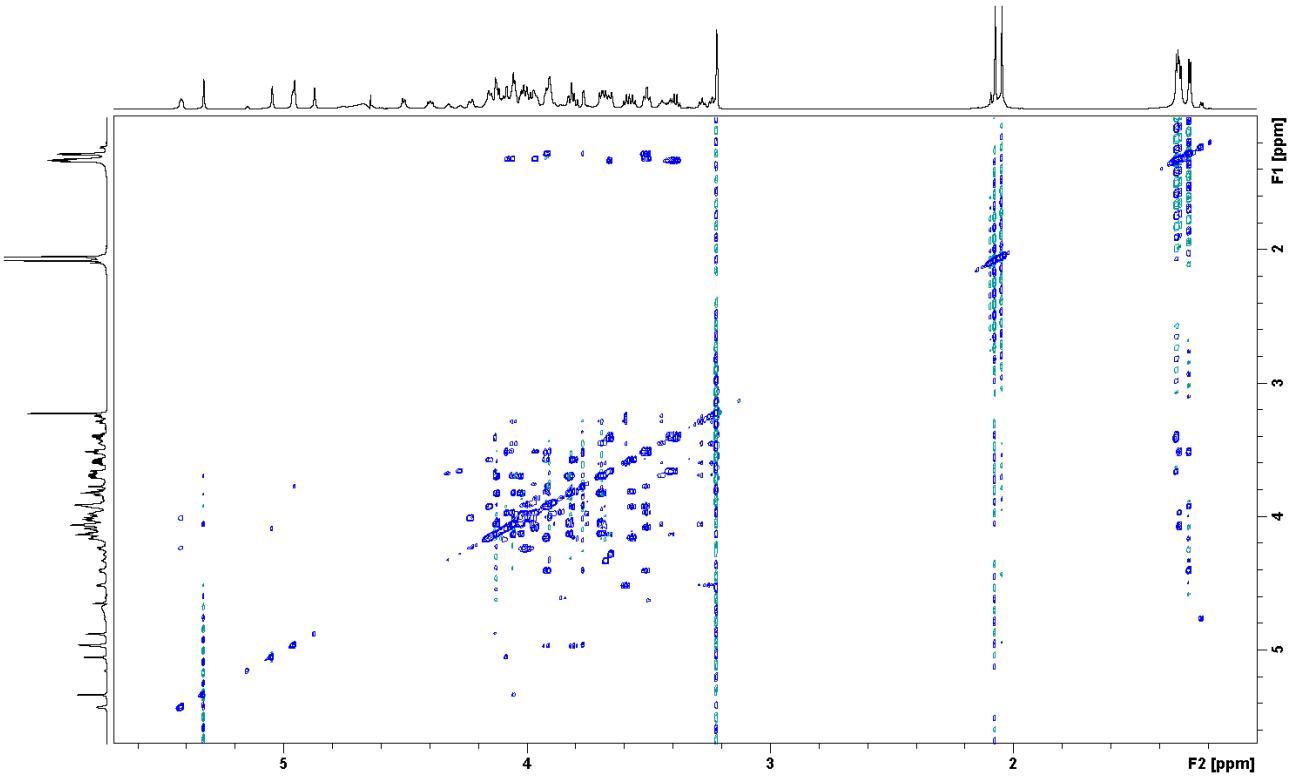


Figure A.11: TOCSY spectrum of the 7C sample. The projections are from separate experiments.

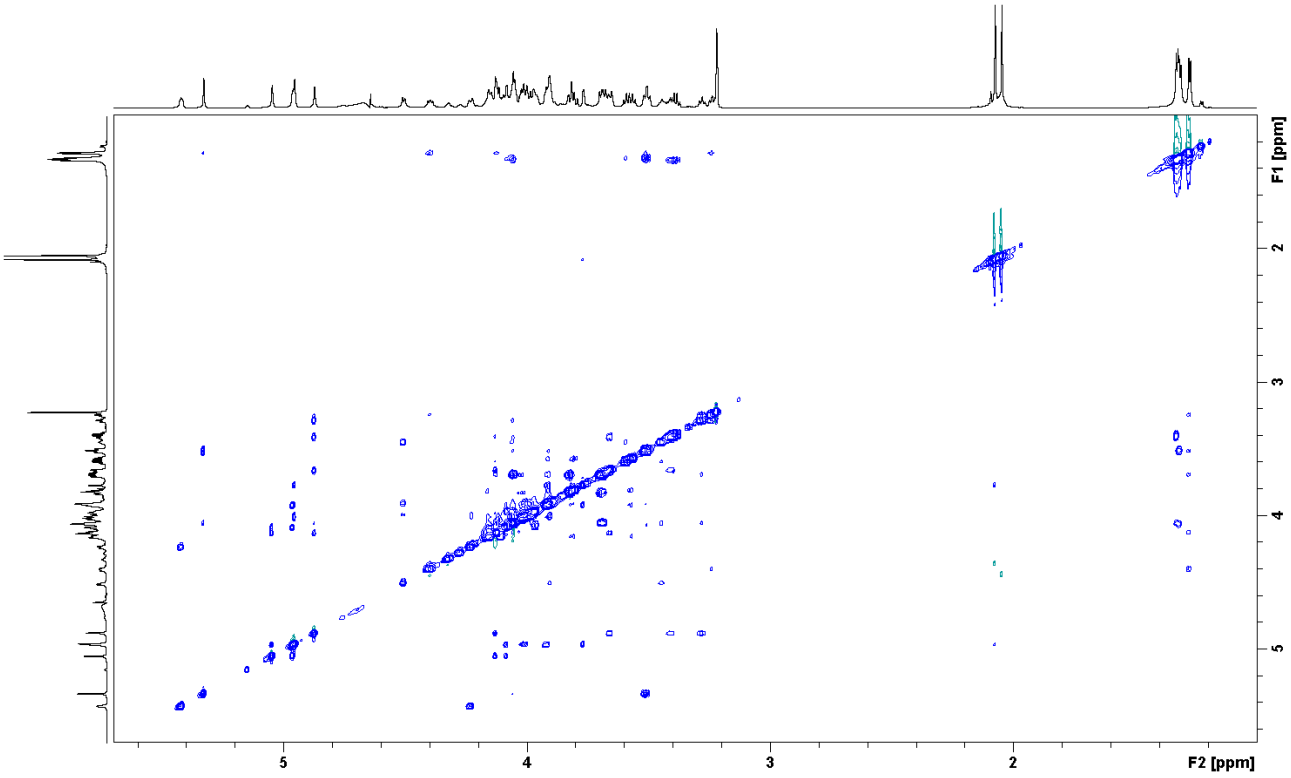


Figure A.12: NOESY spectrum of the 7C sample. The projections are from separate experiments.

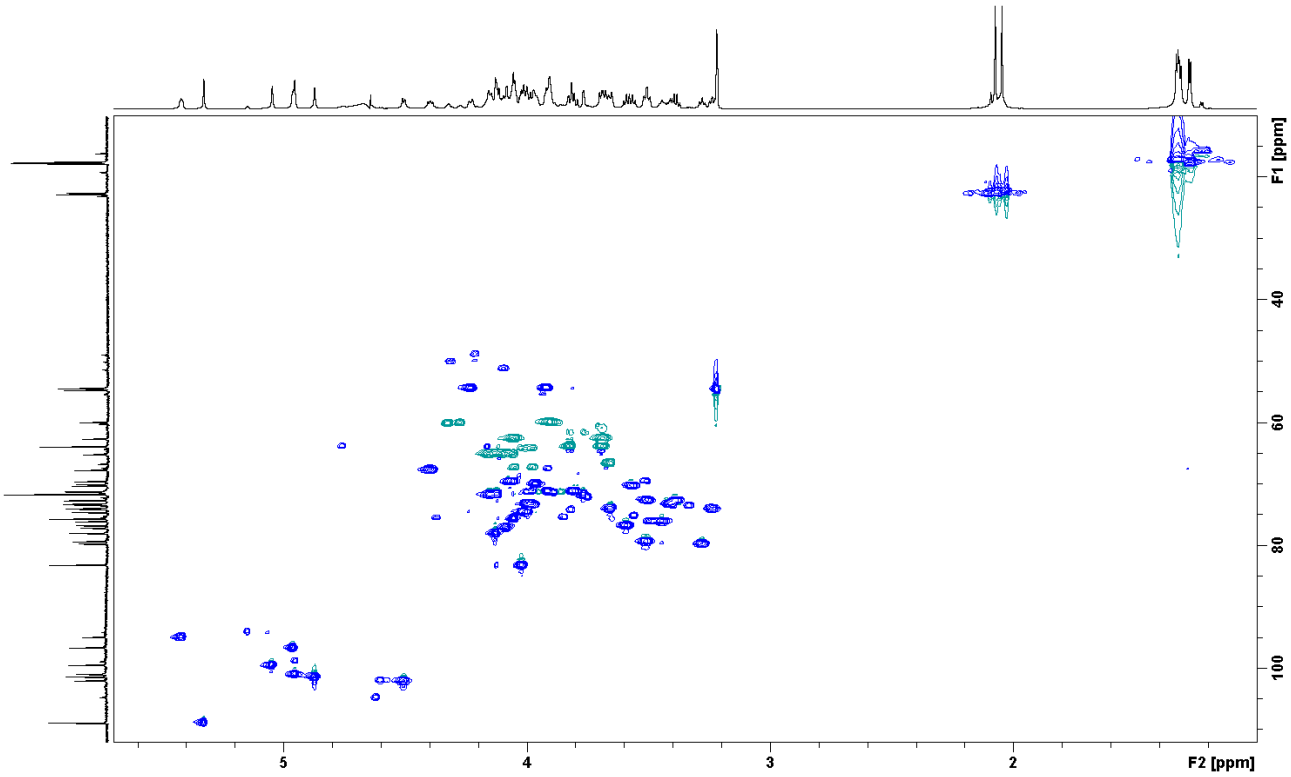


Figure A.13: HSQC spectrum of the 7C sample. The projections are from separate experiments.

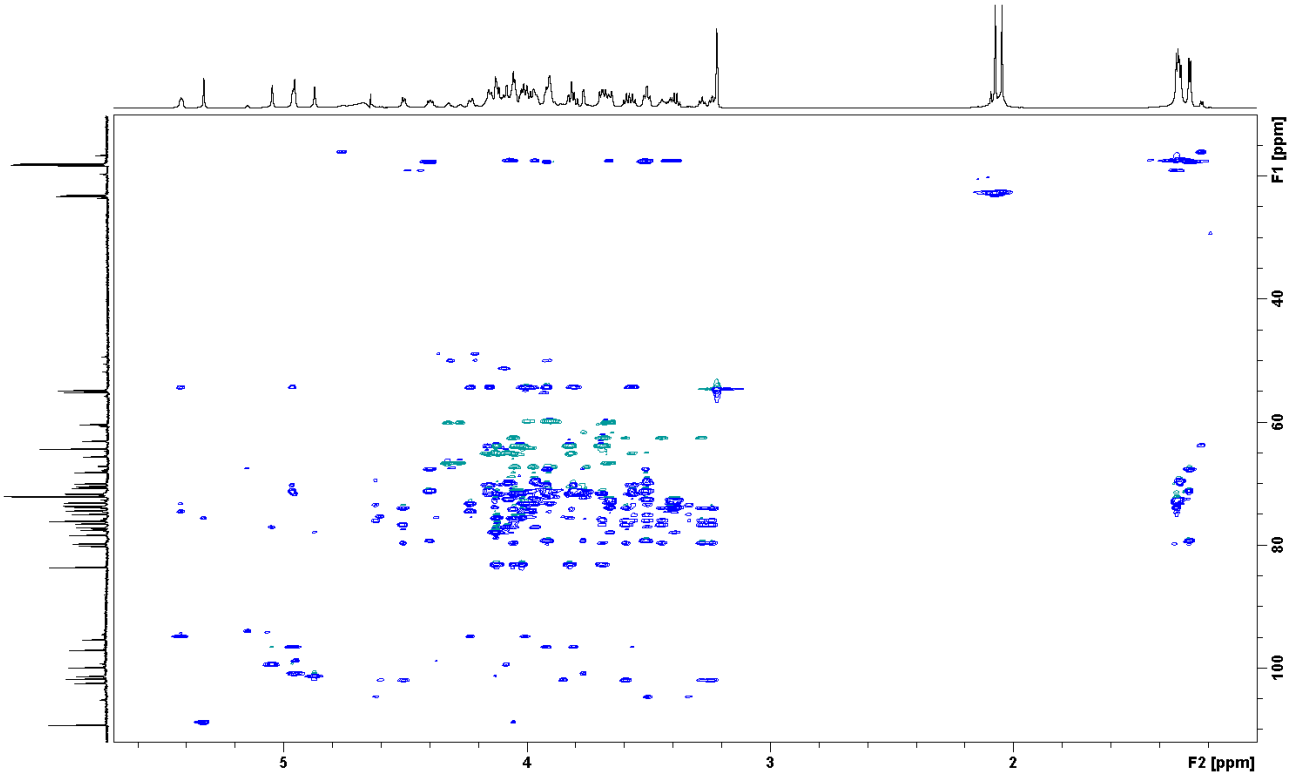


Figure A.14: HSQC-TOCSY spectrum of the 7C sample. The projections are from separate experiments.

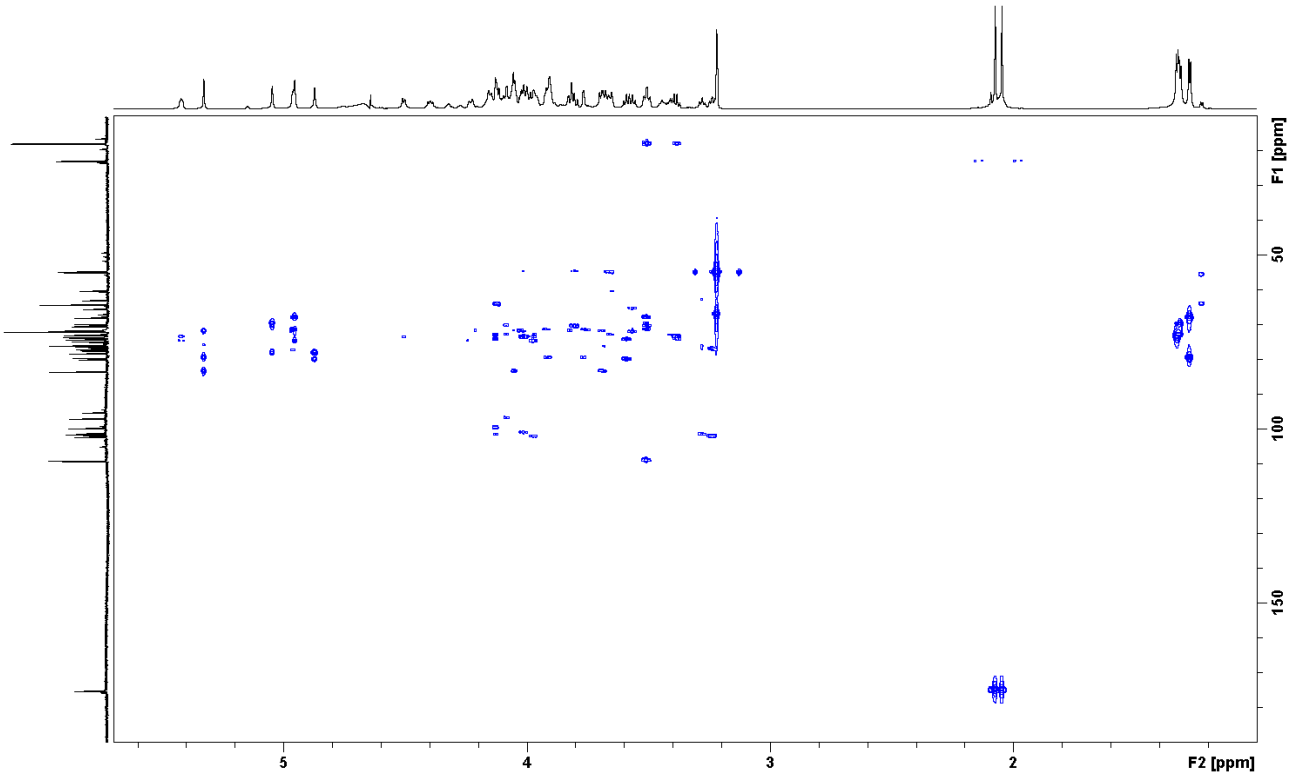


Figure A.15: HMBC spectrum of the 7C sample. The projections are from separate experiments.

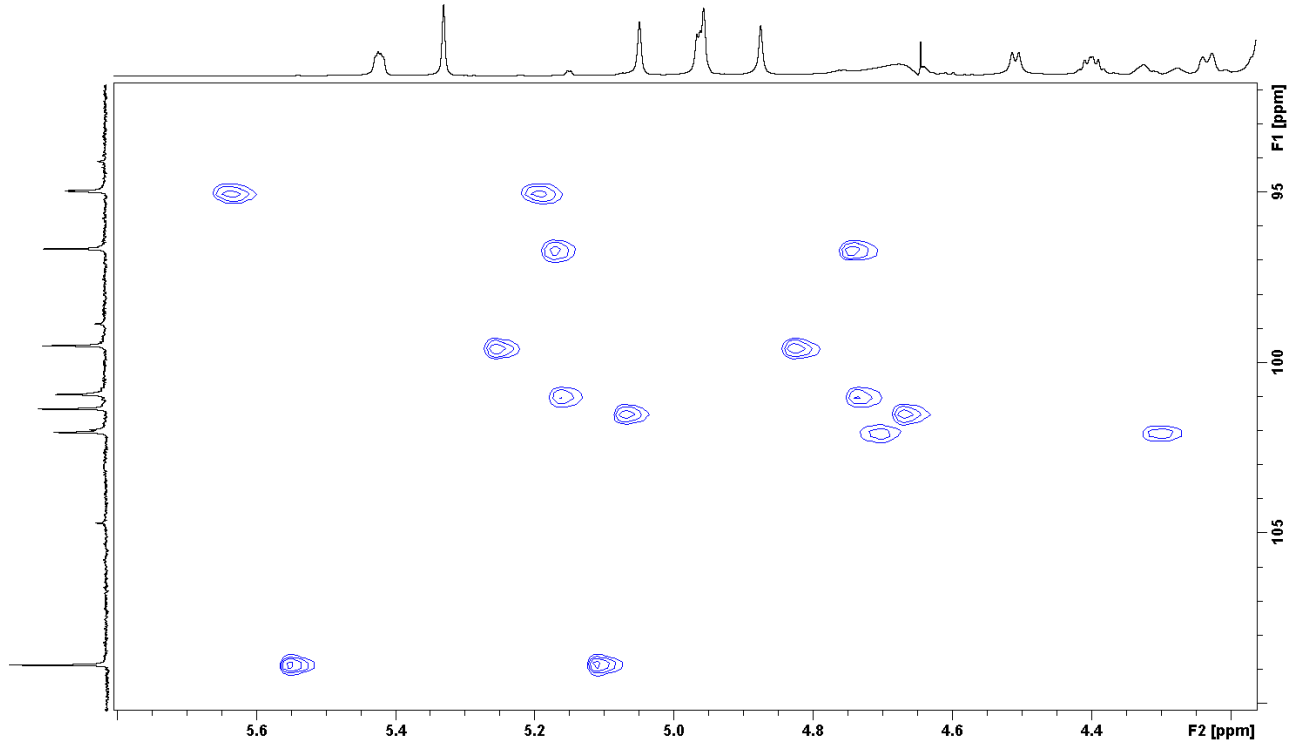


Figure A.16: CLIP-HSQC spectrum of the 7C sample. The projections are from separate experiments.

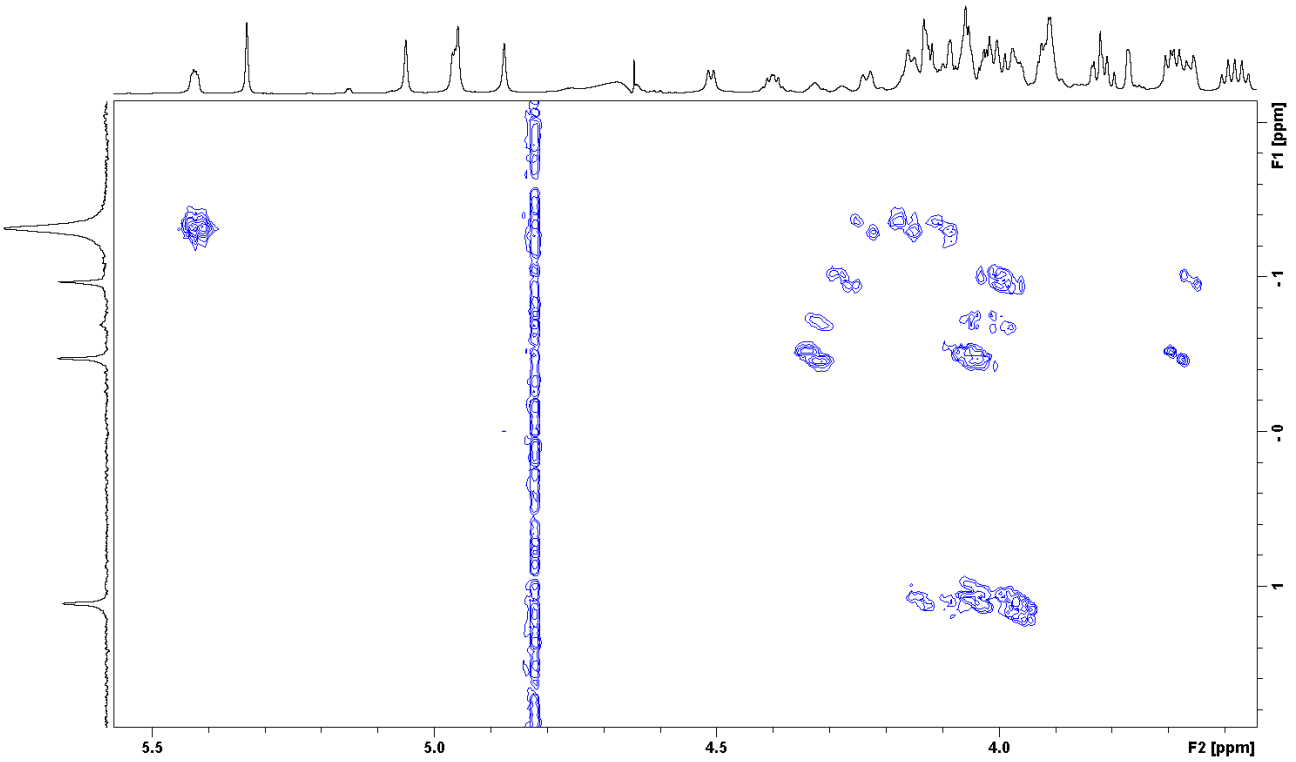


Figure A.17: $^{1}H,^{31}P$ HMBC spectrum of sample 7C. 31P is not calibrated. Projections are from separate experiments.

7C coupling constants

Table A.1: All coupling constants for the 7C cps. N/A indicates when part of the table does not apply and n.d. indicates where it was not possible to determine, usually due to spectral overlap. All $^{1}J_{H,C}$ coupling constants were determined using CLIP-HSQC and all $^{3}J_{H,H}$ coupling constants were determined using DQF-COSY.

7C couplings furanoside	H1-C1	H1-H2	H2-H3	H3-H4	H4-H5	H4-H5’	H5-H5’	N/A
β-Ribf-(1- A	177 Hz	3.1 Hz	5.2 Hz	7.9 Hz	7.4 Hz	3.8 Hz	12.2 Hz	N/A
7C coupling pyranosides	H1-C1	H1-H2	H2-H3	H3-H4	H4-H5	H5-H6	H5-H6’	H6-H6’
-4)-α-Rhap-(1- D	171 Hz	6.3 Hz	4.3 Hz	10.0 Hz	10.1 Hz	6.5 Hz	N/A	N/A
-P-O-6)-α-GlcpNAc-(1- F	171 Hz	4.3 Hz	11.2 Hz	8.3 Hz	10.1 Hz	n.d.	n.d	n.d
-2)-α-Rhap-(1- E	170 Hz	6.2 Hz	4.4 Hz	10.5 Hz	10.2 Hz	6.5 Hz	N/A	N/A
-2)-β-Rhap-(1- C	160 Hz	6.2 Hz	4.1 Hz	9.6 Hz	n.d.	5.9 Hz	N/A	N/A
-4)-β-Glcp-(1- B	162 Hz	7.7 Hz	11.1 Hz	7.5 Hz	12.0 Hz	7.8 Hz	5.0 Hz	11.9 Hz
-3,4)-α-GalpNAc-(1-O-P- G	178 Hz	2.7 Hz	10.2 Hz	n.d.	n.d.	6.7 Hz	N/A	N/A

7C Anomeric HMBC and NOE correlations

Table A.2: Summary of strong anomeric correlations from HMBC and NOESY spectra

Anomeric position	Strong HMBC	Strong NOE
β-Ribf **A**	A2, A3, A4, D4	D4
α-Rhap **D**	D2/D3, D5, G3	G3
α-GlcpNAc **F**	F3, F5, E2	E1, E2
α-Rhap **E**	E3, E5, C2	C2, F1
β-Rhap **C**	C2, B4	C3, C5, B4
β-Glcp **B**	G4	B5, G4, G5
α-GalpNAc **G**	G3, G5	none

7B HSQC spectrum

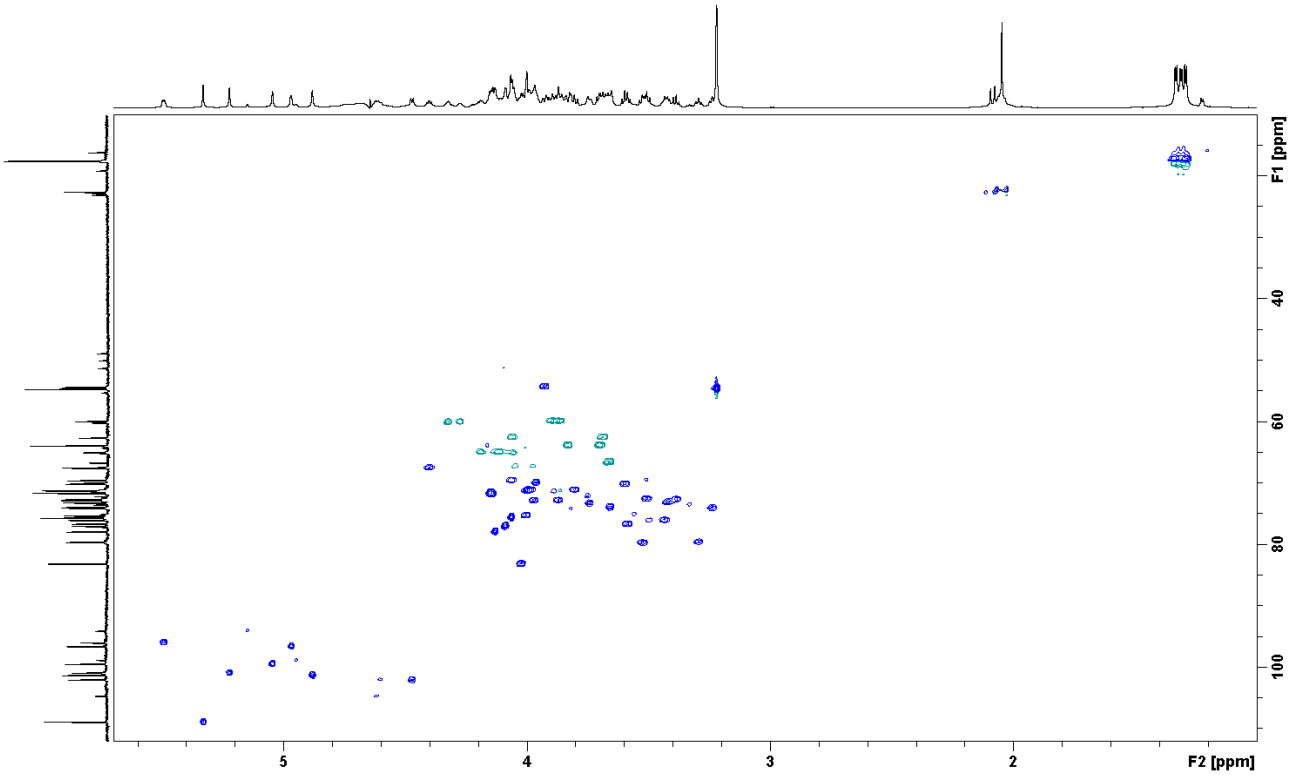


Figure A.18: HSQC spectrum of the 7B sample. The projections are from separate experiments.

7A and 7F HSQC spectrum

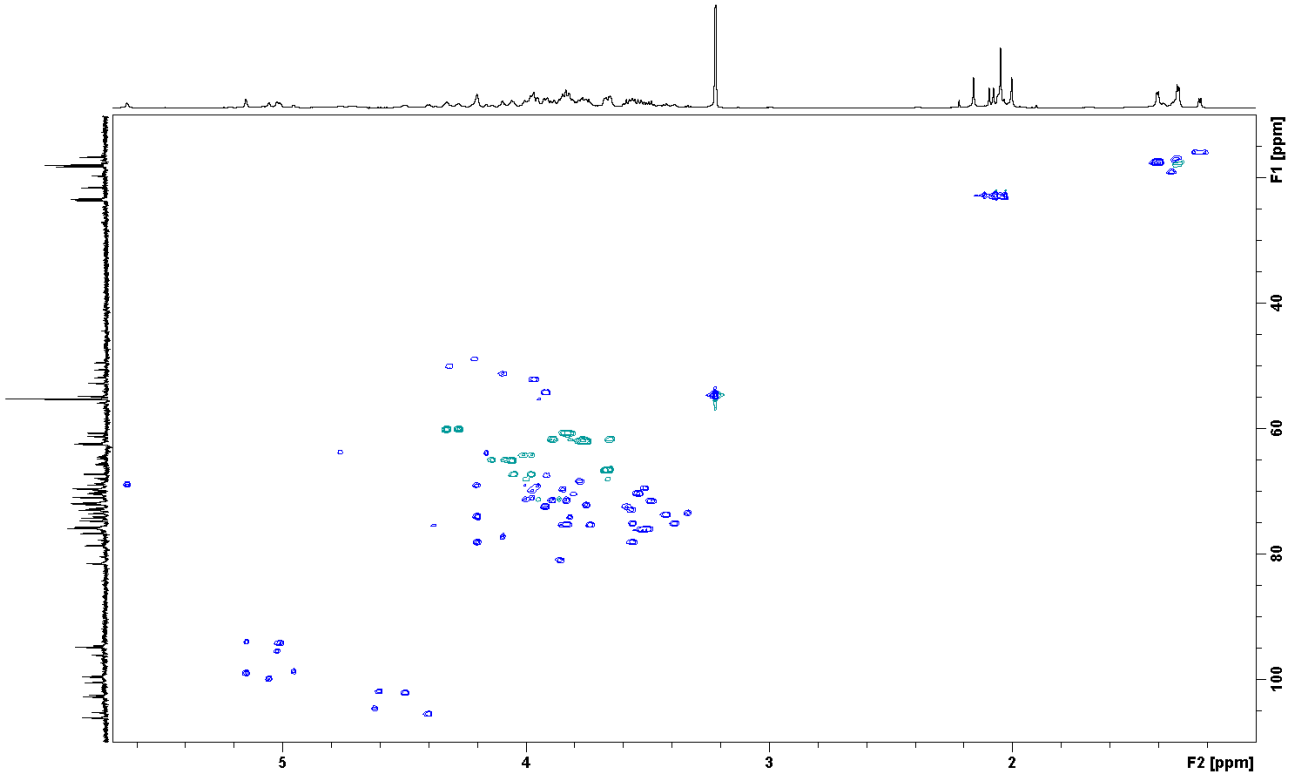


Figure A.19: HSQC spectrum of the 7A sample. The projections are from separate experiments.

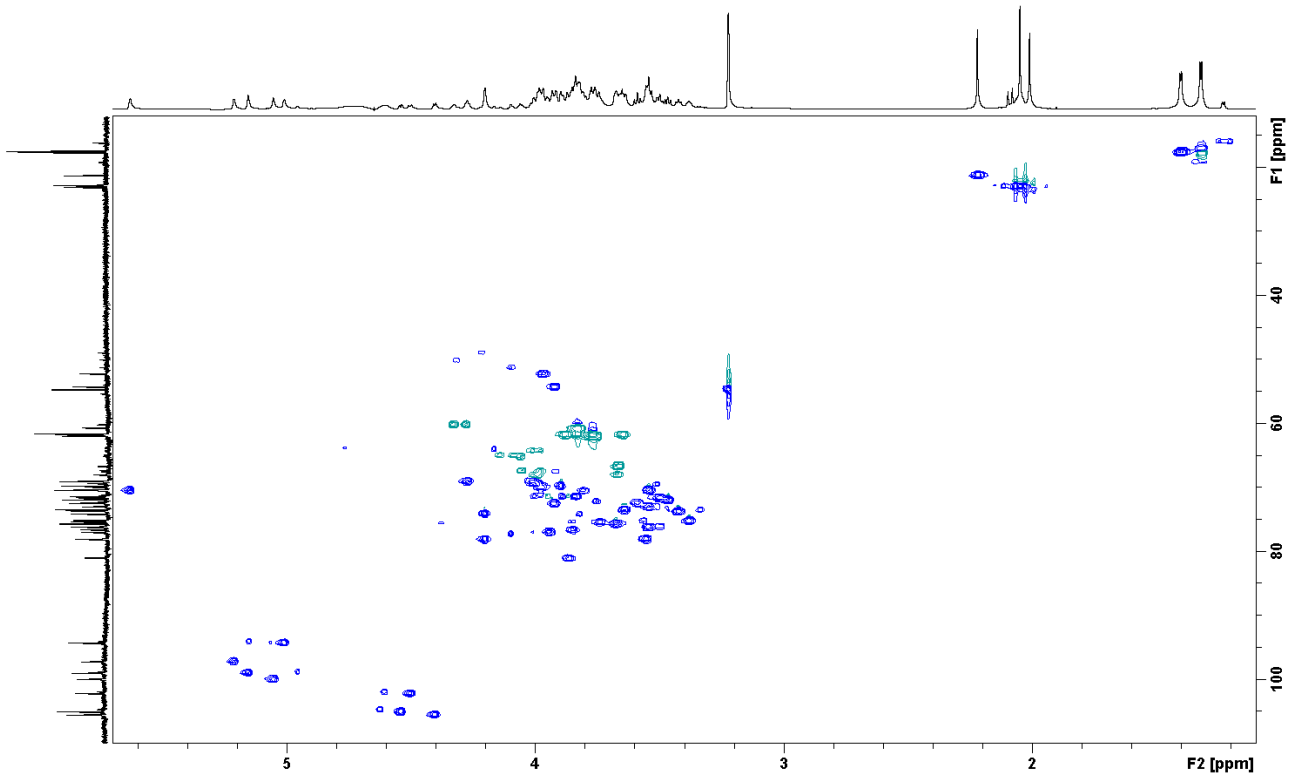


Figure A.20: HSQC spectrum of the 7F sample. The projections are from separate experiments.

7A and 7F tables

Table A.3: Table containing the assignment of 7A.

7A table	1	2	3	4	5	6	Ac-Me	C=O
α-GlcpNAc-(1-	5.009 94.30	3.918 54.30	3.835 71.33	3.538 70.35	3.921 72.50	3.826 60.70	2.049 22.86	N/A 174.46
-2)-α-Rhap-(1-	5.149 99.03	4.199 74.14	3.973 69.99	3.587 72.35	3.807 70.41	1.324 17.46	N/A	N/A
-3)-β-GalpNAc-(1-	4.497 102.20	3.967 52.20	3.859 80.99	4.199 78.18	3.735 75.38	3.763 61.96	2.078 23.10	N/A 175.40
-4)-β-Glcp-(1-	4.400 105.56	3.423 73.72	3.520 76.17	3.561 78.12	3.390 75.12	3.652/3.891 61.77	N/A	N/A
-3)-β-Rhap2Ac-(1-	5.055 99.98	5.637 69.04	3.826 75.26	3.485 71.57	3.568 72.99	1.404 17.70	2.159 22.82	N/A 174.21
-6)-α-Galp-(1-	5.022 95.57	3.781 68.46	3.850 69.79	3.952 69.14	4.199 74.14	3.663/3.997 68.18	N/A	N/A

Table A.4: Table containing the assignment of 7F.

7F table	1	2	3	4	5	6	Ac-Me	C=O
α-GlcpNAc-(1-	5.008 94.32	3.925 54.29	3.835 71.44	3.54 70.48	3.926 72.51	3.834 60.73	2.047 22.86	N/A 174.47
-2)-α-Rhap-(1-	5.153 99.03	4.200 74.14	3.988 69.42	3.588 72.43	3.805 70.49	1.318 17.48	N/A	N/A
β-Galp-(1-	4.541 105.04	3.464 71.95	3.643 73.54	3.895 69.8	3.673 75.7	3.772/3.830 61.81	N/A	N/A
-3)-β-GalpNAc-(1-	4.503 102.19	3.964 52.27	3.866 81.02	4.200 78.09	3.737 75.39	3.77 61.75	2.066 23.08	N/A 175.37
-4)-β-Glcp-(1-	4.403 105.56	3.421 73.71	3.547 76.19	3.551 78.06	3.38 75.16	3.828 61.68	N/A	N/A
-3)-β-Rhap2Ac-(1-	5.051 99.99	5.628 70.45	3.847 76.62	3.496 71.65	3.541 73.03	1.401 17.71	2.22 21.19	N/A 174.21
-2,6)-α-Galp-(1-	5.211 97.24	3.943 76.98	4.005 69.12	3.99 69.75	4.271 69.07	3.669/3.989 68	N/A	N/A

Paper 2

Structural characterization of bioactive heteropolysaccharides from the medicinal fungus *Inonotus obliquus* (Chaga)

Wold, C. W.; **Kjeldsen, C.**; Corthay, A.; Rise, F.; Christensen, B. E.; Duus, J. Ø.; Inngjerdingen, K. T.

Manuscript submitted to Carbohydrate Polymers

Structural characterization of bioactive heteropolysaccharides from the medicinal fungus *Inonotus obliquus* (Chaga)

Christian Winther Wold[a*], Christian Kjeldsen[b], Alexandre Corthay[c], Frode Rise[e], Bjørn E. Christensen[d], Jens Øllgaard Duus[b], Kari Tvete Inngjerdingen[a]

[a] School of Pharmacy, University of Oslo, P.O. Box 1068 Blindern, N-0316 Oslo, Norway
b Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[c]Tumor Immunology Lab, Department of Pathology, Rikshospitalet, Oslo University Hospital, P.O. Box 4950 Nydalen, NO-0424 Oslo, Norway
[d] NOBIPOL, Department of Biotechnology and Food Science, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[e]Department of Chemistry, University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway
* Corresponding author

Abstract

The aim of this paper was to perform a comprehensive characterization of polysaccharides isolated from the interior (IOI) and exterior (IOE) parts of the fungus *Inonotus obliquus.* Pre-extraction with DCM and MeOH, followed by water and alkali extraction and ethanol precipitation gave two water extracts and two alkali extracts. Neutral and acidic polysaccharide fractions were obtained after anion-exchange chromatography of the water extracts. The neutral polysaccharides (60-73 kDa) were heterogeneous and branched and consisted of a (1→3)-linked β-Glc backbone with (1→6)-linked kinks in the chain at approximately every fifth residue, with branches of (1→6)-linked β-Glc in addition to substantial amounts of (1→6)-linked α-Gal with 3-*O*-methylation at about every third Gal residue. The acidic polysaccharide fractions (10-31 kDa) showed similar structural motifs as the neutral fractions differing mainly by the presence of (1→4)-linked α-GalA and α-GlcA. β-Xyl, α-Man and α-Rha were also present in varying amounts in all fractions. No major structural differences between the IOI and IOE fractions were observed. An alkaline polysaccharide fraction (>450 kDa) was obtained from the IOI alkali extract, and consisted mainly of (1→3)- and (1→6)-linked β-Glc and (1→4)-linked β-Xyl. Several of the fractions showed *in vitro* immunomodulatory effect by increasing NO production in the murine macrophage and dendritic cell lines J774.A1 and D2SC/1. Most fractions managed to increase NO production only at the highest concentration tested (100 µg/ml), while the neutral fraction IOE-WN activated potent NO production at 10 µg/ml and was considered the most promising immunomodulating fraction in this study.

1. Introduction

Inonotus obliquus (Fr.) Pilát, a rather unusual polypore fungus found in northern latitudes, has been used as an infusion in traditional medicine in Northern Europe and Russia since the 16th century against a wide range of illnesses such as tuberculosis, heart disease, cancer and stomach issues (Saar, 1991; Shikov et al., 2014). *I. obliquus* infects living, broad-leaved trees from the Betulaceae family, causing white heart rot in the tree, growing for decades while producing massive, sterile conks of sclerotia without forming fruiting bodies (Cha, Lee, Lee, & Chun, 2011; Min-Woong et al., 2008). Scientific studies have revealed many biological properties of extracts and isolated substances from *I. obliquus*, such as anti-inflammatory, anti-tumor, antioxidative and immunomodulatory properties (Chen, Huang, Cui, & Liu, 2015; Glamoclija et al., 2015; Hu et al., 2016; Lee, Lee, Song, Ha, & Hong, 2014; Ma, Chen, Dong, & Lu, 2013). Structural analysis of compounds from this fungus has revealed the presence of triterpenoids, polyphenols, melanin pigments and polysaccharides which are likely to be, at least in part, responsible for its bioactive effects (Chen et al., 2015; Hwang, Lee, & Yun, 2016; Kim et al., 2005; Youn et al., 2009; Zhao et al., 2015; Zheng et al., 2010).

Fungal polysaccharides, for example (1→3/1→6)-β-D-glucans such as lentinan from *Lentinula edodes* (Shiitake), have been associated with immunomodulatory and anti-tumor effects (Zhang, Cui, Cheung, & Wang, 2007). In Japan, lentinan is used in human cancer therapy in combination with conventional treatments such as chemotherapy to improve wellbeing and treatment outcome for patients (Ina, Kataoka, & Ando, 2013; Kapoor, 2014). Various mechanisms of action have been proposed for the anti-tumor effects of such fungal polysaccharides, with both pro- and anti-inflammatory effects being reported in the literature (Du, Lin, Bian, & Xu, 2015; Schwartz & Hadar, 2014; Zhang et al., 2007). A central mechanism of action of these compounds is their interaction with pattern recognition receptors (PRRs) on innate immune cells such as macrophages and dendritic cells (Erwig & Gow, 2016). As such, the immunological effects of the polysaccharides will depend greatly on the specific receptors involved, since different PRRs could induce different downstream effects in the cell when activated (Fraser, Stuart, & Ezekowitz, 2004). In this regard the structure and shape of the polysaccharide, including monosaccharide composition, molecular weight (M_w), branching and anomeric configuration, are important for its activity and could explain some of the conflicting reports in the scientific literature. Further, many alleged polysaccharide extracts from medicinal fungi are shown to contain high levels of polyphenols and proteins (Wei & Van Griensven, 2008). Since some fungal compounds can activate immune cells while others can suppress the same cells a proper purification of the extracts and isolation of pure compounds are highly important to avoid additional scientific confusion (Lu et al., 2016; Wei & Van Griensven, 2008).

Although there are several reports of polysaccharides from *I. obliquus* with immunomodulatory and anti-tumor activity (Chen et al., 2015; Fan, Ding, Ai, & Deng, 2012; Kim et al., 2006; Lee et

al., 2014; Rhee, Cho, Kim, Cha, & Park, 2008), satisfactory characterization of the polysaccharides including anomeric configuration and linkage analysis is lacking, and the purification methods used are often questionable. To our knowledge, the present study is the first extensive characterization of *I. obliquus* immunomodulating polysaccharides.

2. Methods

2.1 Preparation of fungal material

A fresh *Inonotus obliquus (family: Hymenochaetaceae)* specimen was harvested from a birch tree in Oslo, Norway and verified by Prof. Klaus Høiland (Dept. of Biosciences, UiO, Norway). A voucher specimen was deposited in the Pharmacognosy section, School of Pharmacy, University of Oslo, Norway. The fungal material was separated into two parts distinguished by the difference in texture and color - a brown interior part and a black exterior part – before cut into small pieces and lyophilized. The separate parts were then ground to a finely dispersed powder using an industrial blender (RAW® X1500).

2.2 Extraction of polysaccharides

In order to remove low molecular weight compounds, the fungal material (weight: 227g/131g interior/exterior) was pre-extracted in a soxhlet extractor, first with dichloromethane (DCM) for 48 h followed by methanol (MeOH) for 48 h. The remaining dried residue was then extracted first with distilled water (dH_2O, 100 °C, 3x, 2h) and subsequently with boiling sodium hydroxide plus sodium borodeuteride under reflux (1 M NaOH, 0.135 M $NaBD_4$, 2x, 4 h). After each extraction, the supernatant was collected after centrifugation (4000 rpm, 20 min) and concentrated under reduced pressure on a rotavapor. The extracts were then treated with pancreatin (3h, 37 °C, Sigma-Aldrich®) to degrade proteins, before ethanol precipitation (70%, 48 h, 4 °C) and centrifugation. The precipitates were finally dissolved in dH_2O, dialyzed (cut-off 3500 Da, 72 h, 4 °C) and lyophilized to obtain four crude polysaccharide extracts – two water-soluble extracts termed IOI-W and IOE-W (***I****nonotus* ***o****bliquus* **I**nterior/**E**xterior **W**ater-extracted) and two alkali-soluble extracts termed IOI-A and IOE-A (***I****nonotus* ***o****bliquus* **I**nterior/**E**xterior **A**lkali-extracted).

2.3 Fractionation of polysaccharides

The extraction and fractionation scheme used to obtain the purified polysaccharide fractions is presented in Fig. 1. All column chromatography experiments were carried out using an Äkta FPLC system (Pharmacia Äkta, Amersham Pharmacia Biotech, Uppsala, Sweden) with a fraction collector. All extracts and fractions were filtered through 0.45 μm Millipore filters before loaded

onto the columns. The Unicorn 4.0 software (GE Healthcare, Uppsala, Sweden) was used to set-up and monitor the experiments.

2.3.1 Anion-exchange chromatography

The water-soluble crude extracts IOI-W and IOE-W (10 mg/ml) were applied to a column packed with ANX Sepharose™ 4 Fast Flow (high sub) (GE Healthcare). Neutral fractions were obtained by eluting with dH_2O (2 ml/min), while acidic fractions were obtained using a linear NaCl gradient (0–1.5 M, 2 ml/min). Fractions were collected (15 ml/tube) and carbohydrate elution profiles were monitored using the phenol-sulfuric acid method (Dubois, Gilles, Hamilton, Rebers, & Smith, 1956). The related fractions were pooled, dialyzed against dH_2O (cut-off 3500 Da) and lyophilized.

2.3.2 Size exclusion chromatography

Several fractions were subjected to gel filtration/size exclusion column chromatography (SEC) using different SEC columns. The acidic fraction from IOI-W, termed IOI-WAc (20 mg/ml), was fractionated by Hiload™ 16/60 Superdex™ 200 prep grade column (GE Healthcare), eluting with 10 mM NaCl at 0.5 ml/min and collected (2 ml/tube). The neutral fractions IOI-WN and IOE-WN (2 mg/ml) were subjected to SEC using a Superose® 6 prep grade column (GE Healthcare), eluting with 10 mM NaCl at 0.3 ml/min and collected (0.5 ml/tube).
The alkali-soluble crude extracts IOI-A and IOE-A (20 mg/ml) were subjected to SEC using a Sephacryl S-500 HR column (GE Healthcare), eluting with 0.2 M NaOH at 2 ml/min and collected (15 ml/tube). Ultimately, carbohydrate elution profiles were monitored after each experiment using the phenol-sulfuric acid method and related fractions were pooled, dialyzed and lyophilized.

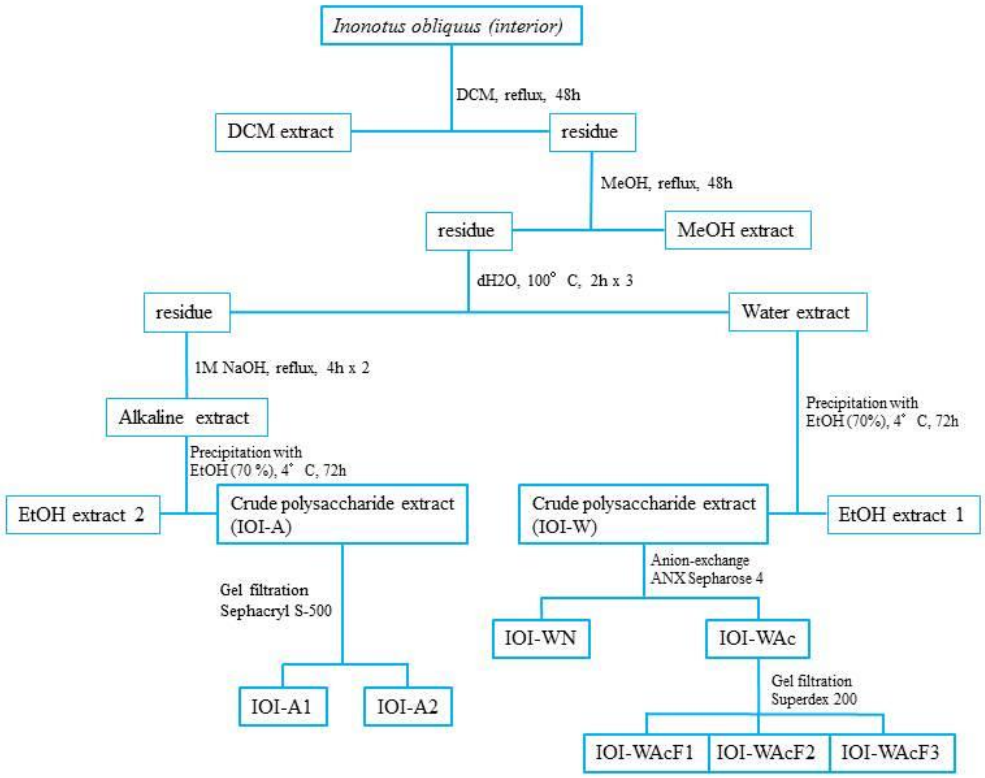


Fig. 1: Flowchart for extraction and fractionation of polysaccharides from the interior part of *Inonotus obliquus*.

2.4 Molecular weight determination by SEC-MALLS

Molecular weight distributions and averages, notably the average molecular weights (M_w) of the polysaccharides, were determined as previously described (Austarheim et al., 2012). Briefly, two SEC columns were coupled in series, coupled to a SIL-10AF auto sampler, a LC-10ADvp pump and a SCL-10Avp system controller (Shimadzu, Japan). Three online detectors were used, a multi-angle laser light scattering (MALLS) Dawin Heleos-II (Wyatt, USA), a refractive index detector (Optilab DSP, Wyatt) and a viscometer (ViscoStar, Wyatt). The column was eluted with aqueous 0.05 M Na_2SO_4/0.01 M EDTA, pH 6 at a flow rate of 0.5 mL/min. Pullulan P8 (8 kDa) was used for normalization of all the MALLS detectors. 100–150 µl sample solution (0.4–1.5 mg/mL) was injected. All samples were dissolved in distilled water 24 h prior to analysis, and diluted with concentrated eluent to reach eluent concentration before injection. The samples were run in duplicates, and M_w for each sample was calculated using the ASTRA V software (Wyatt). A refractive index increment $(dn/dc)_\mu$ of 0.15 ml/g was used in the calculations.

2.5 Determination of proteins, phenolics and lipopolysaccharide in the polysaccharide fractions

The protein content of the purified polysaccharide fractions (1 mg) was determined by the Bio-Rad protein assay (Bio-Rad), based on the method of Bradford (Bradford, 1976). The standard procedure for microtiter plates was used with bovine serum albumin (BSA, Sigma) as a protein standard (30-500 µg/ml). The experiment was carried out in triplicates.

The amount of phenolic compounds in the purified polysaccharide fractions was determined using the Folin-Ciocalteu assay (Swain & Hillis, 1959). Briefly, the samples were dissolved in dH_2O (1 mg/ml) before adding the same amount of Folin-Ciocalteu's phenol reagent (1:2 in dH_2O, Merck/Kebo), mixed and left for 3 min at room temperature. 1 M Na_2CO_3 was added, before the tubes were mixed and allowed to stand for 1 h. The absorbance was measured at 750 nm. A standard curve was plotted using ferulic acid (0-50 µg/ml). The total phenolic content was determined as ferulic acid equivalents. The experiment was carried out in triplicates.

The amount of lipopolysaccharide (LPS) contamination in the purified fractions was determined by a previously described method (de Santana-Filho et al., 2012), with some modifications. Briefly, the lyophilized polysaccharide samples (5 mg) were dissolved in 3 M HCl in MeOH and incubated for 20 h at 80 °C. The solution was then partitioned between hexane (1 ml) and dH_2O (0.5 ml) three times. The hexane phase was collected, pooled and dried under a gentle N_2 stream. The samples were then acetylated at 100 °C for 1 h using a 1:1 mixture of pyridine and acetic anhydride, and thoroughly dried. Finally, acetone (70 µl) was added and the samples were analyzed by GC–MS, using the GC–MS solution software, Version 2.10 (Shimadzu Corporation). The detection limit of this assay was ~2 ng/µl.

2.6 Monosaccharide composition

The monosaccharide composition of the crude extracts and purified fractions was determined using a method originally described by Chambers and Clamp with some modifications (Chambers & Clamp, 1971; Nyman, Aachmann, Rise, Ballance, & Samuelsen, 2016). In short, the samples were subjected to methanolysis with 3 M hydrochloric acid (HCl) in anhydrous methanol (MeOH) for 24 h at 80 °C. To analyze for the presence of amino sugars, the samples were then dried using nitrogen before pyridine and acetic acid anhydride were added for 3 minutes. After drying, 0.5 M HCl in MeOH was added and the samples were incubated for 1 h at 65 °C. Trimethylsilylated (TMS) derivatives of the methyl glycosides obtained after methanolysis were analyzed by capillary gas chromatography on a Trace™ 1300 GC (Thermo Scientific™). Mannitol was used as an internal standard.

205 *2.7 Linkage analysis by methylation/GC-MS*
206
207 Glycosidic linkage analysis was performed by methylation and subsequent gas chromatography-
208 Mass spectrometry (GC-MS) based on previously described methods (I. & Kerek, 1984;
209 Pettolino, Walsh, Fincher, & Bacic, 2012). Briefly, polysaccharide samples (~1 mg) were first
210 reduced using $NaBD_4$ to convert uronic acids to their corresponding neutral sugars. The
211 polymers were then methylated with methyl iodide after dissolving in dimethylsulfoxide (DMSO)
212 and NaOH. The per-O-methylated polysaccharides were hydrolyzed with triflouroacetic acid
213 (TFA, 2.5 M, 4 h) and reduced with $NaBD_4$. The O-methylated alditols were acetylated by
214 adding 1-methylimidazole and acetic anhydride and partially O-methylated alditol acetates were
215 finally extracted with DCM and analyzed on a GCMS-QP2010 (Shimadzu Corporation) with a
216 Restek Rxi-5MS silica column (30 m, i.d. 0.25 mm, 0.25 μm film thickness), split injection and
217 set at a constant pressure mode. Initial flow was 1 ml/min. The injector and interface
218 temperatures were 280°C. At the time of injection, the column temperature was 80°C, then, after
219 5 min, the temperature was increased with 10°C/min up to 140°C followed by 4°C/min to 210°C
220 and then 20°C/min to 310°C, at which it was kept for 4 min. Ion source temperature was 200°C.
221 Helium was used as the carrier gas. Spectra were analyzed using GC–MS solution software,
222 Version 2.10 (Shimadzu Corporation).
223
224 *2.8 Characterization of a (1,6)-α-3-O-methyl-galactose*
225
226 *2.8.1 Demethylation of monosaccharides using boron tribromide*
227
228 The GC chromatograms from the monosaccharide composition analysis revealed three large
229 peaks with unknown identity in some of the samples, which were speculated to come from a
230 methylated hexose. To characterize this unknown compound, a method described in *Methods in*
231 *Carbohydrate Chemistry, Vol II* (Whistler & Wolfrom, 1963) was used. Briefly, IOE-WN was
232 methanolyzed as described in section 2.6, before the residue was dissolved in DCM and cooled
233 to -80°C using a mixture of acetone and dry ice. Boron tribromide (BBr_3, -80°C) was then
234 carefully added to the sample, and left at this temperature for 30 min to induce demethylation.
235 The samples were then left at room temperature overnight with a drying tube with $CaCl_2$, before
236 washed twice with MeOH, dried and analyzed by GC as described in section 2.6. Finally, the
237 chromatogram from the demethylated sample was compared with the untreated counterpart. The
238 standard curve from galactose was used for quantification.
239
240 *2.8.2 Linkage analysis by ethylation*
241
242 To establish the type of linkages and position of the methyl group, IOE-WN was subjected to
243 ethylation and GC-MS analysis, as described in section 2.7 with one modification; ethyl iodide

was used instead of methyl iodide, which would produce unique fragmentation patterns in the GC-MS spectra.

2.9 IR spectroscopy

Infrared (IR) spectra of the purified polysaccharide fractions were obtained on a Nicolet™ FT-IR Spectrometer (Thermo Fischer Scientific). In short, dried samples (~2mm^3) were pressed onto an attenuated total reflectance (ATR) crystal using a high pressure tower to provide consistent results. The results were analyzed using the OMNIC8 software.

2.10 Nuclear magnetic resonance (NMR) spectroscopy

The water-extracted polysaccharide fractions (5-10 mg) were dissolved in 500 µl D_2O (99.9 %, Sigma). All experiments were carried out on a Bruker Avance III (799.90 MHz for 1H and 201.14 MHz for ^{13}C) equipped with a 5 mm TCI ^{1}H/(^{13}C, ^{15}N) cryoprobe using acetone as reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively). The chemical shifts were assigned using the following spectra: 1D ^{1}H with presaturation, 1D ^{13}C, 2D double quantum filter correlated spectroscopy (DQF-COSY), 2D total correlation spectroscopy (TOCSY) with 60 ms mixing time, 2D nuclear Overhauser effect spectroscopy (NOESY) with 200 ms mixing time, 2D ^{13}C heteronuclear single quantum coherence (HSQC) with multiplicity editing, 2D ^{13}C HSQC-[^{1}H,^{1}H] TOCSY (HSQC-TOSCY) with 60 ms mixing time, and 2D ^{13}C heteronuclear multi-bond correlation (HMBC) optimized for 10 Hz long range coupling constants. All spectra were recorded at 313K. Spectra were recorded using TopSpin 3.5 and processed and analyzed using the TopSpin 3.5 software (Bruker).

The alkali-extracted fraction IOI-A1 (10 mg) was dissolved in 0.1 M NaOH with D_2O (99.9%, Sigma). Experiments were carried out on a Bruker Avance III HD Ascend (800.03 MHz for ^{1}H and 201.17 MHz for ^{13}C) equipped with a 5 mm TCI ^{1}H/(^{13}C, ^{15}N) cryoprobe using the internal standard TMSP-$_{d4}$ 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid sodium salt as reference. The chemical shifts were assigned using the following spectra: 1D ^{1}H with presaturation and excitation sculpting solvent suppression, 1D ^{13}C, 2D DQF-COSY, 2D TOCSY with 80 ms mixing time, 2D NOESY with 300 ms mixing time and 2D ^{13}C HSQC with multiplicity editing. Spectra were recorded at 313 K and acquired, processed and analyzed using TopSpin 3.5.

2.11 Periodate oxidation and Smith degradation

Periodate oxidation and Smith degradation of was carried out as previously described (Kamerling & Gerwig, 2007). In short, IOE-WN and IOI-WAc (15 mg) were oxidized with 0.05 M sodium periodate (3 mL) and 0.2 M acetate buffer pH 4.0 (3 mL) at 4°C in the dark with stirring for 48 h. The oxidation was terminated by addition of 1% ethylene glycol and kept for 1

h before desalting using a PD-10 column (GE Healthcare). The samples were reduced with $NaBH_4$ (60 mg) and left at room temperature overnight. The reaction mixture was neutralized to pH 7.0 with 1 M acetic acid on ice and desalted again before hydrolysis with 0.05 M TFA for 30 min at 100°C. The samples were neutralized using 1 M NaOH and analyzed and fractionated using a CarboPac™ PA100 ZOU column (Dionex, ThermoFisher) with galactose, glucose, maltotetraose and maltohexaose as standards (Sigma). Finally the fractions were collected and subjected to methanolysis and GC analysis as described in section 2.6.

2.12 Cell lines

The murine macrophage cell line J774.A1 (a gift from Anders Gammelsrud, Norwegian Veterinary Institute) and the murine dendritic cell line D2SC/1 (a gift from Francesca Granucci, University of Milan) were grown in RPMI medium with 10 % FBS, 5 % streptomycin/penicillin, 1 % NEAA and 1 % mercaptoethanol, and incubated at 37 °C, 5% CO_2 . Prior to the experiments both cell lines were tested and found to be negative for Mycoplasma infection.

2.13 Nitric oxide assay

Nitric oxide (NO) production by activated immune cells was measured using the Griess reagent system (Promega). Briefly, J774.A1 and D2SC/1 cells were seeded at a density of 5×10^5 cells/ml in flat-bottomed 96-well plates to a total volume of 100 µl, and stimulated for 24 h with increasing concentrations of polysaccharide samples together with interferon-gamma (20 ng/ml, recombinant murine IFN-γ, Peprotech). After 24 h, the supernatant was collected and Griess reagents A and B were added to convert NO into nitrite (NO_2^-), which could be quantified colorimetrically using a dilution series of $NaNO_2$ as a standard curve. The absorbance was measured at 540 nm. LPS (1 µg/ml, from *E.coli 055:B5*, Sigma) and Pam3CSK4 (2 µg/ml, InvivoGen) with and without IFN-γ (20ng/ml), IFN-γ alone and untreated cells were used as controls. The nitrite value from the untreated negative control was set to zero and the value was subtracted from the other samples. The experiments were carried out in duplicates, and repeated two (D2SC/1) or three (J774.A1) times.

3. Results and discussion

3.1 Extraction and fractionation of polysaccharides

I. obliquus was separated into an interior and exterior part and extracted using a stepwise procedure with solvents of increasing polarity (Fig. 1). The extracts obtained by DCM and MeOH extraction were saved for later studies. Water- and alkali extraction, protease treatment and ethanol precipitation of the remaining fungal mass gave the four crude polysaccharide extracts, IOI-W, IOE-W, IOI-A and IOE-A. Since this procedure is a widely accepted method for

isolation of crude polysaccharides from plants and fungi, it was expected that the main component in these extracts was polysaccharides (Bouchard, Hofland, & Witkamp, 2007; Zhang et al., 2007). However, all four crude extracts appeared dark brown in color, indicating presence of substances other than polysaccharides. *I. obliquus* is known to contain massive amounts of melanin pigments, reported to be the main content in *I. obliquus* water extracts (Shashkina, Shashkin, & Sergeev, 2006). One study reported a 57 kDa allomelanin pigment isolated from *I. obliquus*, which could explain why the color remained in the extracts even after dialysis and ethanol precipitation (Kukulyanskaya, Kurchenko, Kurchenko, & Babitskaya, 2002). Indeed, the fractions obtained after anion-exchange chromatography of the crude water-extracts revealed that only a minor part of the extracts were polysaccharides. The crude extract IOI-W (3.2 g) gave two polysaccharide fractions - IOI-WN (80 mg) and IOI-WAc (75 mg) – which combined were responsible for 4.8 % of the total extract weight, or 0.45 % of total interior *I. obliquus* dry weight. Although most of the pigment was trapped in the ANX sepharose resin when eluting with water, some of the color came out with the acidic fractions IOI-WAc and IOE-WAc. The presence of phenolics was later detected in the acidic fractions (Table 1). Previous studies have reported fungal melanin to have a polyphenolic structure (Eisenman & Casadevall, 2012; Prados-Rosales et al., 2015), and in agreement with this phenolics were not detected in the fractions where the color was absent, such as IOI-WN. After anion-exchange chromatography, IOI-WN, IOE-WN and IOI-WAc were subjected to SEC. IOI-WAc was fractionated using Superdex 200, giving the fractions IOI-WAcF1, IOI-WAcF2 and IOI-WAcF3 in that order (Fig. S30). The other samples could not be fractionated further (Fig. S27-29).

The alkali-extracted interior part yielded a relatively pure polysaccharide fraction after SEC - IOI-A1 - accounting for 12.5 % of the alkaline crude extract (~ 1 % of total dry interior weight). In contrast, the extract from the exterior part yielded only heavily pigmented fractions after SEC with low amounts of polysaccharide (IOE-A1 and IOE-A2) and these were not analyzed further in this study. The presence of melanin in the polysaccharide extracts highlights the importance of proper fractionation when isolating polysaccharides from *I. obliquus*, especially if biological assays will be conducted.

3.2 Molecular weight determination by SEC-MALLS

The weight-average molecular weights (M_w) of the polysaccharide fractions and weight average radius of gyration (R_G) values were determined by SEC-MALLS analysis (Table 1, Fig. S19-S26). The neutral polysaccharides IOI/IOE-WN had higher M_w than the acidic ones (60/73 kDa vs 10-31 kDa), while the alkaline-extracted IOI-A1 gave the highest M_w of ~2000 kDa. However, the M_w of IOI-A1 must be interpreted with caution, since the SEC-MALLS analysis was conducted at pH 6 while IOI-A1 is poorly soluble at pH<8. An HPLC analysis of this sample was therefore carried out in addition to SEC-MALLS, dissolving IOI-A1 in 10 mM NaOH. The HPLC experiment indicated a polymer size larger than the biggest dextran standard utilized,

which was 450 kDa (data not shown). IOI-WN and IOE-WAc were suggested to contain more than one type of polymer, but when using prep grade SEC it was not possible to fractionate these any further. IOI-WAc was also suggested to contain more than one polymer. In the SEC-MALLS chromatogram of IOI-WAc (Fig S21), only one main peak can be observed. However, the peak was quite broad, indicating that it could contain polymers of more than one size. This seemed to be the case, as IOI-WAc was fractionated into the three sub-fractions IOI-WAcF1-F3 by gel filtration (Fig. 1). In general, the SEC chromatograms, notably the light scattering profiles, were too irregular for conventional data analysis, making it difficult to estimate the radius of gyration (R_G) values, as well as the number-average molecular weights (M_n). Nevertheless, the R_G (weight average) ranges observed across the peaks indicated that the polymers were in some cases rather compact structures, possibly aggregates, which could explain why we could not manage to further fractionate IOI-WN and IOE-WAc by SEC. There are a few reports in the literature on molecular weight of polysaccharides from *I. obliquus.* One group found a 48.8 kDa water-soluble polysaccharide from wild *I. obliquus*, in reasonable agreement with our own data (Chen et al., 2015). Another study reported water-soluble *I. obliquus* polysaccharides with a wide range of sizes, from <10 kDa up to 1.100 kDa (Kim et al., 2005), but the use of cultivated *I. obliquus* mycelium instead of wild fungus makes it difficult to compare these results directly with our own.

3.3 Purity of the polysaccharide fractions and LPS contamination

Table 1 gives an overview of the purity of the polysaccharide fractions, showing carbohydrate content, protein content, phenolic content and LPS contamination in the fractions. The fractions had a carbohydrate content ranging from 41.9-89 %. The percentages are based on the GC analysis, using mannitol as internal standard. It should be noted that there is some uncertainty to the calculations of these percentages. The peak coming from glucose in the GC chromatogram was too big for accurate quantification, possibly leading to an underestimation of actual content of this monosaccharide, but when diluting the samples the peaks from several of the other monosaccharides became too small. However, there was a high uncertainty in these calculations. In addition, the calculation for 3-*O*-Me-Gal was based on the galactose standard curve, meaning it could have been inaccurately quantified. All fractions were essentially free from proteins, as would be expected since the extracts were treated with a protease prior to fractionation. The phenolic content was related to the presence of melanin pigments as discussed in section 3.1. The acidic fractions had a higher content of phenolics than the neutral fractions. For example, IOI-WN did not contain any phenolics, while IOI-WAc had a phenolic content of 6.7 %. This suggests that the pigments are not covalently bound to the polysaccharides, although a possible explanation could be that the galacturonic acid (which is present in large amounts in the acidic fractions) interacts with the pigments in some way. When IOI-WAc was further fractionated with SEC, the three sub fractions IOI-WAcF1 – F3 contained 4,2 %, 5.8 % and 9.7 % phenolics, respectively, suggesting that the melanin pigments range in size from at least 10-31 kDa. The

alkali extracted IOI-A1 also contained some phenolic traces, accounting for 0.4 % of total sample weight. The samples were analyzed for the presence of LPS, a constituent of the outer cell membrane in gram-negative bacteria and a common contaminant in natural products research (Lieder, Petersen, & Sigurjonsson, 2013). Only the acidic fraction IOE-WAc from the exterior part was found to contain LPS, accounting for 0.16 % of the total sample weight which is equivalent to 160 ng/ml LPS in a 1 mg/ml sample. Biological assays are often highly sensitive to LPS, and it has been reported that LPS at less than 1 ng/ml can be recognized by human immune cells (Lieder et al., 2013). This underlines the need for caution when interpreting results when using natural products in such assays, and the analysis for the presence/absence of LPS should always be carried out prior to the experiments.

Table 1: M_w, R_G, carbohydrate content, protein content (expressed as albumin), total phenolic content (expressed as ferulic acid) and LPS content in the isolated polysaccharide fractions.

Fraction	M_w (kDa)	R_G (nm)	Total carbohydrate (%)	Total protein (%)	Total phenolics (%)	LPS contamination
IOI-WN	60	57[1]	89	<0.1	<0.1	n.d.
IOI-WAc	15	<10-15[2]	69.1	<0.1	6.7	n.d.
IOI-WAcF1	28	17	81.9	<0.1	4.2	n.d.
IOI-WAcF2	14	<10-15[2]	61.2	<0.1	5.8	n.d.
IOI-WAcF3	10	<10-15[2]	58.3	<0.1	9.7	n.d.
IOE-WN	73	36	54.7	<0.1	<0.1	n.d.
IOE-WAc	31	29[1]	47.3	<0.1	2.0	0.16 %
IOI-A1	>450	n.e.	41.9	<0.1	0.4	n.d.

[1] may be grossly overestimated due to irregular light scattering data; [2] below detection limit; n.e. = not established; n.d. = not detected; Limit of detection for LPS: 2 ng/µl (de Santana-Filho et al., 2012).

3.4 Monosaccharide composition

The monosaccharide composition of the polysaccharide fractions is presented in Table 2. The most common monosaccharide in all water-soluble fractions was glucose, with galactose, mannose and xylose present in substantial amounts. The acidic fractions had high levels of galacturonic acid and increased amounts of rhamnose compared to the neutral fractions, while arabinose, fucose and glucuronic acid were present in minor amounts in all fractions. 3-*O*-methyl-galactose (3-*O*-Me-Gal) was found in varying amounts in all fractions, and the amount

decreased with decreasing size of the polymer, as seen by comparing the acidic fractions IOI-WAcF1, F2 and F3 (Table 1 and 2). The alkali-extracted IOI-A1 contained mainly glucose (79.2 %) and xylose (12.1 %), with traces of mannose, fucose and N-acetylglucosamine (GlcNAc). There are a few publications on isolated and purified *I. obliquus* polysaccharides. Chen et al. isolated a polysaccharide consisting of rhamnose, arabinose, glucose and galactose in a molar ratio of 2.5:4.6:1.0:2.6 with approx. 30 % uronic acid (Chen et al., 2015), while Fan et al. isolated a polysaccharide consisting of rhamnose, mannose and glucose in ratio 1.0:2.3:1.7 with 7.5 % uronic acid (Fan et al., 2012). The high amounts of arabinose and rhamnose in these studies are in contrast to our own data. There are several possible explanations for the differences, such as isolation methods, harvest location and age of the fungus, and which part of the fungus was used. Since arabinose and rhamnose are typical plant monosaccharides (Liu, Willför, & Xu, 2015), it could also be possible that the *I. obliquus* samples were "contaminated" with birch material after harvesting, which is quite common when collecting wild-grown *I. obliquus* sclerotia. Apart from these papers, most of the studies on *I. obliquus* polysaccharides have been carried out using cultured mycelium rather than wild-grown fungi, and the results are therefore not directly comparable (Kim et al., 2006; Xu, Quan, & Shen, 2015).

Table 2: Monosaccharide composition (mol %) after methanolysis and GC of the polysaccharide fractions isolated from *I. obliquus*.

Monosaccharide	**IOI-WN**	**IOE-WN**	**IOI-WAc**	**IOI-WAcF1**	**IOI-WAcF2**	**IOI-WAcF3**	**IOE-WAc**	**IOI-A1**
Arabinose	1.4	2.0	2.1	2.3	2.3	2.4	0.9	n.d.
Rhamnose	1.3	1.3	5.1	4.8	5.4	7.7	3.4	n.d.
Fucose	1.7	2.5	1.3	1.5	1.0	0.7	0.5	2.8
Xylose	6.9	8.9	8.8	4.7	5.2	19.0	6.0	12.2
Mannose	6.8	7.2	8.2	10.8	13.4	4.8	8.4	4.0
Galactose	17.5	17.2	11.2	13.5	9.6	8.1	15.6	n.d.
Glucose	53.1	49.3	37.9	34.3	43.1	39.2	31.6	78.2
Glucuronic acid	1.0	2.1	5.1	4.4	4.0	4.4	3.1	n.d.
Galacturonic acid	1.8	1.2	15.7	17.7	13.1	12.9	24.3	n.d.
3-*O*-methyl galactose	8.7	8.3	4.6	6.0	2.9	1.1	6.2	n.d.
N-acetylglucosamine	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.8
Sum %	100	100	100	100	100	100	100	100

(n.d. = not detected)

3.5 Chemical determination of (1→6)-α-*3*-O-*Me-Gal*

The results from the monosaccharide composition analysis by GC revealed three unknown peaks in the chromatogram, which based on previous discoveries in our lab was suggested to come from a methylated hexose (Fig. 2). To identify the unknown compound, IOE-WN was demethylated using BBr_3 and analyzed by GC after methanolysis before compared with the original GC data. The three peaks in question disappeared after demethylation, while the ratio between mannose, galactose and glucose had changed from ~1:2:7.5 to ~1:3:6, suggesting that the methylated species was a galactose. The IOE-WN fraction was ethylated and analyzed by GC-MS to find the position of the methyl group. From the linkage analysis data (Table 3) it was seen that most of the galactose was (1→6)-linked, and therefore we searched for fragments which matched this configuration, theoretically placing the methyl group in position C2, C3 or C4 to see which fragments were expected after ethylation. We located a large peak containing the fragments 132, 190, 203 and 261 m/z, which were expected if a hexose was methylated at C3. Ultimately, this finding was verified by NMR spectroscopy including anomeric configuration (Fig. 3, Table 4), suggesting presence of (1→6)-α-3-*O*-Me-Gal residues in most fractions. The present study is the first to report this uncommon monosaccharide in *I. obliquus.* There are some papers describing (1→6)-α-3-*O*-Me-Gal from other fungi, such as species from the genera *Phellinus* and *Pleurotus* (Carbonero et al., 2008; Yang et al., 2007). With the presence of this residue in *Inonotus* as well, it seems that extensive 3-*O*-methylation of galactose is more common in polypore fungi than previously thought, although the biological role of such methylation is still largely unknown (Staudacher, 2012).

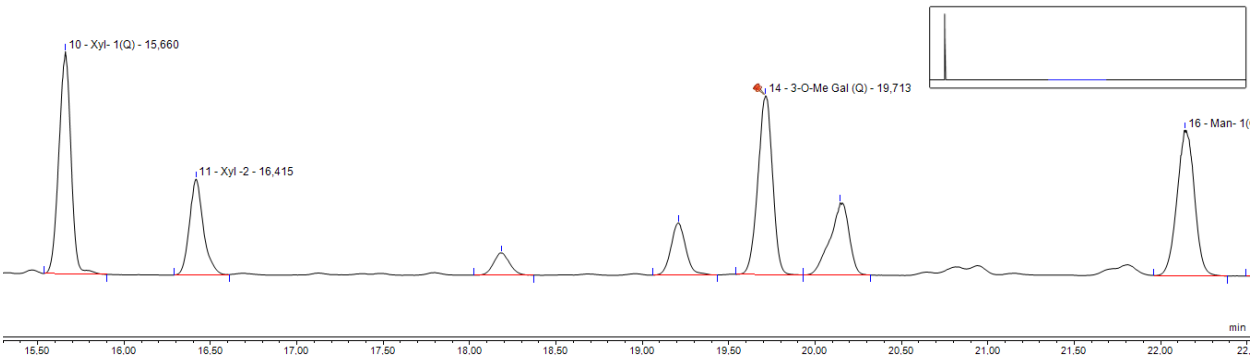


Fig. 2: The three chromatographic peaks (19-20.50 min) from 3-*O*-Me-Gal after methanolysis and GC analysis of IOE-WN, located between the peaks from xylose (15.5-16.5 min) and mannose (22 min).

3.6 Linkage analysis by methylation/GC-MS

To reveal how the different monosaccharides were linked, the methylated alditol acetates of the monosaccharides were analyzed by GC-MS, as seen in Table 3. The results are presented as total mol% of each sub-species based on the percentage of each monosaccharide initially found by GC analysis after methanolysis (Table 2). The most common glucose species in the water-extracted fractions were (1→3)-linked Glc and (1→6)-linked Glc. There were also substantial amounts of

terminal (non-reducing) Glc (T-Glc) and (1→3, 6)-linked Glc, indicating a highly branched structure of the polysaccharides. The presence of (1→2, 3)-linked Glc (~1%) and (1→2, 3, 6)-linked Glc (~1%) suggest that glucose could be substituted at position C2 by arabinose, fucose, xylose and glucuronic acid, which were all present in minor amounts and were mostly of terminal origin. Galactose was the second most common monosaccharide, being almost exclusively (1→6)-linked, with small amounts of T-Gal and (1→2, 6)-linked Gal indicating that also galactose was substituted at C2. The percentage of the different galactose species found from GC-MS is a combination of native and methylated galactose, since these two types would appear identical after methylation of the samples, except for (1→3) - and (1→3, 6)-linked Gal which only could come from native galactose. Mannose was mostly (1→2)- and (1→3)-linked, with some T-Man as well. The acidic fractions had substantial amounts of (1→4)-linked GalA, which was virtually absent in the neutral fractions IOI-WN and IOE-WN.

The presence of (1→4)-linked GalA, (1→2)- and (1→2, 4)-linked Rha, as well as some (1→3)-linked Gal and T-Ara indicates that a part of the polysaccharides had a pectin-like structure (Yapo, 2011). It could be that some birch material was mixed with the fungal material prior to extraction and fractionation. If this were true, the traces should only be prominent in the fractions isolated from the interior part of the fungus, since this part is in direct contact with the birch. Since fractions isolated from both the interior and exterior parts contained the pectin-resembling linkages, it is likely that there exists some other explanation for the linkages observed.

Table 3: Glycosidic linkages (mol %) in the polysaccharide fractions isolated from *I. obliquus* detected by GC-MS analysis of methylated alditol acetates.

Linkage	IOI-WN	IOE-WN	IOI-WAc	IOI-WAcF1	IOI-WAcF2	IOI-WAcF3	IOE-WAc	IOI-A1
Ara								
T	1.4	2	2.1	2.3	2.3	2.4	Trace	n.d.
Rha								
T	n.d.	Trace	Trace	Trace	2.3	n.d.	1.4	n.d.
1,2	Trace	Trace	2.1	2.1	1.1	5	Trace	n.d.
1,3	Trace	Trace	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
1,2,4	n.d.	Trace	2.1	2	2	2.7	1	n.d.
Fuc								
T	1.7	2.5	1.3	1.5	1	Trace	Trace	1.9
Xyl								
T	3.1	4.7	5.3	3.3	3.5	3.7	2.6	2.8
1,4	3.8	4.2	3.5	1.4	1.7	16.3	3.4	9.4

Man								
T	1.6	1.4	1.8	1	1.8	2.5	2.6	n.d.
1,2	2.4	2.4	3.2	5.3	1.8	n.d.	2.4	n.d.
1,3	2.8	3.4	3.2	4.5	9.8	2.3	3.4	3.5
Gal								
T	Trace	2.8	2.2	1.3	3.6	6.1	1.1	n.d.
1,4	Trace	n.d.	1.2	Trace	1.4	n.d.	Trace	n.d.
1,3	Trace	1.1	Trace	Trace	Trace	n.d.	Trace	n.d.
1,6	22.1	17.9	9.4	14.4	6.1	3.1	17.2	n.d.
1,2,6	2.2	2	1.2	1.8	n.d.	n.d.	2.7	n.d.
1,3,6	Trace	1.1	Trace	1.3	Trace	n.d.	Trace	n.d.
Glc								
T	9.5	9.7	7.1	6.9	9.7	7.9	4.8	7.8
1,3	19	17.2	12.3	12	12.8	22.1	10.1	55.6
1,4	1.7	3.1	3.2	1.6	7	Trace	2	n.d.
1,6	16.2	12.8	9.6	8.2	10.3	6.1	9.1	3.2
1,3,4	Trace	Trace	Trace	Trace	n.d.	n.d.	Trace	2.6
1,2,3	Trace	1.3	Trace	1	n.d.	n.d.	1.3	2.5
1,3,6	5.3	5.1	4.6	3.8	2.9	2.1	3.3	5.9
1,4,6	Trace	Trace	Trace	Trace	Trace	n.d.	Trace	n.d.
1,2,3,6	1	Trace	1.1	Trace	Trace	Trace	1.0	Trace
GlcA								
T	Trace	1.2	1.7	1	n.d.	n.d.	2.5	n.d.
1,4	Trace	trace	3.4	3.4	4	4.4	Trace	n.d.
GalA								
T	Trace	Trace	3.8	1.1	n.d.	1.4	3.2	n.d.
1,4	n.d.	Trace	10.2	11.2	13.1	11.5	16.5	n.d.
1,3,4	1.3	Trace	1.6	5.4	n.d.	n.d.	4.6	n.d.

(n.d. = not detected; trace = less than 1 %; T = terminal, non-reducing end)

3.7 IR spectroscopy

The polysaccharide fractions were analyzed by IR spectroscopy. There were several bands in the IR spectra that could be assigned to structural motifs from carbohydrates. There was a strong, but broad band at 3360 cm^{-1} coming from –OH groups, and several bands around 900-1050 cm^{-1} that probably came from (C-O-C) in glycosidic linkages of various configurations. The bands from C-O-C will usually shift slightly depending on type of linkage (α, β, 1→6 vs 1→3 etc.) and it

was therefore expected that this area was crowded with signals, as observed in the spectra (Wiercigroch et al., 2017). In addition, a band with strong intensity could be observed at 1650 cm^{-1} in the acidic fractions, probably from a C=O in a carboxylic group coming from the uronic acids.

3.8 NMR spectroscopy

The NMR data were interpreted by comparing and matching chemical shift values from the 1D spectra ^{1}H and ^{13}C, and 2D spectra DQF-COSY, TOCSY, NOESY, HSQC, HSQC-TOSCY and HMBC. It was possible to deduce the monosaccharide structure for many of the major signals, including possible linkages. The polysaccharide fractions were analyzed separately, and differences between them were deduced by comparing the different NMR spectra and by comparison with GC- and GC-MS data. CASPER predictions and reference values from the literature were also used to assign the chemical shift values (Agrawal, 1992; Duus, Gotfredsen, & Bock, 2000; Jansson, Stenutz, & Widmalm, 2006; Lundborg & Widmalm, 2011). In general, the signals belonging to the different monosaccharide species were assigned using the following procedure: First, an anomeric (H1) ^{1}H-^{1}H signal was located in a DQF-COSY spectrum. The H1 signal was then correlated with the H2 signal in the same spin system, and it was attempted to find further correlations to H3 and so on, using DQF-COSY and TOCSY. If the rest of the spin system could not be determined this way, HSQC-TOCSY, HMBC and NOESY spectra were used to find the last ^{1}H signals for that species and to look for long-range correlations and coupling to other monosaccharide species. HSQC and ^{13}C spectra were used continuously to assign the ^{13}C signals that correlated with the specific ^{1}H signals. The different samples were usually compared to each other using overlay plots, for example overlaying the HSQC spectra from IOE-WN and IOI-WAcF1 to look for differences. If some signals could still not be assigned, either due to overlap with other signals or because of weak signal intensity, CASPER prediction was used to get an idea of the location (Jansson et al., 2006; Lundborg & Widmalm, 2011). In general it was difficult to find evidence of linkages between the different monomers in the HMBC and NOESY spectra, such as direct linkages from Xyl to Glc, or from GalA to Gal. Since all fractions were highly heterogeneous, it means that either these linkage signals were too weak to observe by NMR analysis, or that the fractions consisted of several separate polymers which were diluted together during fractionation. However, the latter explanation seems unlikely. During column chromatography, we loaded sample onto the column step-wise, a small amount each time, and monitored the fractionation using the phenol-sulfuric acid assay. The polysaccharides were eluted at the same point each time, and if the samples were to be a mix of "entangled" polymers, there would have been variation in elution times from experiment to experiment. It therefore seems more likely that the fractions contained heteropolysaccharides but that the linkages were too difficult to capture by NMR spectroscopy due to low relative abundance of each specific linkage signal. All water-soluble samples had presence of major peaks in the anomeric region, defined as having chemical shift values between 4.5-5.5 ppm for

^{1}H and 95-105 ppm for ^{13}C (Fig. 3) (Duus et al., 2000). Typically, α-anomeric signals will appear downfield of β-anomeric signals for ^{1}H chemical shifts, and opposite for ^{13}C chemical shifts. The chemical shift values of the major signals found are presented in Table 4. There was evidence of reducing ends present in some fractions (Fig. S9-S14). Although these signals correlated with the relative sizes of the fractions (lower M_w means more reducing ends), the signals were stronger than they should be according to SEC-MALLS. For example, in IOI-WAcF2 the total signal strength of the reducing ends was about 6.7% compared to the anomeric signals, theoretically giving a M_w of ~2.5 kDa instead of 14 kDa as seen from SEC-MALLS (Table 1). One possible explanation for the observed effect is that reducing ends located near the surface would give stronger NMR signals than the monomers located inside the polymer, since they would be more flexible. This could be particularly true for the polysaccharides in this study, since the R_G values suggest that the structures are very compact (Table 1). This could result in the reducing ends having T2 relaxation times more similar to oligo- or monosaccharides, and longer T2 relaxation times would in turn result in lower loss of magnetization during mixing times (Duus et al., 2000).

Table 4: Chemical shift values of the major signals found in the water-extracted polysaccharide fractions from *I. obliquus*. The values are expressed as ppm from each position in that species, from C1/H1 (1) to C6/H6 (6), and are based mainly on spectra from IOE-WN, IOI-WAc and IOI-WAcF1-F3. Some of the species could only be partially assigned due to overlaps and/or weak signal intensities.

Type of monosaccharide w/linkage	1	2	3	4	5	6	3-*O*-Me
(1→3)-β-Glc-(1→3)							
δ^{13}C	102.48	73.17	84.59	68.17	75.44	60.80	
δ^{1}H	4.87	3.65	3.85	3.61	3.6	3.83/4.03	
1→6)-β-Glc-(1→3)							
δ^{13}C	102.92	73.35	75.42	69.62	74.59	68.85	
δ^{1}H	4.80	3.47	3.62	3.51	3.80	3.96/4.31	
(1→6)-β-Glc-(1→6)							
δ^{13}C	102.72	73.05	75.62	69.52	74.82	68.85	
δ^{1}H	4.61	3.42	3.59	3.55	3.72	3.96/4.31	
(1→3)-β-Glc-(1→6)							
δ^{13}C	102.40	72.58	84.79	68.11	75.44	60.80	
δ^{1}H	4.63	3.61	3.84	3.60	3.58	3.83/4.03	
T-β-Glc-(1→6)							
δ^{13}C	102.72	73.05	75.62	69.52	74.82	60.80	
δ^{1}H	4.61	3.42	3.59	3.55	3.72	3.83/4.03	

(1→4)-β-Xyl							
$\delta^{13}C$	101.87	74.21			62.9		
$\delta^{1}H$	4.55	3.34	3.66	3.82	3.46/4.18		
T-β-Xyl-(1→4)							
$\delta^{13}C$	101.69				65.16		
$\delta^{1}H$	4.56	3.38			3.4/4.06		
(1→6)-α-Gal-(1→6)							
$\delta^{13}C$	97.72	67.95	69.83	69.23	69.11	66.76	
$\delta^{1}H$	5.08	3.93	3.97	4.11	4.26	3.78/4.01	
(1→6)-α-3-O-Me-Gal-(1→6)							
$\delta^{13}C$	97.72	67.42	78.77	65.1	68.77	66.76	56.02
$\delta^{1}H$	5.07	3.97	3.63	4.38	4.27	3.78/4.01	3.53
(1→4)-α-GalA-(1→4)							
$\delta^{13}C$	98.92	68.02	68.62	77.72	71.12	174.70	
$\delta^{1}H$	5.17	3.85	4.09	4.52	4.85		
(1→3)-α-Man							
$\delta^{13}C$	102.1	70.38	78.43			62.84	
$\delta^{1}H$	5.21	3.86	4.12			3.82/3.98	
(1→4)-α-GlcA							
$\delta^{13}C$	98.9						
$\delta^{1}H$	5.34		4.21				
(1→2)-α-Rha							
$\delta^{13}C$	102.3					16.93	
$\delta^{1}H$	5.09	4.18		3.5	3.9	1.33	

3.8.1 β-glucose

The HSQC spectra revealed the presence of two main signal clusters in the β-anomeric region located at ($^{1}H/^{13}C$) ~4.85/102.5 ppm and ~4.62/102.9 ppm (Fig. 3). With the use of DQF-COSY, TOCSY as well as HSQC-TOCSY spectra, the signals were suggested to come from β-glucose: β-Glc-(1→3)-species at ~4.85/103 ppm and β-Glc-(1→6) species at ~4.6/103 ppm (Table 4). The peaks were too broad to contain just one signal, and the HSQC and ^{13}C spectra indicated presence of signals from several different anomeric carbons, probably variations of the different

β-Glc species, such as (1→3-β-Glc-1→3), (1→6-β-Glc-1→3), (1→3-β-Glc-1→6), and (1→6-β-Glc-1→6). Anomeric signals from even more complex species including (1→3, 6)-branching points, or with C2 in glucose additionally substituted by another monosaccharide as suggested from the linkage analysis data (Table 3) would also be located in the same anomeric signal clusters. Further, signals from terminal glucose would also be located here, as these could be defined as pure β-Glc-(1→3) or β-Glc-(1→6) species with no other substitutions. The two main clusters were defined on the basis of the assignment methods described above, and also because the two groups could be differentiated by the relatively large difference in shift values for C6 in the HSQC spectra, where the (1→6)-β-Glc has a ^{13}C chemical shift downfield of its (1→3)-β-Glc counterpart (Table 4). In addition, the C3 signal from (1→3)-β-Glc has a very characteristic chemical shift value, in our samples observed at ~3.85/85 ppm in the HSQC spectra, thus confirming that the signals come from glucose. It was attempted to find correlations between the various β-Glc species by comparing integrals in the HSQC spectra (which could not be done in 1D spectra due to overlaps). Although 2D integrals are less accurate and overlap also occurred in these spectra, some clues were found that could indicate a possible configuration. In IOE-WN, the HSQC integral from the anomeric β-Glc-(1→3) cluster was approximately the same size as the signal from C3/H3 in (1→3)-β-Glc. The signal from C4/H4 in (1→3)-β-Glc was about 25 % smaller than the others, but a signal located at same ^{13}C ppm but 0.1 ^{1}H ppm downfield corresponded well with these 25 % and according to CASPER this signal could come from (1→3, 6)-β-Glc-(1→), suggesting several branching points in the polymer. The relative proportion of (1→3)-linked and (1→6)-linked β-Glc was approx. 1:1 for most samples according to GC-MS (Table 3), while the signal cluster in the HSQC spectra from β-Glc-(1→6)-species was twice as big as the β-Glc-(1→3)-cluster (Fig. 3). Methylation/GC-MS analysis does not give information on how terminal residues are linked but NMR does, and a possible explanation for the difference in signal strength could therefore be that most of the terminal Glc in our samples is (1→6)-linked.

3.8.2 β-xylose

Some signals close to the two major peaks in the β-anomeric region were assigned to (1→4)-linked β-xylose species (Fig. 3, Table 4). β-Xyl has axial protons for H1, H2, H3 and H4, making it easier to detect than some other monosaccharides since the H,H couplings will give strong signals which can be correlated through DQF-COSY and TOCSY. The anomeric ^{1}H signals could therefore be correlated throughout the molecule from H1-H5, but some of the ^{13}C signals could not be assigned due to overlaps (C3 and C4). In the HSQC spectra, opposite phase chemical shifts for H5a and H5b were in accordance with CASPER prediction for (1→4)-β-Xyl and terminal β-Xyl(1→4). The H-H split of C5 in xylose is quite large and characteristic compared to some other monosaccharides (Agrawal, 1992). There were some differences between the samples for the H5 signals. For example, IOE-WN had one clear ^{1}H-^{1}H-split from H5 in (1→4)-β-Xyl, while in IOI-WN there were two such splits differing by 0.05 ppm, suggesting more than one (1→4)-β-Xyl species in IOI-WN. This could mean that xylose chains

of various lengths or xylose being in different local environments are present, giving rise to more signals. From GC-MS data it was found that the IOE-WN fraction had slightly more T-Xyl (4.7 %) than (1→4)-Xyl (4.2 %), suggesting that xylose is substituting a main polysaccharide chain as either mono- or disaccharides with (1→4) linkages. The IOI-WN fraction had slightly more (1→4)-Xyl (3.8 %) than T-Xyl (3.1 %), suggesting that the xylose side chains would be of varying lengths, which would fit well with the NMR spectra observed. However, we could not observe any HMBC correlations from xylose to glucose or galactose, but this could also be due to the relative rarity of these linkages in the samples. As seen from Table 2, xylose is present in relatively high amounts in all fractions, also when galactose, mannose and galacturonic acid levels are low. It therefore seems likely that xylose is somehow linked to glucose within the polymers chain.

3.8.3 α-galactose and α-3-O-Me-Gal

All samples had presence of several strong signals in the α-anomeric region (Fig. 3). The signal at 5.07/97.72 ppm was suggested to come from (1→6)-α-3-*O*-Me-Gal. Using DQF-COSY it was found that this signal actually correlated with two separate H2 signals, at 3.97 and 3.93 ppm. Using DQF-COSY and TOCSY the signals at 3.97 ppm could be assigned through the spin system from H2-H6. The HMBC spectrum showed a correlation between the H3/C3 of this species to a methyl peak at 3.53/56.02 ppm, and as such this species was designated (1→6)-α-3-*O*-Me-Gal. The other signal in the same anomeric peak was assigned to (1→6)-α-Gal. According to CASPER prediction, the anomeric ^{13}C chemical shift values from these two species should actually be separated by 2 ppm. However, since the signal cluster ranged from ~96.80-99 ppm, and others also have reported the same findings as us, it seems logical that the two signals would be located in the same area (Smiderle et al., 2008). This would also explain the relatively strong intensity of the anomeric peak, which had equal signal intensity as the β-Glc-(1→6) signal cluster. In both species the rest of the ^{13}C signals could be assigned using HSQC and HMBC. The ^{13}C signal from C6 could not be differentiated between the two species, and had a relatively downfield shift value compared to the expected value for galactose (Agrawal, 1992). This was probably due to the substitution at C6 with another Gal residue, which was further strengthened by finding a strong correlation from C1 to C6 in HMBC as well, suggesting that a part of the polymer chain has a (1→6)-α-Gal-(1→6) motif with about every second or third residue being 3-*O*-methylated.

3.8.4 α-GalA

In the acidic fractions only, a signal at 5.17/98.92 ppm with strong intensity was found in the α-anomeric region (Fig. 3a, Fig S9-S14). The signal was suggested to come from (1→4)-α-GalA since this species should be present in high amounts in acidic fractions and absent from the neutral fractions according to GC-MS data (Table 3). The signal could be correlated throughout

the ^{1}H spin system from H1-H5 and it was also possible to assign the ^{13}C signals to each ^{1}H signal (Table 4). According to CASPER prediction, GalA was expected to have ^{1}H chemical shifts values for H4 and H5 close to the β-anomeric region but with highly different ^{13}C values, and this seemed to be the case in our samples as well, with H4 at 4.52 ppm and H5 at 4.85 ppm (Agrawal, 1992). These two signals could be used to distinguish GalA from GlcA, since GlcA would have ^{1}H shift values at much lower ppm according to CASPER. In the ^{13}C and HMBC spectra there was a signal at 174.70 ppm which indicated presence of a C6 carboxyl group coming from an uronic acid. Further, in several HMBC spectra we could correlate C4 and H1 and H4 and C1 to each other by HMBC, indicating that this species was a long chain of (1→4)-α-GalA-(1→4). No linkage correlations between GalA and other species could be observed, but this could be due to a very low relative abundance of monomers linking the GalA chain to the rest of the polymer.

3.8.5. Other α-linked species

In most fractions there were also several minor signals that could not be completely assigned to specific species due to signal overlap and the relative weakness of the signals (Figure 3a). GC and GC-MS data indicated that glucuronic acid, rhamnose and mannose were present in various amounts in all samples (Table 3) and several of the signals in the α-anomeric region probably came from these monosaccharides. As for GlcA it was difficult to locate most signals, probably because of the low abundance in all samples (Table 4). According to CASPER prediction, (1→4)-α-GlcA is expected to have anomeric ^{1}H chemical shift values downfield of GalA, and the anomeric signal at 5.33/98.62 was consistently present in the acidic fractions where (1→4)-GlcA was present. Although clear correlations to other signals could not be seen using DQF-COSY, the anomeric signal could be correlated to a ^{1}H signal at 4.21 ppm using the TOCSY spectrum, which could be from H3 in the same spin system, and this also was in relative accordance with CASPER prediction. Other signals from (1→4)-GlcA could not be observed due to unknown reasons, thus the signal assignments remains unclear. GC-MS data also indicated the presence of (1→2)- and (1→3)-linked mannose, and mannose is usually in the α-configuration in fungi (Arana et al., 2009). The HSQC signal at 5.21/102.27 ppm could come from (1→3)-linked α-Man (Fig. 3). However, this signal could not be correlated further to the H2 signal, which could be due to the fact that mannose can be quite difficult to deduce by NMR spectroscopy, as the equatorial H2 gives weak couplings to H1 and H3. This means that high amounts of each mannose species probably would be required for sufficient signal intensities to be produced. The absence of clear correlations is in this sense in accordance with GS-MS data which suggested that only small amounts of each mannose species were present in the samples (Table 3). Additionally, it was only in IOI-WAcF2 – the fraction with the highest amount of (1→3)-α-Man - that the anomeric signal could be located clearly, with some traces in the IOI-WAc sample as well. There was also a slight indication of a correlation between ^{1}H-signals at 4.155 and 5.369 ppm in NOESY, which could come from mannose (Fig. S15). However, this

could not be found in HSQC-TOCSY or other spectra, which makes it possible that it could come from across the glycosidic linkage and not from the same monomer.
Rhamnose was thought to be in α-configuration due to the presence of several non-assigned anomeric signals in the α-anomeric region. Using spectra from IOI-WAcF3, a signal was located at 1.33/16.93 ppm in HSQC which likely was from a methyl in a 6-deoxy sugar. In DQF-COSY, this signal had correlations to another ^{1}H signal at ~3.90 ppm, which could be the H5 signal from α-rhamnose (Agrawal, 1992). This signal could be correlated to signals at ~3.5 ppm and 4.18 ppm using TOCSY, and although the peaks were quite broad this suggests the presence of at least one α-Rha species in the samples, probably (1→2)-α-Rha (Fig. S16 and S17).

3.8.6 IOI-A1

The alkali-extracted fraction IOI-A1 was also subjected to NMR analysis. However, since the fraction was dissolvable only in alkali and neither H_2O or dimethyl sulfoxide (DMSO), the sample was dissolved in 0.1 M NaOH in D_2O which led to some results not consistent with the other samples. The HSQC spectrum revealed presence of some strong signals coming from β-Glc species (Fig. S18). Some of the signals could be assigned with a high degree of certainty, such as two (1→3)-β-Glc signals at 3.79/87.7 ppm (from H3/C3) and 3.83/4.02/61.1 ppm (from H6a/H6b/C6). Other signals seemed to come from (1→6)-linked species. Some of the signals had shifted notably compared to the other samples, possibly due to the difference in pH. In addition, the large size of the IOI-A1 fraction (>450 kDa) and the presence of Na^{+} in the sample could give weaker signals than expected, especially for the backbone of the polymer which is primarily made up of (1→3)-linked β-Glc, whereas the side-chains of higher mobility would be less affected.
The GC-MS data suggested that IOI-A1 contained mainly (1→3)-Glc (57.3 %), with presence of (1→6)-Glc (3.3 %), (1→3, 6)-Glc (5.6 %) and T-Glc (8 %) (Table 3). The strongest signals seen in ^{13}C (Fig. S31) and HSQC spectra of IOI-A1 came from what seemed to be either (1→6)-linked β-Glc or (1→3, 6)-β-Glc species, and it therefore seems likely that the terminal β-Glc would be (1→6)-linked, thus accounting for some of the signals seen in the NMR spectra. This would give a polymer with a backbone consisting of (1→3)-linked β-Glc with short side chains of (1→6)-linked β-Glc, possibly substituted with β-Xyl in between.

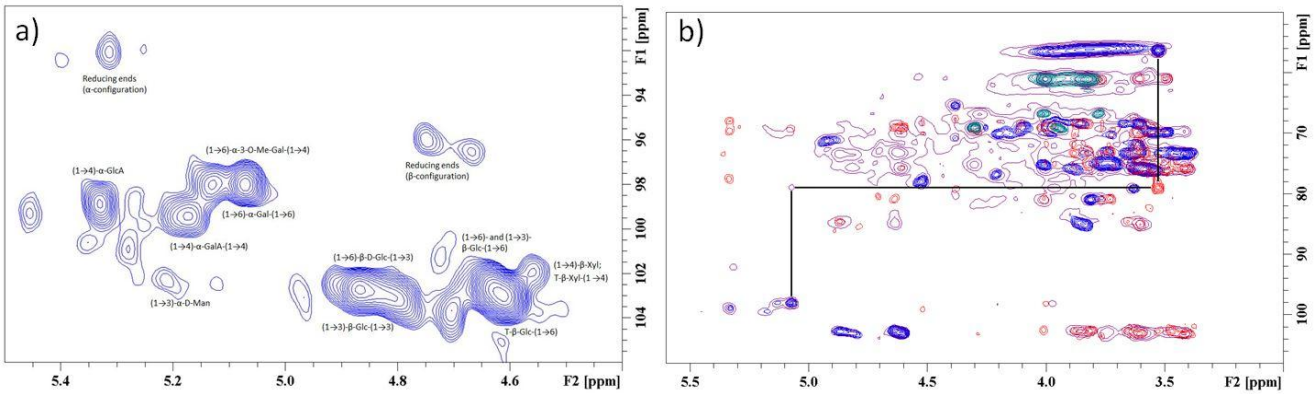


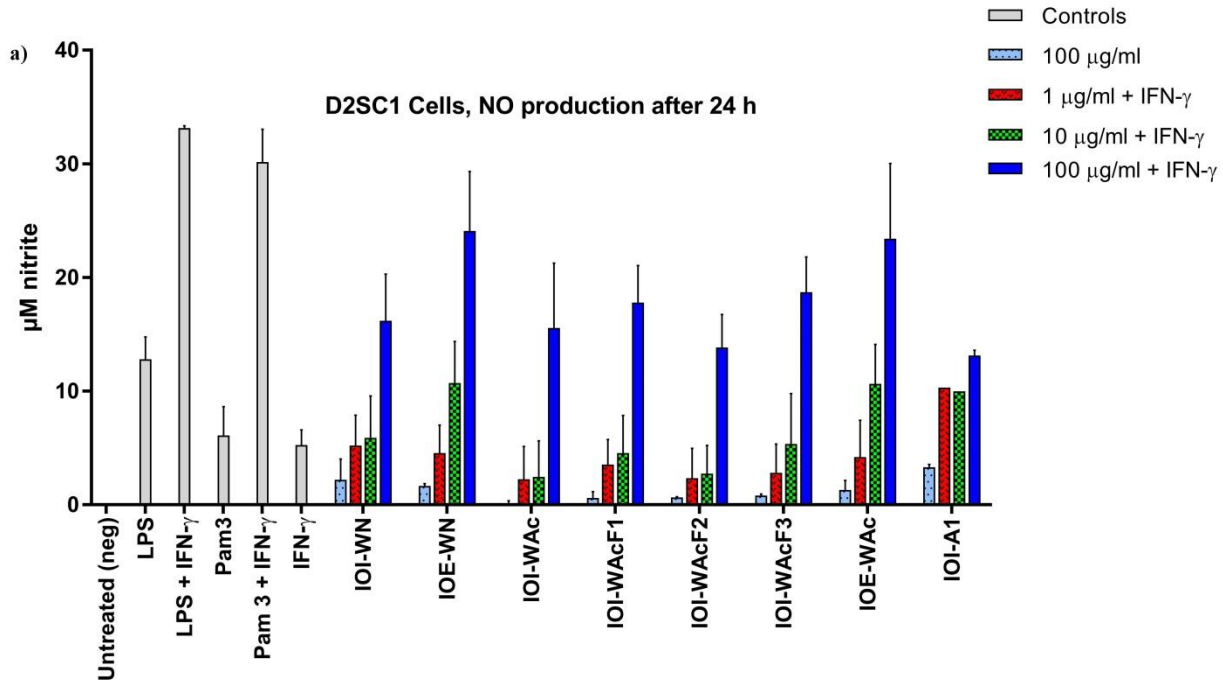
Fig. 3: (a) NMR analysis of fraction IOI-WAcF2, showing the anomeric region from the HSQC spectrum with assignments of the major signals in this region. Chemical shifts are expressed in ppm. (b) Overlay plot from IOI-WAcF2 of HSQC (blue)/HSQC-TOCSY (purple)/HMBC (red), showing the main carbohydrate region between 3-6 ppm/50-110 ppm $^{1}H/^{13}C$. This plot gives an idea of the complexity of the sample. The black line shows a direct link from the *O*-methyl signal at 3.53/56.02 ppm to a C3 signal from α-Gal at 3.63/78.77 ppm, which could be traced back to the C1 signal at 5.07/97.72 ppm. This species was assigned as (1→6)-3-*O*-Me-α-Gal.

3.9 Periodate oxidation and Smith degradation

IOE-WN was subjected to periodate oxidation and Smith degradation to find clues about the relationship between homologous stretches in the polymer chain and linkages. Since the (1→3) linked sugars should not be hydrolyzed by the weak acid utilized after periodate oxidation (0.05 M TFA), it would be expected that some oligosaccharides were produced if stretches of (1→3)-linked β-Glc existed (Kamerling & Gerwig, 2007). Fractionation of the degraded sample gave three main fractions which were analyzed by GC. While the two first fractions consisted of monomers (3-*O*-Me-Gal and Glc, respectively), the last fraction had a retention time in the middle of the maltotetraose and maltohexaose standards, and contained only glucose. This indicated that there were stretches of about 4-6 (1→3)-linked glucose monomers before a change in the polymer chain occurs. Given the relatively large proportion of (1→6) linked Glc and Gal in the samples, it could be that the main motif of the polysaccharides is (1→3)-linked β-Glc five monomers long with (1→6)-linked "kinks" in between with either Glc or Gal residues. There are some reports of similar structures in the literature, such as galactoglucans from the fungus *Lentinus fusipes* (Manna et al., 2017) and from the bacteria *Rhizobium meliloti* (Zevenhuizen, 1997). In addition, alternating (1→4)/(1→3)-linkages of glucose is the main motif in oat and barley β-glucans (Brennan & Cleary, 2005), suggesting that the possibility of alternating linkages in the main chain of polysaccharides from other species such as fungi may exist as well.

3.10 Nitric oxide (NO) assay

The polysaccharide fractions were screened for their ability to induce nitric oxide (NO) production in the murine dendritic cell line D2SC/1 and the murine macrophage cell line J774A1. NO regulates many processes in the immune system and is an important molecule in human physiology (Hickok & Thomas, 2010), and high NO production by macrophages and dendritic cells can be considered a marker for a pro-inflammatory phenotype (Curren Smith, 2015). To investigate the immunomodulating effect of *I. obliquus* polysaccharides, cells were incubated with polysaccharide fractions with or without IFN-γ for 24 h. IFN-γ has been shown to have a synergistic effect with TLR ligands on NO production in immune cells (Qiao et al., 2013; Totemeyer et al., 2006). The results are presented in Fig. 4. Although the fractions failed to activate NO production when used alone, most fractions demonstrated potent activity with IFN-γ at the highest concentration (100 μg/ml), indicating a synergistic effect which also was observed using the positive control Pam3CSK4. The neutral fraction IOE-WN gave a clear dose-response relationship, and was therefore considered the most promising fraction in this assay. IOE-WAc initially seemed promising; however it was found that this fraction contained LPS equivalent to 16 ng/ml at 100 μg/ml (0.16 %). The acidic fractions from the interior part contained polyphenolic traces which could influence the results. In preliminary experiments a phenolic fraction from *I. obliquus* had a repressive effect on NO production in J774.A1 cells pretreated with LPS and IFN-γ (data not shown), whereas others have found the same effect using polyphenolic extracts from various plants and fungi (Diaz, Jeong, Lee, Khoo, & Koyyalamudi, 2012). The absence of a clear dose-response relationship in the acidic fractions from the interior part (containing 4.2 - 9.7 % polyphenols) could therefore be due to the presence of polyphenols although it would not explain why the same fractions induced potent NO production at 100 μg/ml. The alkali-extracted IOI-A1 was not able to induce NO production in a dose-dependent manner, but demonstrated some activation even at the lowest concentration (1 μg/ml with IFN-γ) in the D2SC/1 cell line.



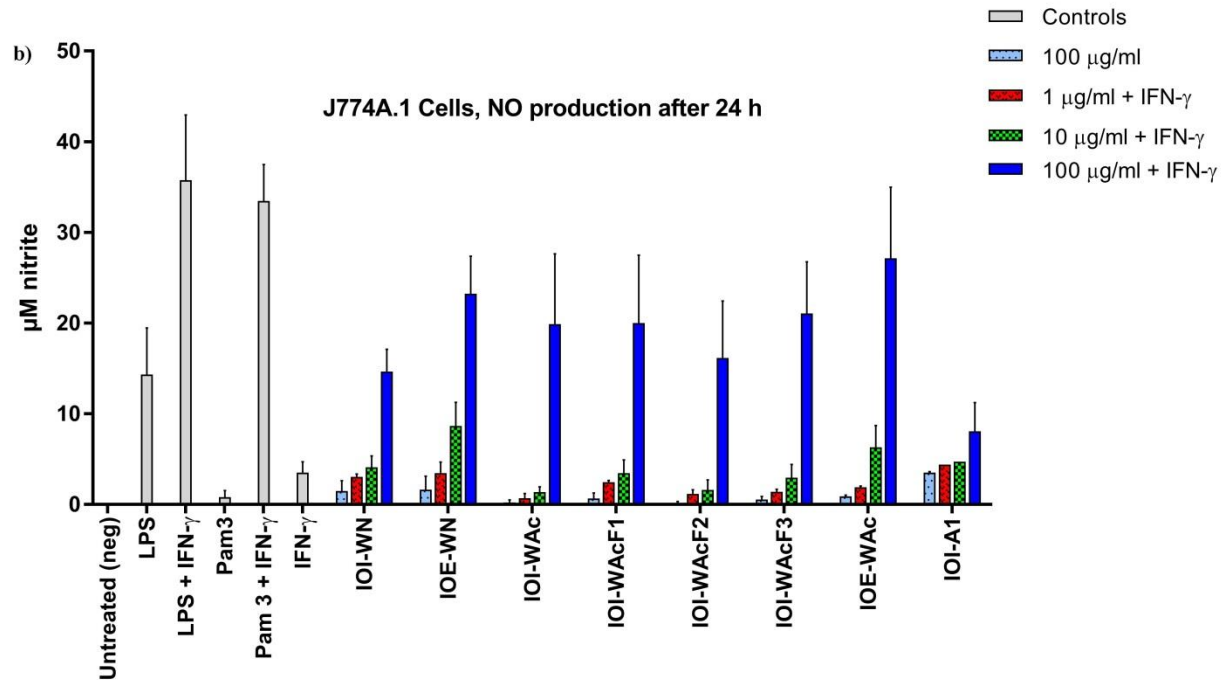


Fig. 4: NO production expressed as µM nitrite in the supernatant from D2SC/1 cells (a) and J774A.1 cells (b) after treatment with *I. obliquus* polysaccharide fractions with/without IFN-γ for 24 h. LPS (1 µg/ml), Pam3CSK4 (2 µg/ml) and IFN- γ (20 ng/ml) were used as positive controls alone and in combination. The untreated negative control had nitrite values ranging from 0.2-1 µM but this was set to zero for the presentation of the results.

3.11 Structure-function relationship

A clear structure-function relationship of the polysaccharides and their immunomodulating effects could not be established using the NO assay alone, but some suggestions were made. Since the neutral fractions seemed more active than the acidic ones, the GalA residues were probably not the most important for immune cell activation. Further, there was no obvious correlation between the amount of galactose and activity, since IOI-WAcF1 (~20% Gal/3-*O*-Me-Gal) and IOI-WAcF3 (~9% Gal/3-*O*-Me-Gal) gave similar results. Since the alkaline fraction IOI-A1 was not very active, it seems logical to conclude that soluble polysaccharides were better immune cell activators than particulate ones in this assay, in contrast with previous reports (Goodridge et al., 2011). However, the relative inactivity of IOI-A1 could also be due to its large size rather than solubility, or some other factors like the presence of xylose (14 %) in this fraction. The two neutral fractions IOI-WN and IOE-WN were very similar in terms of monosaccharide composition and types of linkages found, but IOE-WN seemed more promising in the NO assay. IOE-WN was slightly larger, having an M_w of 73 kDa vs. 60 kDa for IOI-WN. In addition, IOE-WN had a narrower weight-distribution according to SEC-MALLS which means a more homogenous polysaccharide fraction, and this could also explain some of the difference in activity. It seems likely that the (1→3)/(1→6)-β-Glc motif is, at least in part, responsible for the main effects observed in the experiments, but other factors such as three-dimensional shape could also be important. Fungal β-glucans of various kinds have previously been shown to be potent immune activators, and many different immune cell receptors have been associated with the recognition of such fungal structures (Erwig & Gow, 2016). Dectin-1 has for example been established as the main receptor for particulate (1→3)-β-glucans (Goodridge et al., 2011), but it could very well be that other receptors are involved in the recognition of soluble β-glucans such as the ones from *I. obliquus* described herein. One study found an increased production of TNF-α by RAW264.7 cells after treatment with a polysaccharide extract from *I. obliquus*, and this effect was sugggested to be via Toll-like receptor 2 (TLR2) interaction and not Dectin-1 or TLR-4 (Won et al., 2011). Although the effects in the mentioned study were measured indirectly, and the experiment was conducted using a crude polysaccharide extract obtained by ethanol precipiation only (probably containing substantial amount of melanin as well), the results encourage further testing of purified *I. obliquus* polysaccharides in similar assays.

4. Conclusion

To our knowledge, this study is the first comprehensive characterization of polysaccharides from the medicinal fungus *I. obliquus*. Water-extracted neutral and acidic polysaccharide fractions as well as an alkali-extracted polysaccharide fraction were isolated from the interior (IOI) and exterior (IOE) parts of *I. obliquus* sclerotia and characterized by chromatographic and spectroscopic methods. No major structural differences between fractions from IOI and IOE

were observed. The neutral polysaccharides IOI-WN (60 kDa) and IOE-WN (73 kDa) were heterogeneous and branched and consisted of a (1→3)-linked β-Glc backbone with (1→6)-linked kinks in the chain at approximately every fifth residue, with branches of (1→6)-linked β-Glc in addition to substantial amounts of (1→6)-linked α-Gal with 3-*O*-methylation at about every third Gal residue. The acidic polysaccharide fractions IOI-WAc, IOI-WAcF1, F2 and F3 and IOE-WAc (10-31 kDa) showed similar structural motifs as the neutral ones, differing mainly by the presence of (1→4)-linked α-GalA and α-GlcA. β-Xyl, α-Man and α-Rha were present in varying amounts in all fractions. The alkaline polysaccharide fraction IOI-A1 (>450 kDa) consisted mainly of (1→3)- and (1→6)-linked β-Glc and (1→4)-linked β-Xyl. GC-MS data from all fractions indicated that the polymer branching points were mainly found on β-Glc residues, thus these residues possibly connect the different parts of the polymers. However, such linkages could not be directly observed by NMR, and it should be attempted to clarify this in future studies. Several of the polysaccharide fractions showed promising results as immunomodulators as seen from the synergistic effect with IFN-γ on NO production in the murine cell lines J774.A1 and D2SC/1. The neutral polysaccharide fraction IOE-WN showed the most potent effect in the assay, increasing NO production at 10 μg/ml. It seems likely that the (1→3)/(1→6)-β-Glc motif within the polymer is important for bioactivity, but other factors such as three-dimensional shape could be equally important for the immunomodulating effect. The results presented herein encourage further exploration of the fine-structure, the structure-function relationship and bioactivity of polysaccharides from *I. obliquus.*

Acknowledgements

The authors would like to thank professor emerita Berit Smestad Paulsen for valuable discussions and interpretation of the results in general, Hoai Thi Nguyen Aas for help with the LPS determination assay, and Ellen Hanne Cohen for help with the Dionex column for the periodate oxidation/Smith degradation experiment. The work was partially funded by Novo Nordisk Foundation: Biotechnology-based synthesis and production research, program grant 5371. The 800 Mhz NMR spectra for most samples were recorded on the NMR Center – DTU, supported by the Villum Foundation. The work was partly supported by the Research Council of Norway through the Norwegian NMR platform, NNP (226244/F50). Finally the authors would like to thank the Norwegian PhD School of Pharmacy for funding a 3-week research stay at DTU for the NMR experiments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in Appendix A in the online version.

References

Agrawal, K. A. (1992). NMR spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry,* 31(10), 3307-3330.

Arana, D. M., Prieto, D., Roman, E., Nombela, C., Alonso-Monge, R., & Pla, J. (2009). The role of the cell wall in fungal pathogenesis. *Microbial Biotechnology,* 2(3), 308-320.

Austarheim, I., Christensen, B. E., Hegna, I. K., Petersen, B. O., Duus, J. O., Bye, R., . . . Paulsen, B. S. (2012). Chemical and biological characterization of pectin-like polysaccharides from the bark of the Malian medicinal tree Cola cordifolia. *Carbohydrate Polymers,* 89(1), 259-268.

Bouchard, A., Hofland, G. W., & Witkamp, G. J. (2007). Properties of Sugar, Polyol, and Polysaccharide Water-Ethanol Solutions. *Journal of Chemical & Engineering Data,* 52, 1838-1842.

Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry,* 72, 248-254.

Brennan, C. S., & Cleary, L. J. (2005). The potential use of cereal (1→3,1→4)- β -d-glucans as functional food ingredients. *Journal of Cereal Science,* 42(1), 1-13.

Carbonero, E. R., Gracher, A. H., Rosa, M. C., Torri, G., Sassaki, G. L., Gorin, P. A., & Iacomini, M. (2008). Unusual partially 3-O-methylated alpha-galactan from mushrooms of the genus Pleurotus. *Phytochemistry,* 69(1), 252-257.

Cha, J. Y., Lee, S. Y., Lee, S. Y., & Chun, K. W. (2011). Basidiocarp formation by Inonotus obliquus on a living paper birch tree. *Forest Pathology,* 41(2), 163-164.

Chambers, R. E., & Clamp, J. R. (1971). An Assessment of Methanolysis and Other Factors Used in the Analysis of Carbohydrate-Containing Materials. *Biochemical Journal,* 125, 1009-1018.

Chen, Y., Huang, Y., Cui, Z., & Liu, J. (2015). Purification, characterization and biological activity of a novel polysaccharide from Inonotus obliquus. *International Journal of Biological Macromolecules,* 79, 587-594.

Curren Smith, E. W. (2015). Macrophage Polarization and Its Role in Cancer. *Journal of Clinical and Cellular Immunology,* 06(04).

de Santana-Filho, A. P., Noleto, G. R., Gorin, P. A. J., de Souza, L. M., Iacomini, M., & Sassaki, G. L. (2012). GC–MS detection and quantification of lipopolysaccharides in polysaccharides through 3-O-acetyl fatty acid methyl esters. *Carbohydrate Polymers,* 87(4), 2730-2734.

Diaz, P., Jeong, S. C., Lee, S., Khoo, C., & Koyyalamudi, S. R. (2012). Antioxidant and anti-inflammatory activities of selected medicinal plants and fungi containing phenolic and flavonoid compounds. *Chinese Medicine,* 7(26).

Du, B., Lin, C., Bian, Z., & Xu, B. (2015). An insight into anti-inflammatory effects of fungal beta-glucans. *Trends in Food Science & Technology,* 41(1), 49-59.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Analytical Chemistry,* 28, 350-356.

Duus, J. Ø., Gotfredsen, C. H., & Bock, K. (2000). Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations. *Chemical Reviews,* 100, 4589-4614.

Eisenman, H. C., & Casadevall, A. (2012). Synthesis and assembly of fungal melanin. *Applied Microbiology and Biotechnology,* 93(3), 931-940.

Erwig, L. P., & Gow, N. A. (2016). Interactions of fungal pathogens with phagocytes. *Nature Reviews Microbiology,* 14(3), 163-176.

Fan, L., Ding, S., Ai, L., & Deng, K. (2012). Antitumor and immunomodulatory activity of water-soluble polysaccharide from Inonotus obliquus. *Carbohydrate Polymers,* 90(2), 870-874.

Fraser, I. P., Stuart, L., & Ezekowitz, R. A. (2004). TLR-independent pattern recognition receptors and anti-inflammatory mechanisms. *Journal of Endotoxin Research,* 10(2), 120-124.

Glamoclija, J., Ciric, A., Nikolic, M., Fernandes, A., Barros, L., Calhelha, R. C., . . . van Griensven, L. J. (2015). Chemical characterization and biological activity of Chaga (Inonotus obliquus), a medicinal "mushroom". *Journal of Ethnopharmacology,* 162, 323-332.

Goodridge, H. S., Reyes, C. N., Becker, C. A., Katsumoto, T. R., Ma, J., Wolf, A. J., . . . Underhill, D. M. (2011). Activation of the innate immune receptor Dectin-1 upon formation of a 'phagocytic synapse'. *Nature,* 472(7344), 471-475.

Hickok, J. R., & Thomas, D. D. (2010). Nitric Oxide and Cancer Therapy: The Emperor has NO Clothes. *Current Pharmaceutical Design,* 16(4), 381-391.

Hu, Y., Sheng, Y., Yu, M., Li, K., Ren, G., Xu, X., & Qu, J. (2016). Antioxidant activity of Inonotus obliquus polysaccharide and its amelioration for chronic pancreatitis in mice. *International Journal of Biological Macromolecules,* 87, 348-356.

Hwang, B. S., Lee, I. K., & Yun, B. S. (2016). Phenolic compounds from the fungus Inonotus obliquus and their antioxidant properties. *The Journal of Antibiotics,* 69(2), 108-110.

I., C., & Kerek, F. (1984). A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research,* 131, 209-217.

Ina, K., Kataoka, T., & Ando, T. (2013). The Use of Lentinan for Treating Gastric Cancer. *Anti-Cancer Agents in Medicinal Chemistry,* 13, 681-688.

Jansson, P. E., Stenutz, R., & Widmalm, G. (2006). Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program CASPER. *Carbohydrate Research,* 341(8), 1003-1010.

Kamerling, J. P., & Gerwig, G. J. (2007). *Strategies for the Structural Analysis of Carbohydrates*: Elsevier Ltd.

Kapoor, S. (2014). Lentinan: clinical benefit in the management of systemic malignancies. *Surgery Today,* 44(7), 1389.

Kim, Y. O., Han, S. B., Lee, H. W., Ahn, H. J., Yoon, Y. D., Jung, J. K., . . . Shin, C. S. (2005). Immuno-stimulating effect of the endo-polysaccharide produced by submerged culture of Inonotus obliquus. *Life Sciences,* 77(19), 2438-2456.

Kim, Y. O., Park, H. W., Kim, J. H., Lee, J. Y., Moon, S. H., & Shin, C. S. (2006). Anti-cancer effect and structural characterization of endo-polysaccharide from cultivated mycelia of Inonotus obliquus. *Life Sciences,* 79(1), 72-80.

Kukulyanskaya, T. A., Kurchenko, N. V., Kurchenko, V. P., & Babitskaya, V. G. (2002). Physicochemical Properties of Melanins Produced by the Sterile Form of Inonotus obliquus ("Chagi") in Natural and Cultivated Fungus. *Applied Biochemistry and Microbiology,* 38(1), 58-61.

Lee, K. R., Lee, J. S., Song, J. E., Ha, S. J., & Hong, E. K. (2014). Inonotus obliquus-derived polysaccharide inhibits the migration and invasion of human non-small cell lung carcinoma cells via suppression of MMP-2 and MMP-9. *International Journal of Oncology,* 45(6), 2533-2540.

Lieder, R., Petersen, P. H., & Sigurjonsson, O. E. (2013). Endotoxins-the invisible companion in biomaterials research. *Tissue Engineering Part B: Reviews,* 19(5), 391-402.

Liu, J., Willför, S., & Xu, C. (2015). A review of bioactive plant polysaccharides: Biological activities, functionalization, and biomedical applications. *Bioactive Carbohydrates and Dietary Fibre,* 5(1), 31-61.

Lu, C. C., Hsu, Y. J., Chang, C. J., Lin, C. S., Martel, J., Ojcius, D. M., . . . Young, J. D. (2016). Immunomodulatory properties of medicinal mushrooms: differential effects of water and ethanol extracts on NK cell-mediated cytotoxicity. *Innate Immunity,* 22(7), 522-533.

Lundborg, M., & Widmalm, G. (2011). Structural analysis of glycans by NMR chemical shift prediction. *Analytical Chemistry,* 83(5), 1514-1517.

Ma, L., Chen, H., Dong, P., & Lu, X. (2013). Anti-inflammatory and anticancer activities of extracts and compounds from the mushroom Inonotus obliquus. *Food Chemistry,* 139(1-4), 503-508.

975 Manna, D. K., Maity, P., Nandi, A. K., Pattanayak, M., Panda, B. C., Mandal, A. K., . . . Islam, S. S. (2017).
976 Structural elucidation and immunostimulating property of a novel polysaccharide extracted
977 from an edible mushroom Lentinus fusipes. *Carbohydrate Polymers,* 157, 1657-1665.
978 Min-Woong, L., Hyeon-Hur, Kwang-Choon, C., Tae-Soo, L., Kang-Hyeon, K., & Jankovsky, L. (2008).
979 Introduction to Distribution and Ecology of Sterile Conks of Inonotus obliquus. *Mycobiology,*
980 36(4), 199-202.
981 Nyman, A. A., Aachmann, F. L., Rise, F., Ballance, S., & Samuelsen, A. B. (2016). Structural
982 characterization of a branched (1-->6)-alpha-mannan and beta-glucans isolated from the fruiting
983 bodies of Cantharellus cibarius. *Carbohydrate Polymers,* 146, 197-207.
984 Pettolino, F. A., Walsh, C., Fincher, G. B., & Bacic, A. (2012). Determining the polysaccharide composition
985 of plant cell walls. *Nature Protocols,* 7(9), 1590-1607.
986 Prados-Rosales, R., Toriola, S., Nakouzi, A., Chatterjee, S., Stark, R., Gerfen, G., . . . Casadevall, A. (2015).
987 Structural Characterization of Melanin Pigments from Commercial Preparations of the Edible
988 Mushroom Auricularia auricula. *Journal of Agricultural and Food Chemistry,* 63(33), 7326-7332.
989 Qiao, Y., Giannopoulou, E. G., Chan, C. H., Park, S. H., Gong, S., Chen, J., . . . Ivashkiv, L. B. (2013).
990 Synergistic activation of inflammatory cytokine genes by interferon-gamma-induced chromatin
991 remodeling and toll-like receptor signaling. *Immunity,* 39(3), 454-469.
992 Rhee, S. J., Cho, S. Y., Kim, K. M., Cha, D.-S., & Park, H.-J. (2008). A comparative study of analytical
993 methods for alkali-soluble β-glucan in medicinal mushroom, Chaga (Inonotus obliquus). *LWT-*
994 *Food Science and Technology,* 41(3), 545-549.
995 Saar, M. (1991). Fungi in Khanty Folk Medicine. *Journal of Ethnopharmacology*, 175-179.
996 Schwartz, B., & Hadar, Y. (2014). Possible mechanisms of action of mushroom-derived glucans on
997 inflammatory bowel disease and associated cancer. *Annals of Translational Medicine,* 2(2), 19.
998 Shashkina, M. Y., Shashkin, P. N., & Sergeev, A. V. (2006). Chemical and medicobiological properties of
999 chaga (review). *Pharmaceutical Chemistry Journal,* 40(10), 560-568.
1000 Shikov, A. N., Pozharitskaya, O. N., Makarov, V. G., Wagner, H., Verpoorte, R., & Heinrich, M. (2014).
1001 Medicinal plants of the Russian Pharmacopoeia; their history and applications. *Journal of*
1002 *Ethnopharmacology,* 154(3), 481-536.
1003 Smiderle, F. R., Olsen, L. M., Carbonero, E. R., Marcon, R., Baggio, C. H., Freitas, C. S., . . . Iacomini, M.
1004 (2008). A 3-O-methylated mannogalactan from Pleurotus pulmonarius: structure and
1005 antinociceptive effect. *Phytochemistry,* 69(15), 2731-2736.
1006 Staudacher, E. (2012). Methylation – an uncommon modification of glycans. *The Journal of Biological*
1007 *Chemistry,* 393(8), 675-685.
1008 Swain, T., & Hillis, W. E. (1959). The phenolic constituents of Prunus domestica. I. - The quantitative
1009 analysis of phenolic constituents. *Journal of the Science of Food and Agriculture,* 10(1), 63-68.
1010 Totemeyer, S., Sheppard, M., Lloyd, A., Roper, D., Dowson, C., Underhill, D., . . . Bryant, C. (2006). IFN-
1011 γ Enhances Production of Nitric Oxide from Macrophages via a Mechanism That Depends on
1012 Nucleotide Oligomerization Domain-2. *The Journal of Immunology,* 176(8), 4804-4810.
1013 Wei, S., & Van Griensven, L. J. (2008). Pro- and Antioxidative Properties of Medicinal Mushroom
1014 Extracts. *International Journal of Medicinal Mushrooms,* 10(4), 315-324.
1015 Whistler, R. L., & Wolfrom, M. L. (1963). *Reactions of Carbohydrates*. **New York**: Academic Press Inc.
1016 Wiercigroch, E., Szafraniec, E., Czamara, K., Pacia, M. Z., Majzner, K., Kochan, K., . . . Malek, K. (2017).
1017 Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochimica Acta Part A:*
1018 *Molecular and Biomolecular Spectroscopy,* 185, 317-335.
1019 Won, D. P., Lee, J. S., Kwon, D. S., Lee, K. E., Shin, W. C., & Hong, E. K. (2011). Immunostimulating activity
1020 by polysaccharides isolated from fruiting body of Inonotus obliquus. *Molecules and Cells,* 31(2),
1021 165-173.

1022 Xu, X., Quan, L., & Shen, M. (2015). Effect of chemicals on production, composition and antioxidant
1023 activity of polysaccharides of Inonotus obliquus. *International Journal of Biological*
1024 *Macromolecules,* 77, 143-150.
1025 Yang, Y., Zhang, J., Liu, Y., Tang, Q., Zhao, Z., & Xia, W. (2007). Structural elucidation of a 3-O-methyl-D-
1026 galactose-containing neutral polysaccharide from the fruiting bodies of Phellinus igniarius.
1027 *Carbohydrate Research,* 342(8), 1063-1070.
1028 Yapo, B. M. (2011). Pectic substances: From simple pectic polysaccharides to complex pectins—A new
1029 hypothetical model. *Carbohydrate Polymers,* 86(2), 373-385.
1030 Youn, M. J., Kim, J. K., Park, S. Y., Kim, Y., Park, C., Kim, E. S., . . . Park, R. (2009). Potential anticancer
1031 properties of the water extract of Inonotus obliquus by induction of apoptosis in melanoma B16-
1032 F10 cells. *Journal of Ethnopharmacology,* 121(2), 221-228.
1033 Zevenhuizen, L. P. T. M. (1997). Succinoglycan and galactoglucan. *Carbohydrate Polymers,* 33, 139-144.
1034 Zhang, M., Cui, S. W., Cheung, P. C. K., & Wang, Q. (2007). Antitumor polysaccharides from mushrooms:
1035 a review on their isolation process, structural characteristics and antitumor activity. *Trends in*
1036 *Food Science & Technology,* 18(1), 4-19.
1037 Zhao, F., Mai, Q., Ma, J., Xu, M., Wang, X., Cui, T., . . . Han, G. (2015). Triterpenoids from Inonotus
1038 obliquus and their antitumor activities. *Fitoterapia,* 101, 34-40.
1039 Zheng, W., Miao, K., Liu, Y., Zhao, Y., Zhang, M., Pan, S., & Dai, Y. (2010). Chemical diversity of
1040 biologically active metabolites in the sclerotia of Inonotus obliquus and submerged culture
1041 strategies for up-regulating their production. *Applied Microbiology and Biotechnology,* 87(4),
1042 1237-1254.

1043

Structural characterization of bioactive heteropolysaccharides from the fungus *Inonotus obliquus.*

Christian Winther Wold[a], Christian Kjeldsen[b], Alexandre Corthay[c], Frode Rise[e], Bjørn Erik Christensen[d], Jens Øllgaard Duus[b], Kari Tvete Inngjerdingen[a]

[a] Department of Pharmaceutical Chemistry, School of Pharmacy, University of Oslo, Oslo, Norway
b Department of Chemistry, Technical University of Denmark, Copenhagen, Denmark
[c]Tumor Immunology Lab, Department of Pathology, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[d] Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Norway
[e]Department of Chemistry, University of Oslo, Oslo, Norway
e-mail: c.w.wold@farmasi.uio.no

Supplementary data

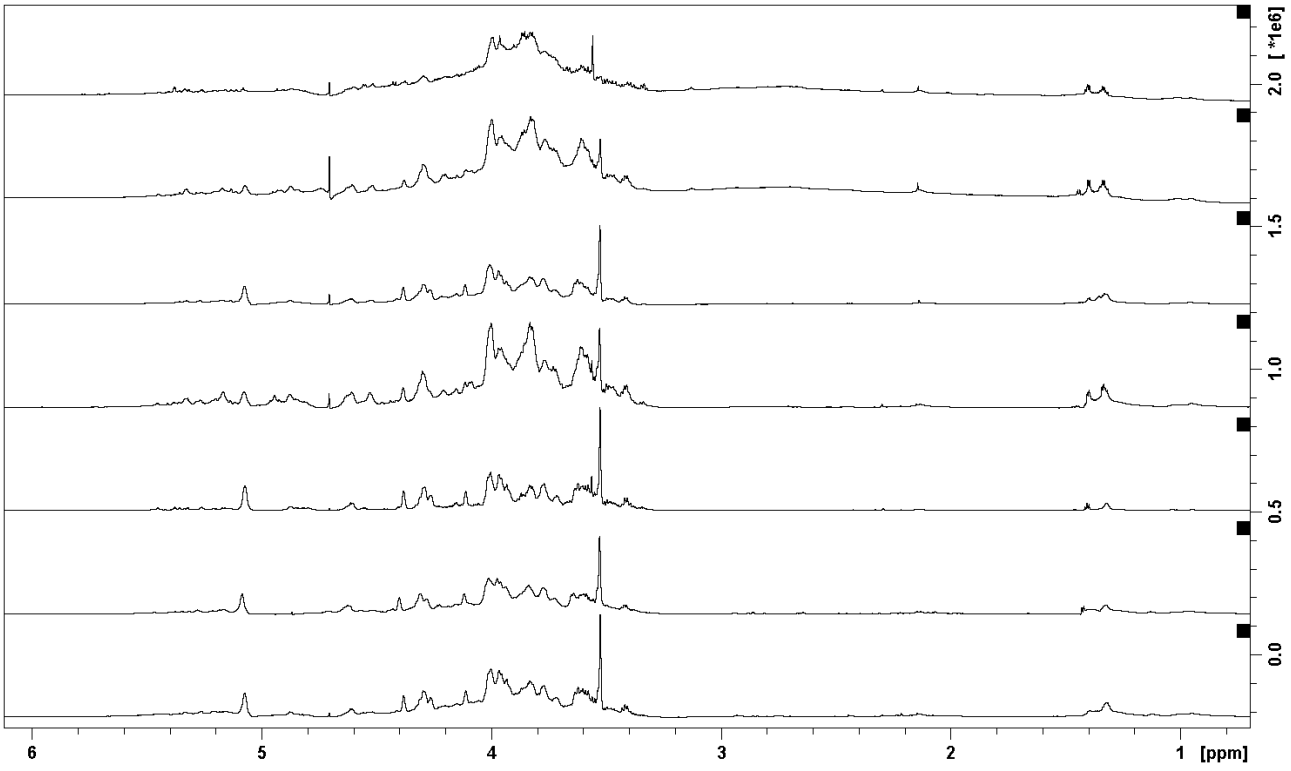


Fig. S1. ^{1}H spectra of the following samples from bottom to top: IOI-WAcF3, IOI-WAcF2, IOI-WAcF1, IOI-WAc, IOI-WN, IOE-WAc, IOE-WN

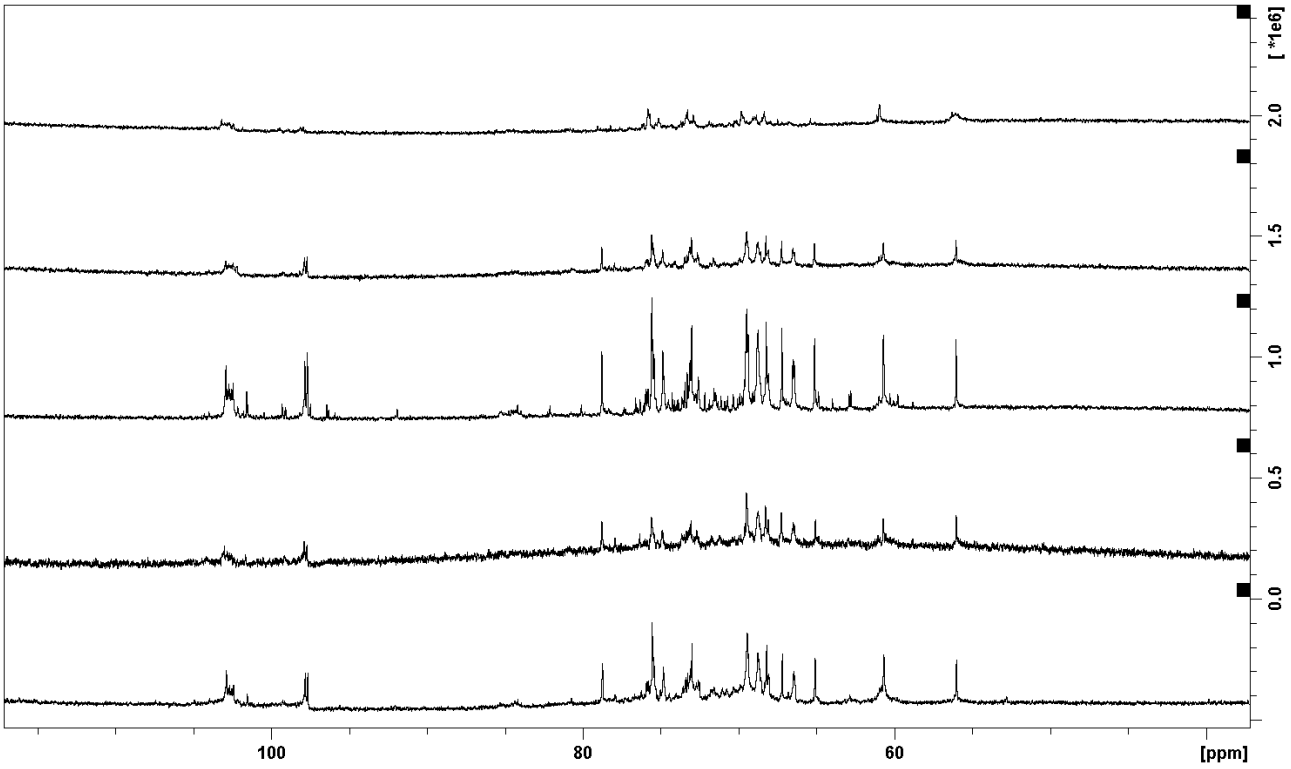


Fig. S2. ^{13}C spectra of the following samples from top to bottom: IOI-WAcF2, IOI-WAcF1, IOI-WN, IOE-WAc, IOE-WN

The following is an example of each type of 2D spectrum used for the assignment: DQF-COSY, TOCSY, NOESY, HSQC, HSQC-TOCSY and HMBC for the IOE-WN sample.

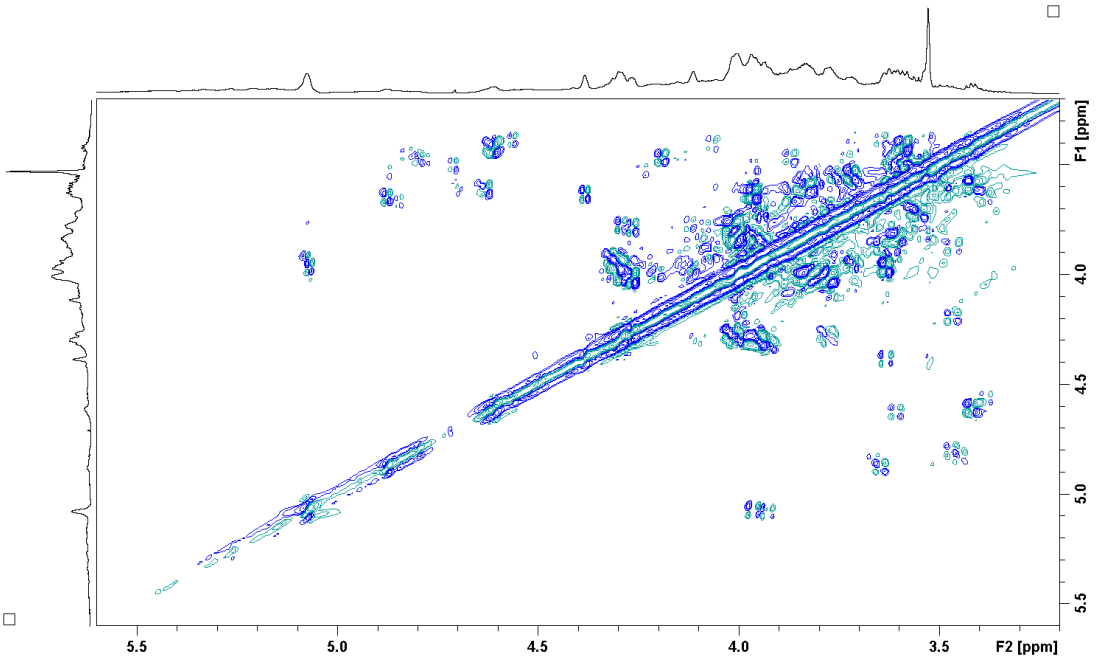


Fig. S3. DQF-COSY spectrum from the IOE-WN sample

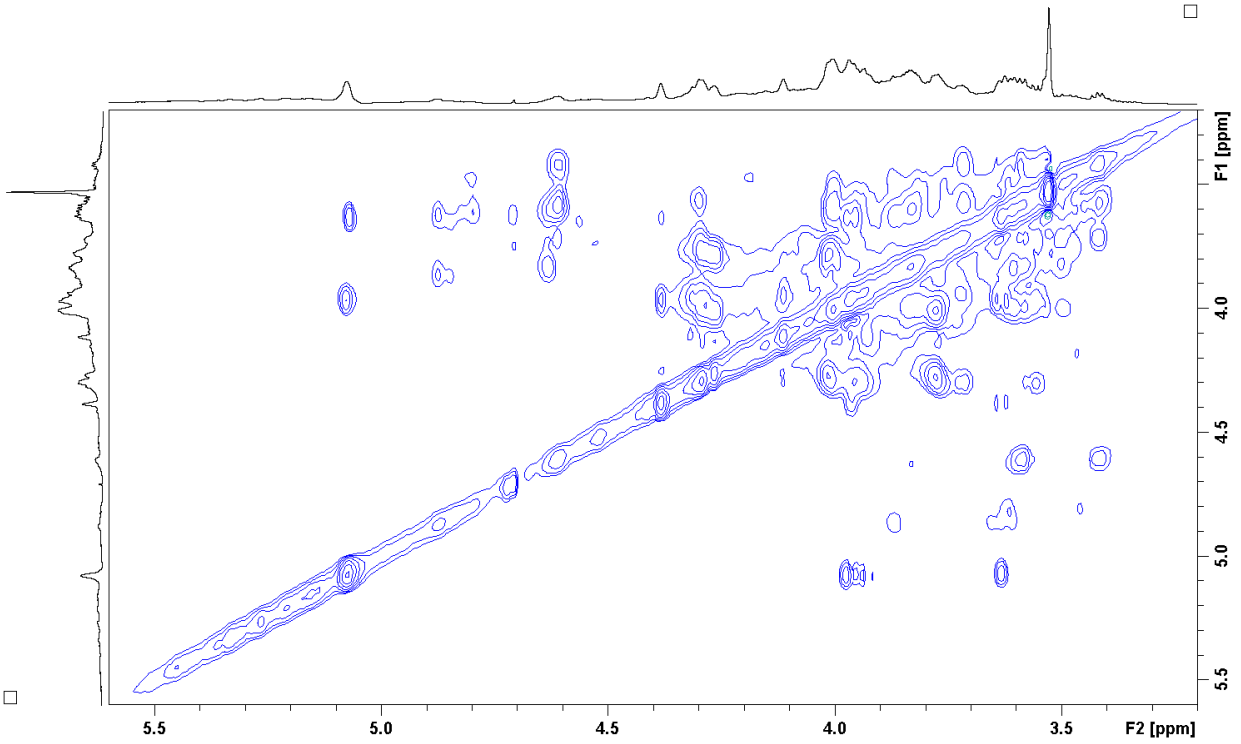


Fig. S4. TOCSY spectrum from the IOE-WN sample

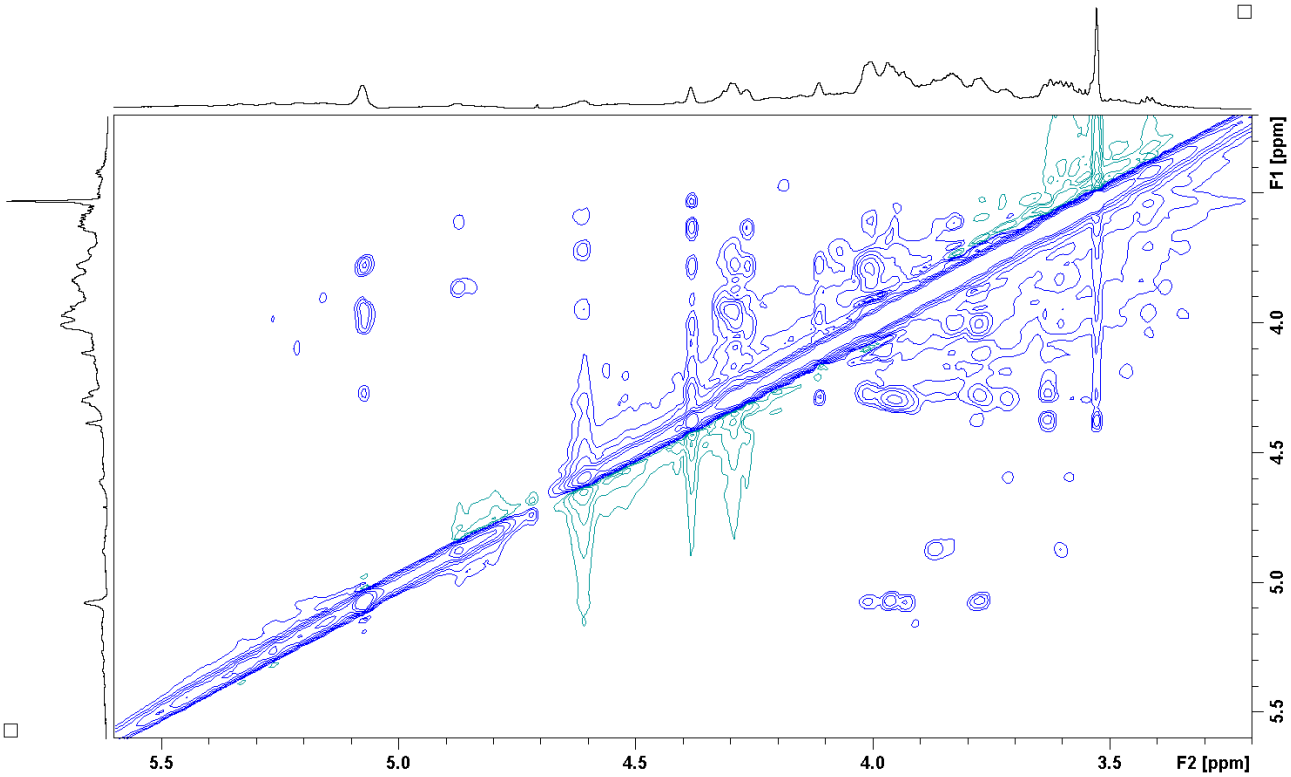


Fig. S5. NOESY spectrum from the IOE-WN sample

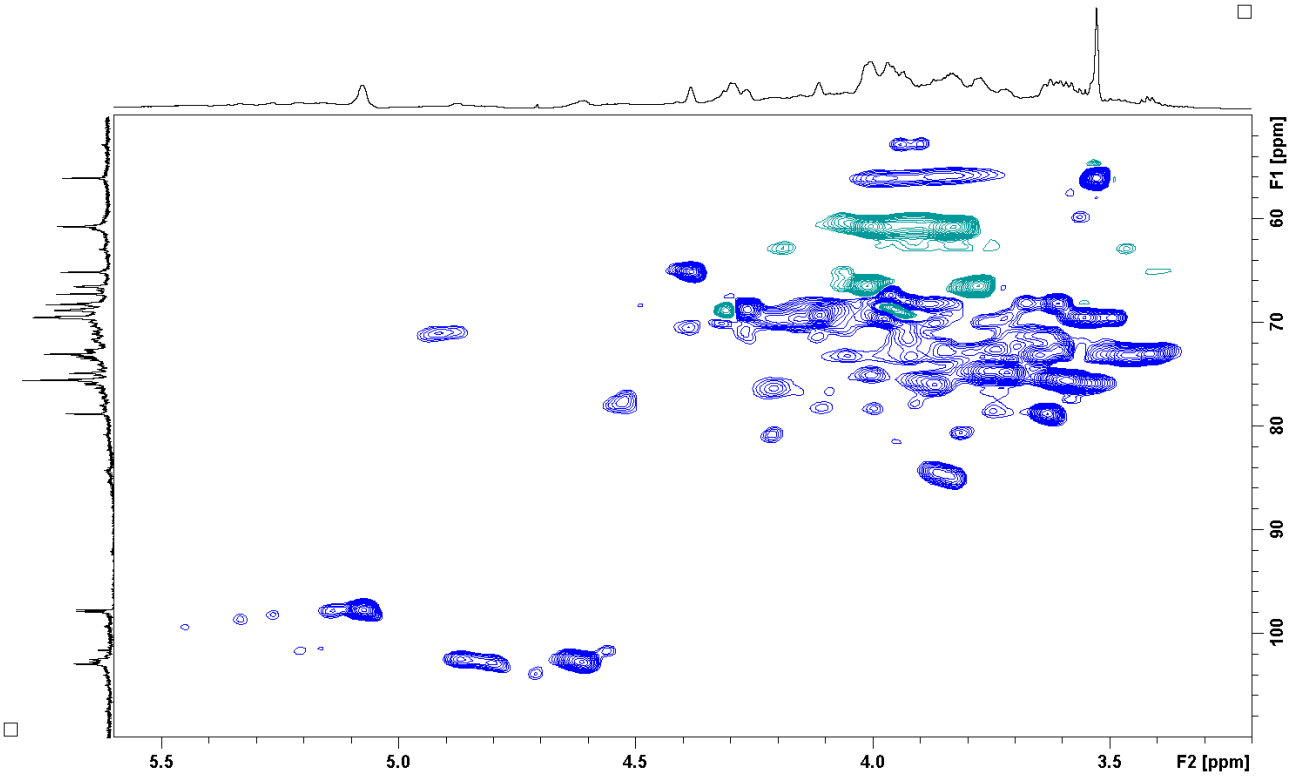


Fig. S6. HSQC spectrum from the IOE-WN sample

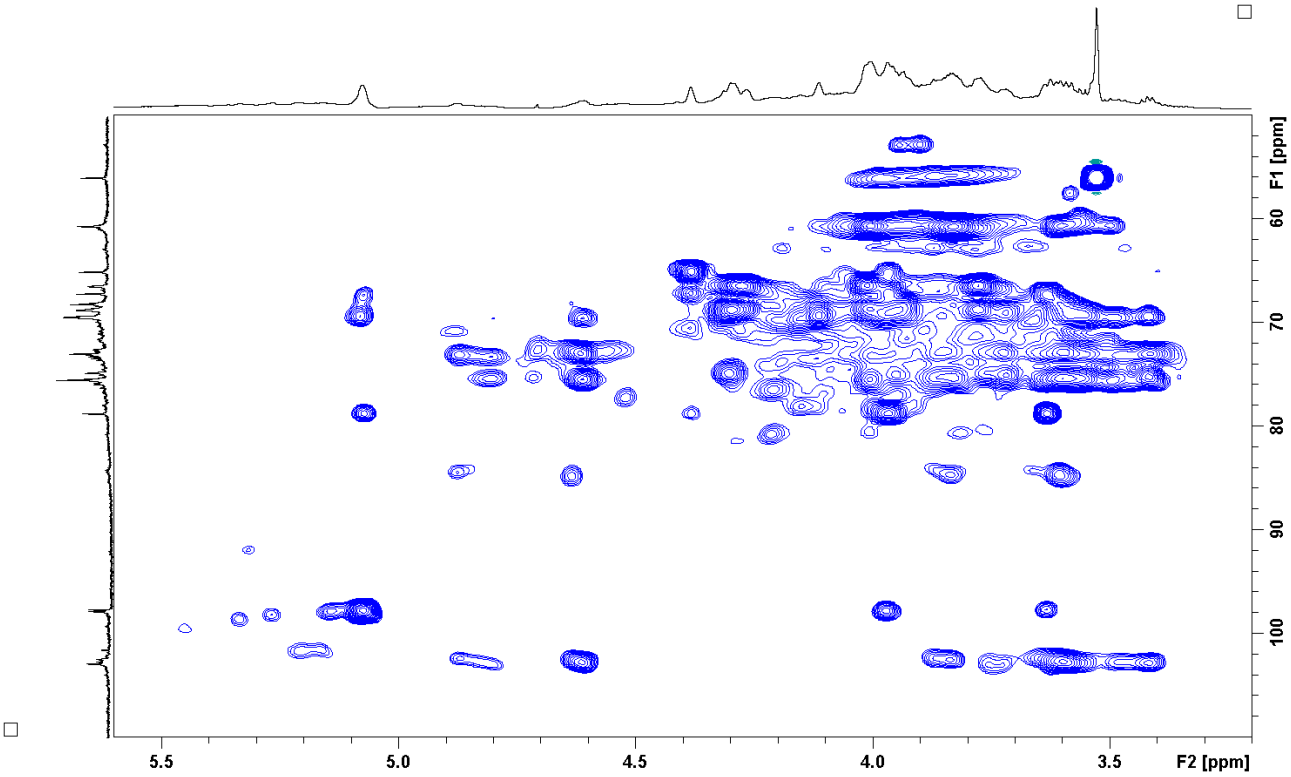


Fig. S7. HSQC-TOCSY spectrum from the IOE-WN sample

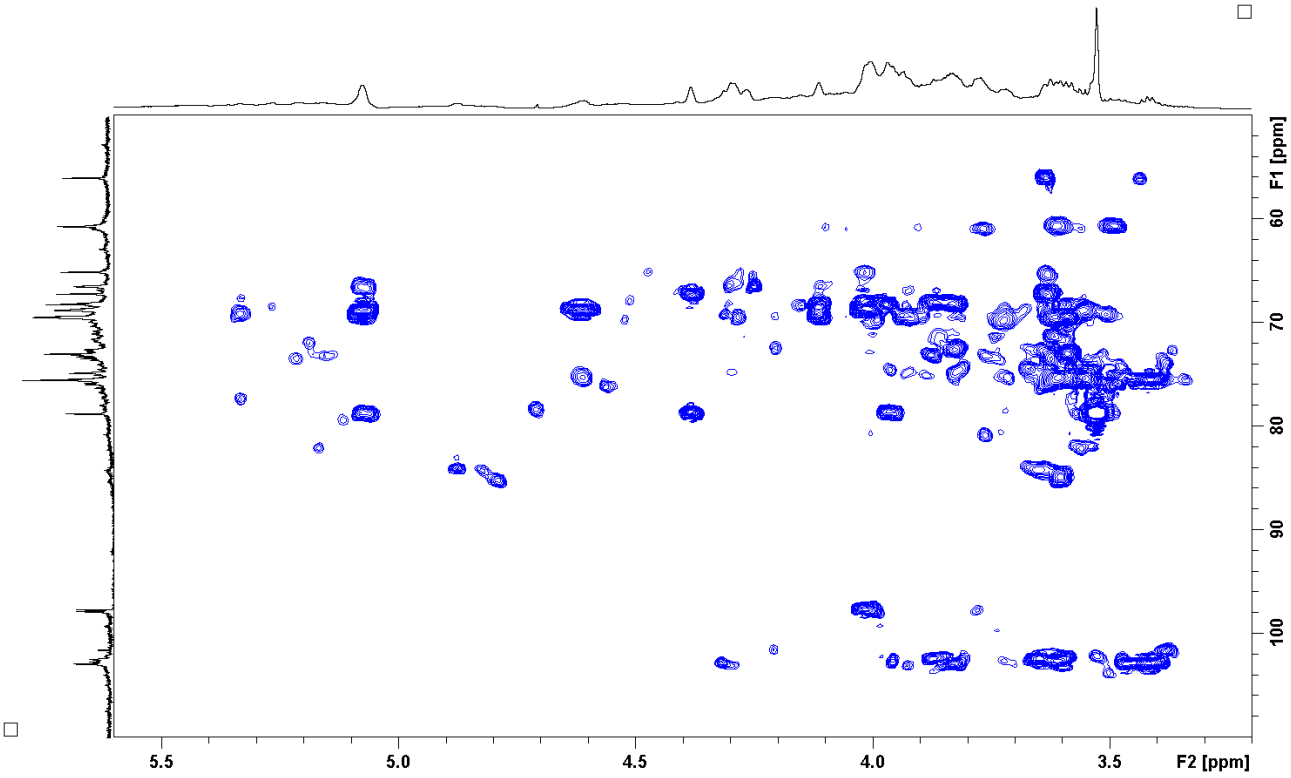


Fig. S8. HMBC spectrum from the IOE-WN sample

The following is the HSQC spectra of the remaining samples, for comparison

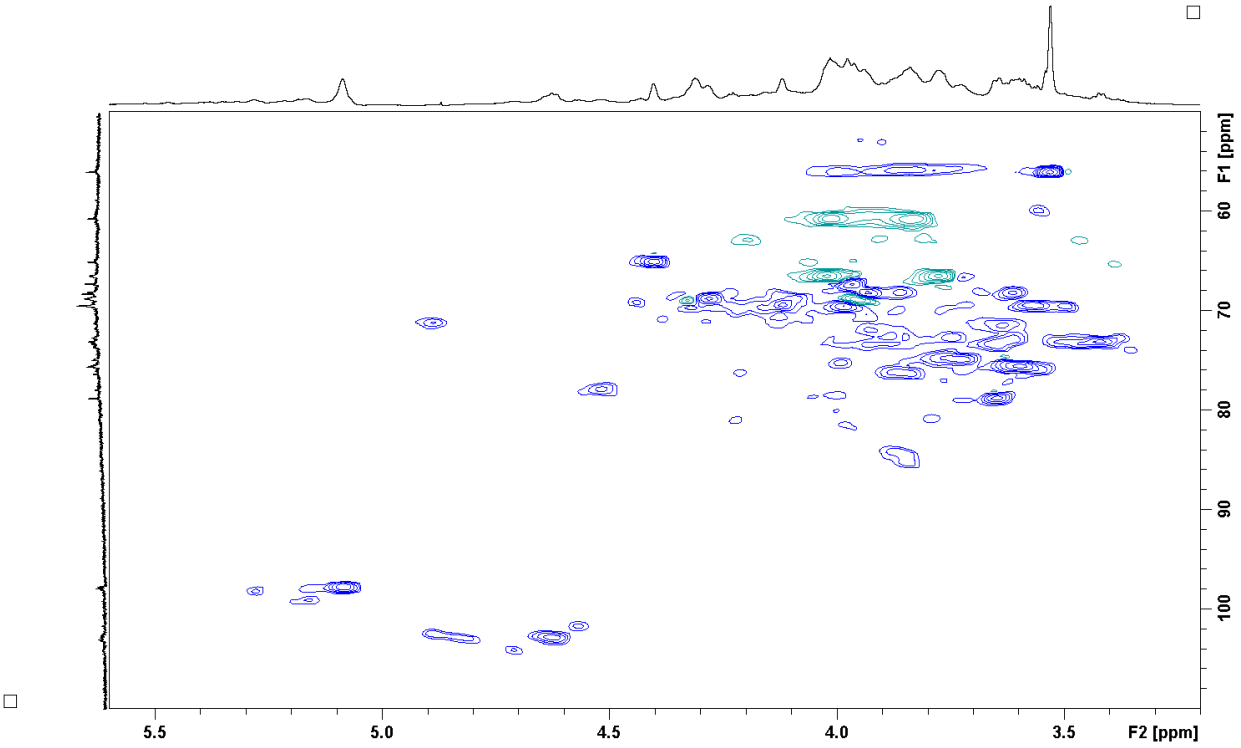


Fig. S9. HSQC spectrum from the IOE-WAc sample

.

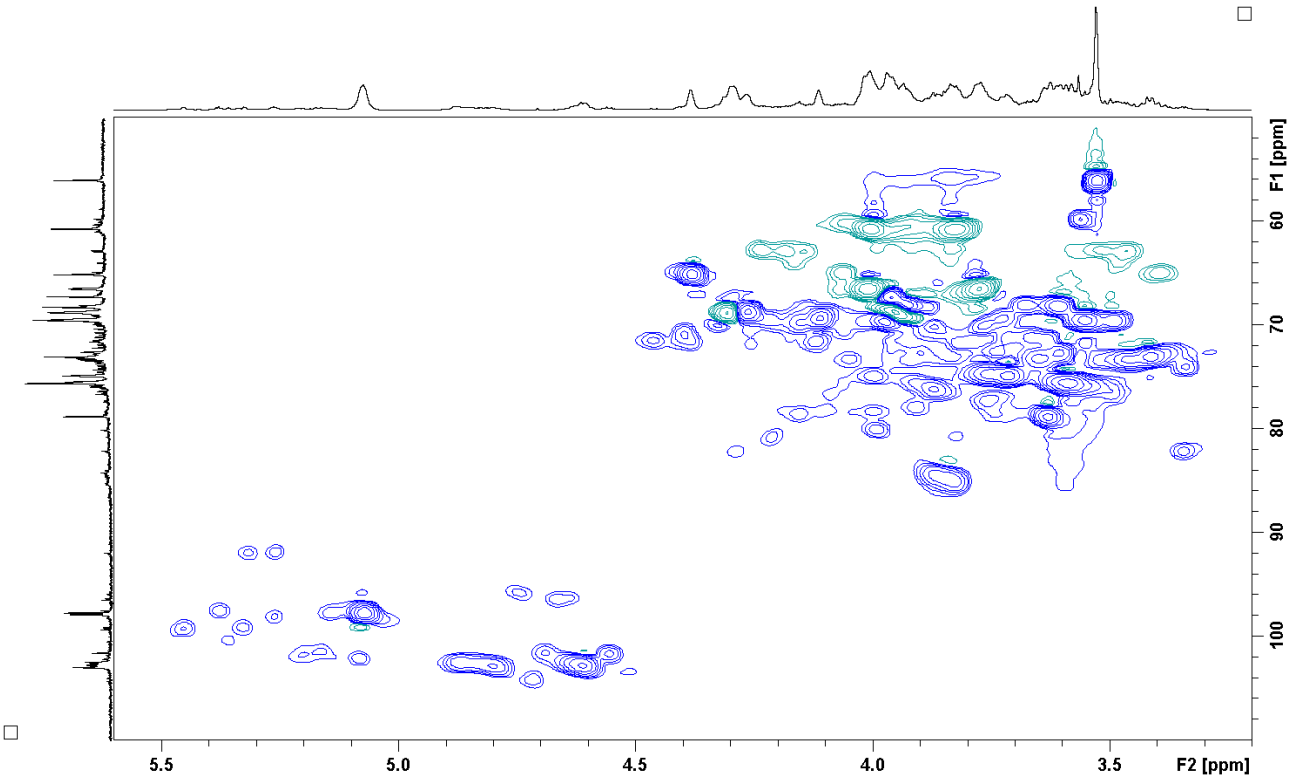


Fig. S10. HSQC spectrum from the IOI-WN sample

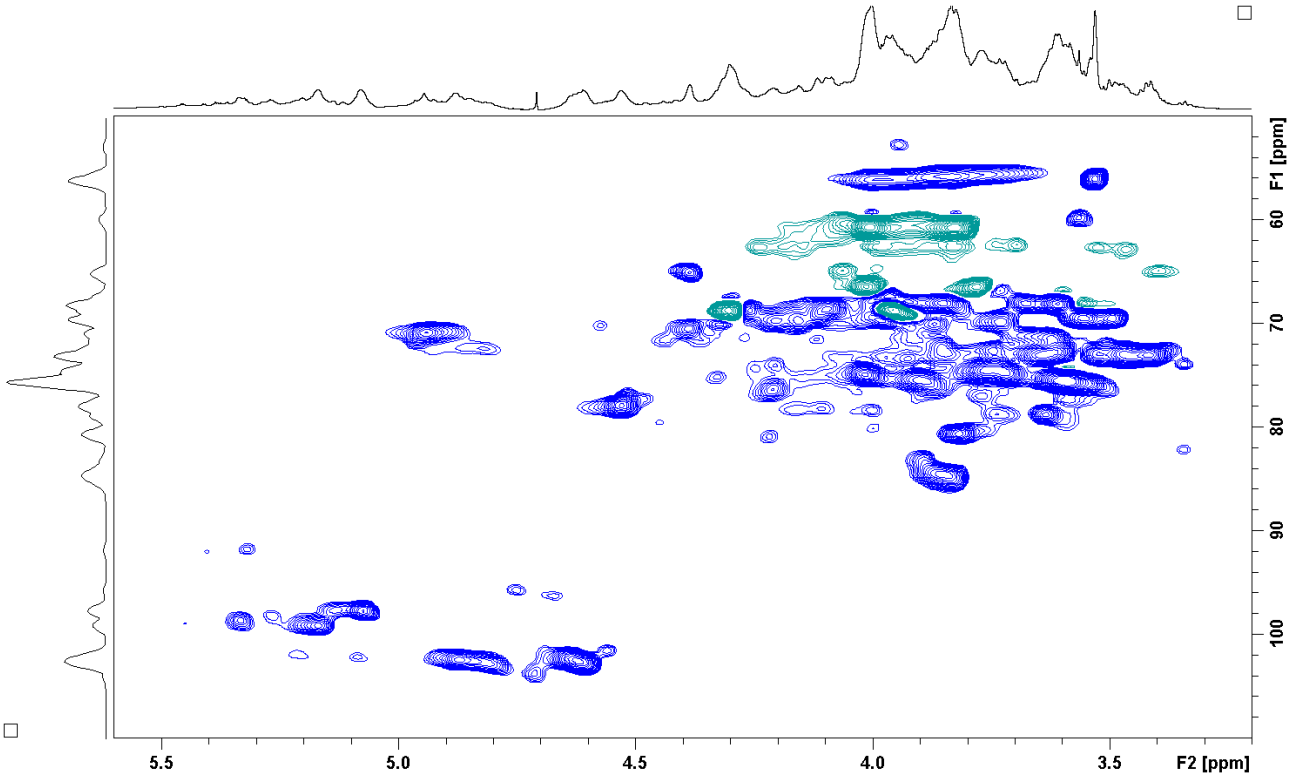


Fig. S11. HSQC spectrum from the IOI-WAc sample

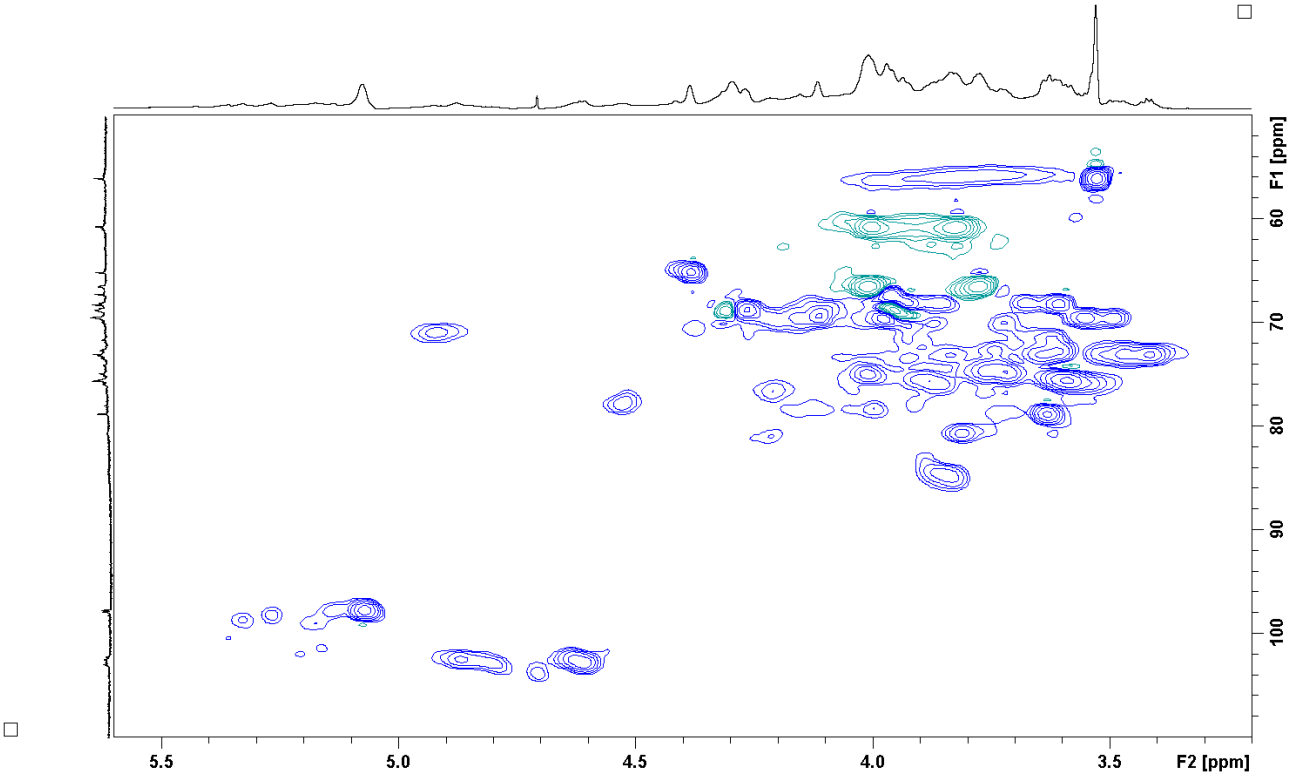


Fig. S12. HSQC spectrum from the IOI-WAcF1 sample

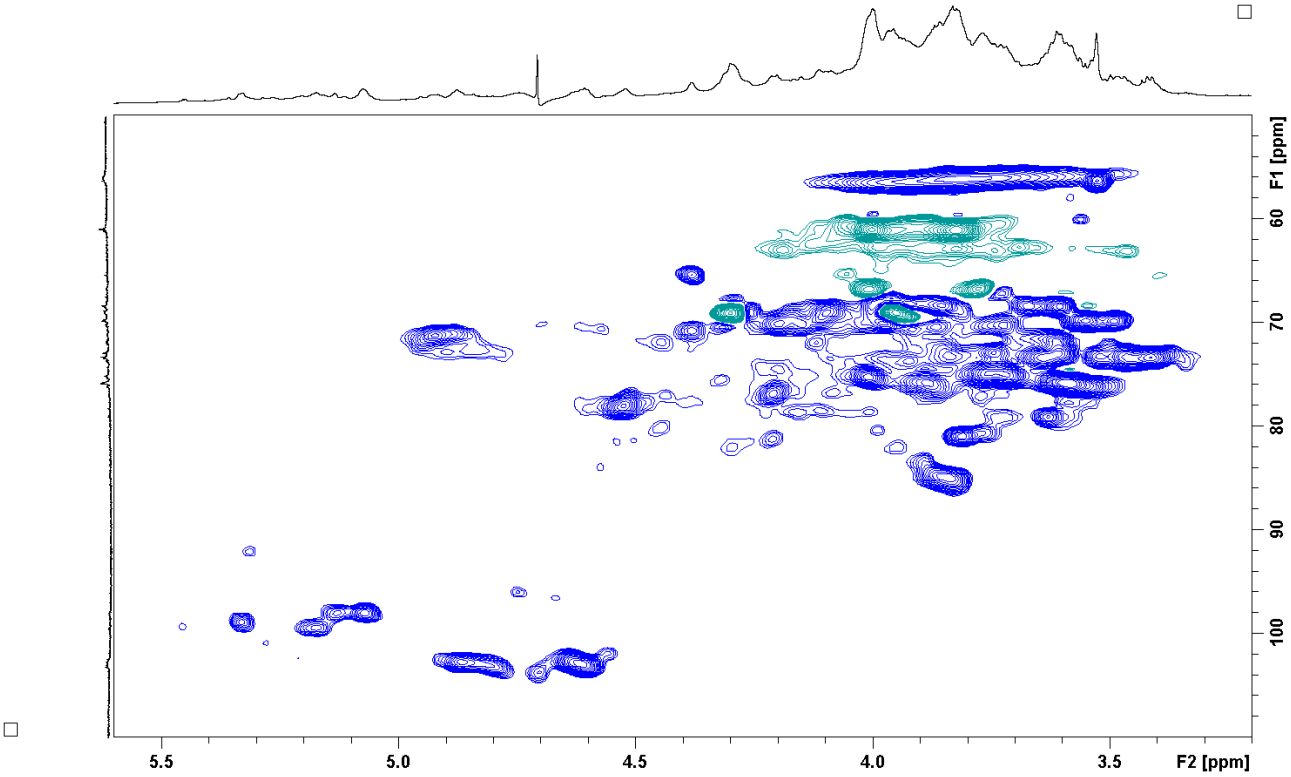


Fig. S13. HSQC spectrum from the IOI-WAcF2 sample

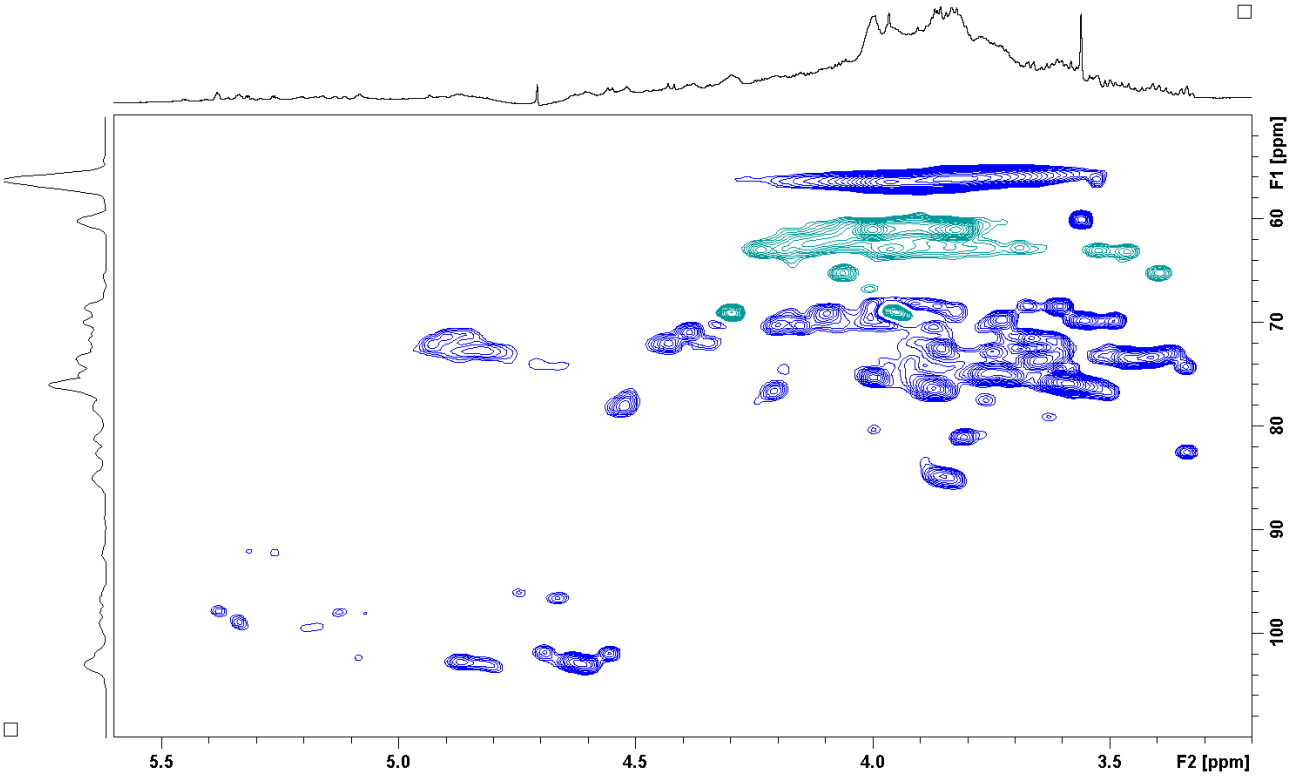


Fig. S14. HSQC spectrum from the IOI-WAcF3 sample

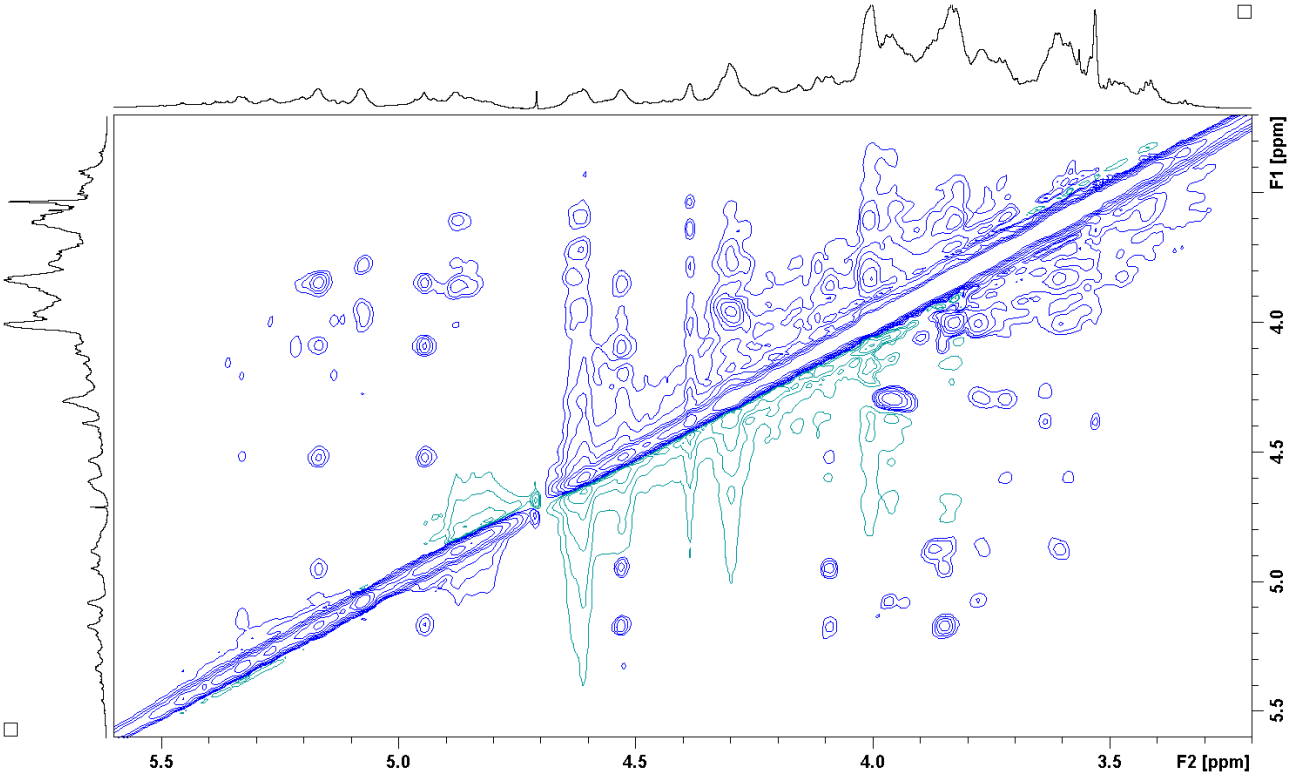


Fig. S15. NOESY spectrum from the IOI-WAc sample

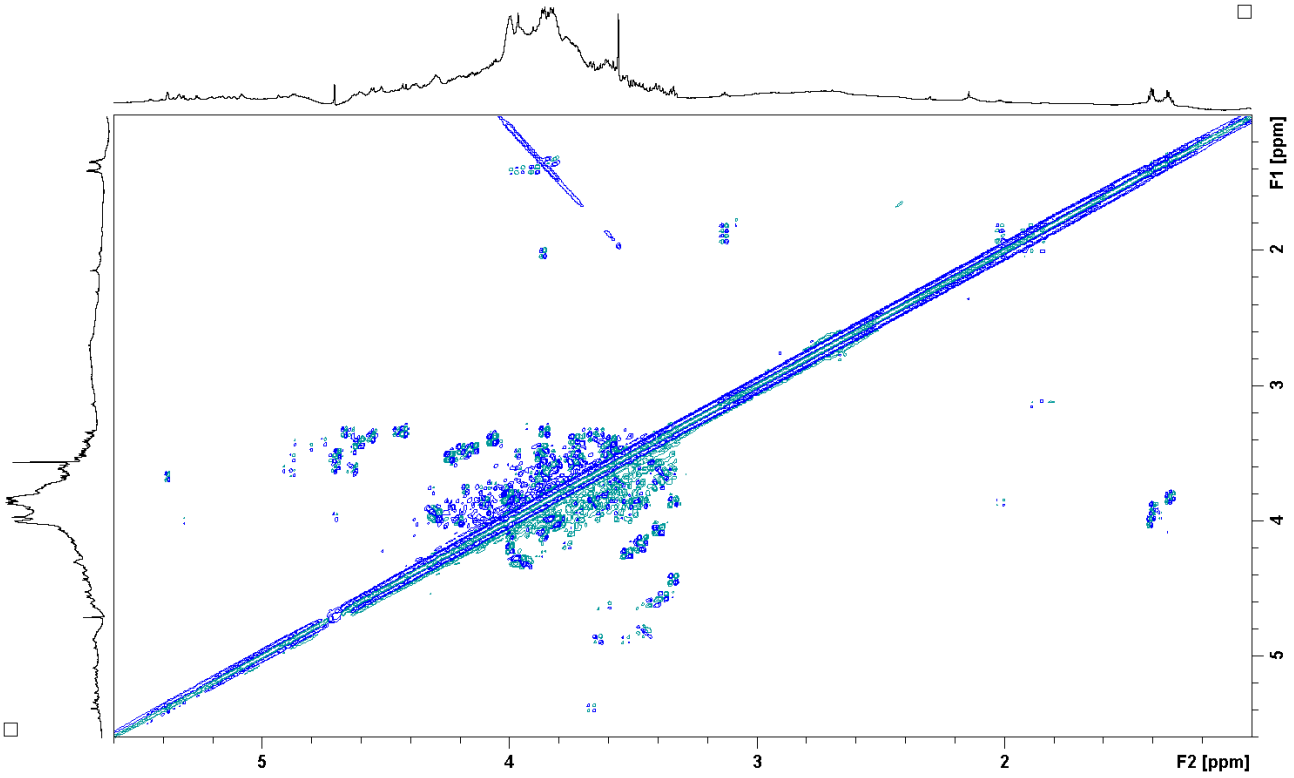


Fig. S16. DQF-COSY spectrum from the IOI-WAcF3 sample, zoomed out to show the suggested H6-H5 correlation from α-Rha

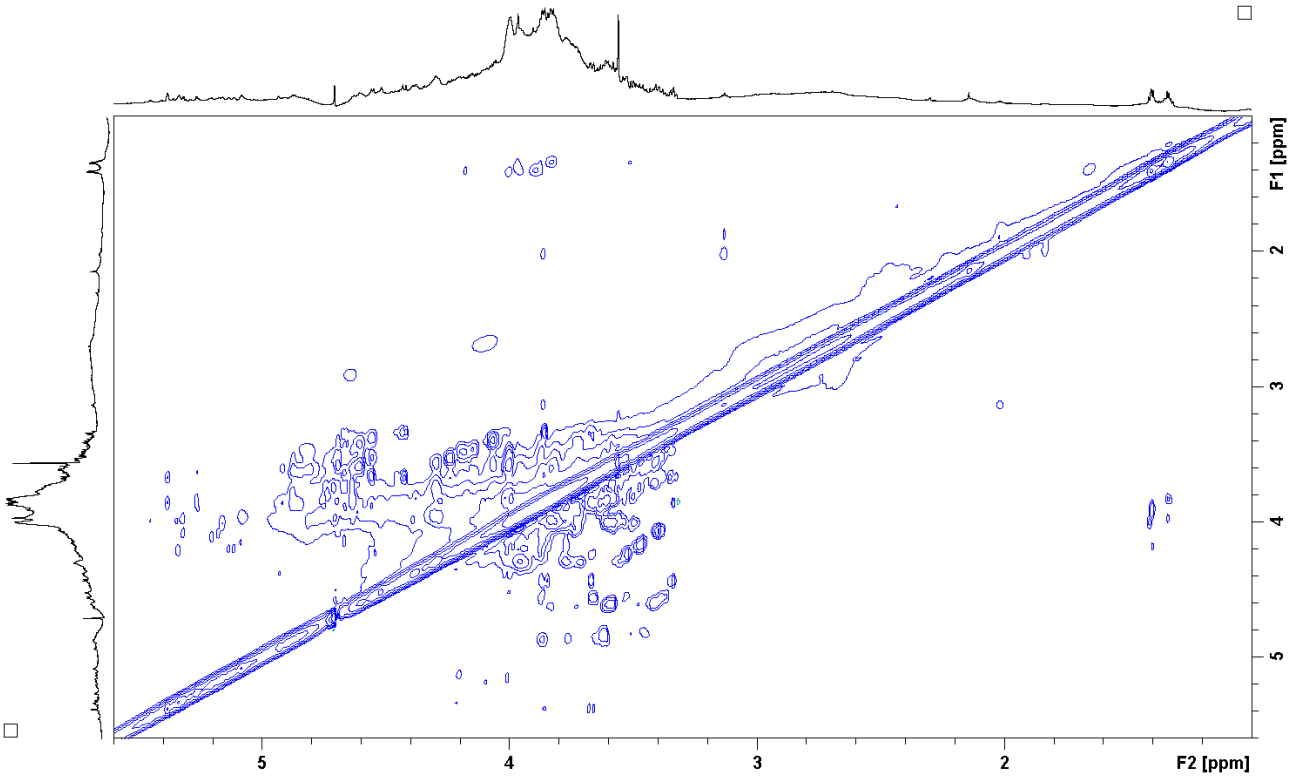


Fig. S17. TOCSY spectrum from the IOI-WAcF3 sample, zoomed out to show the suggested H6 correlations from α-Rha

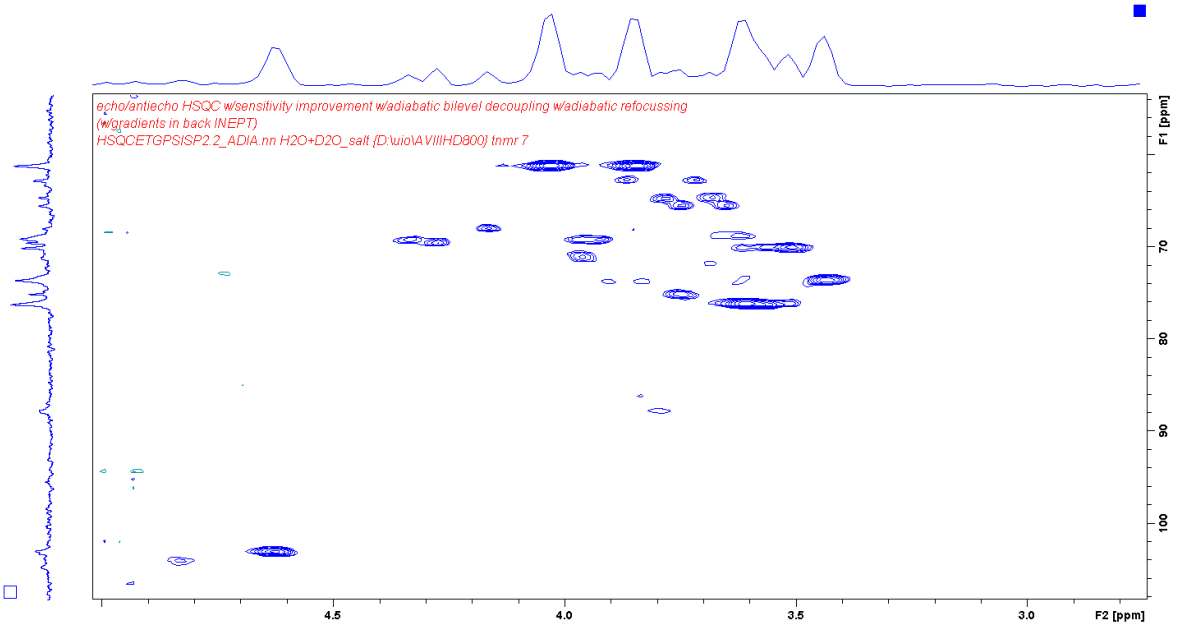


Fig. S18. HSQC spectrum from the IOI-A1 sample

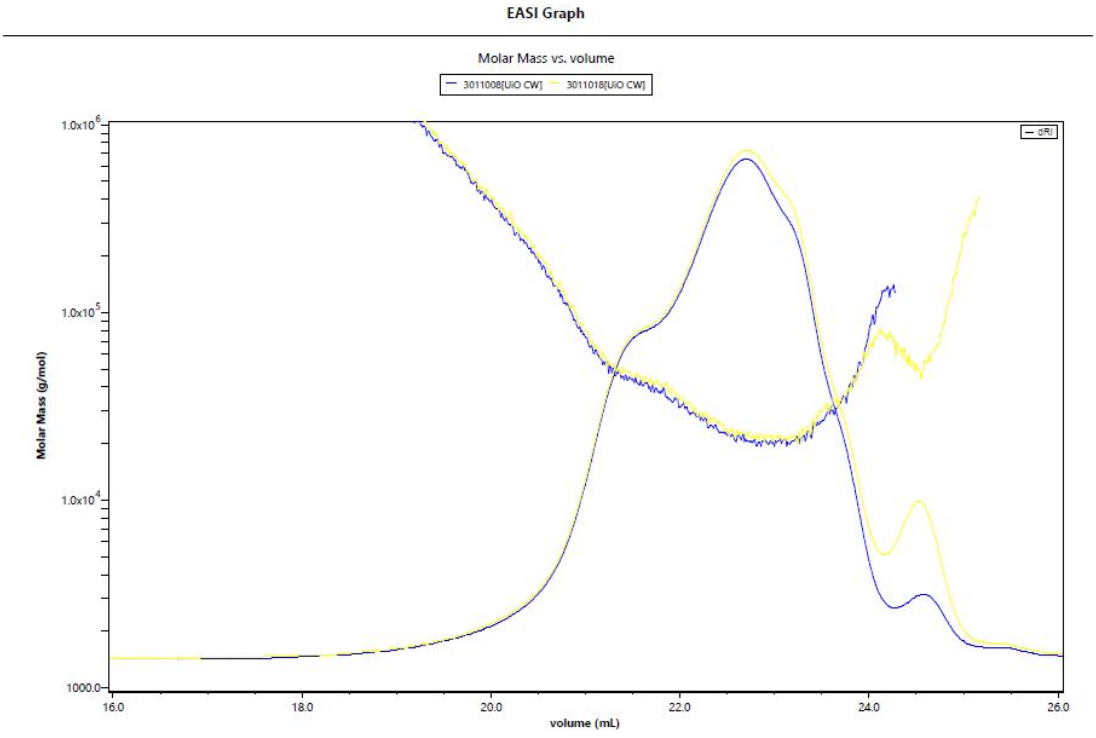


Fig. S19. SEC-MALLS chromatogram from the IOI-WN sample

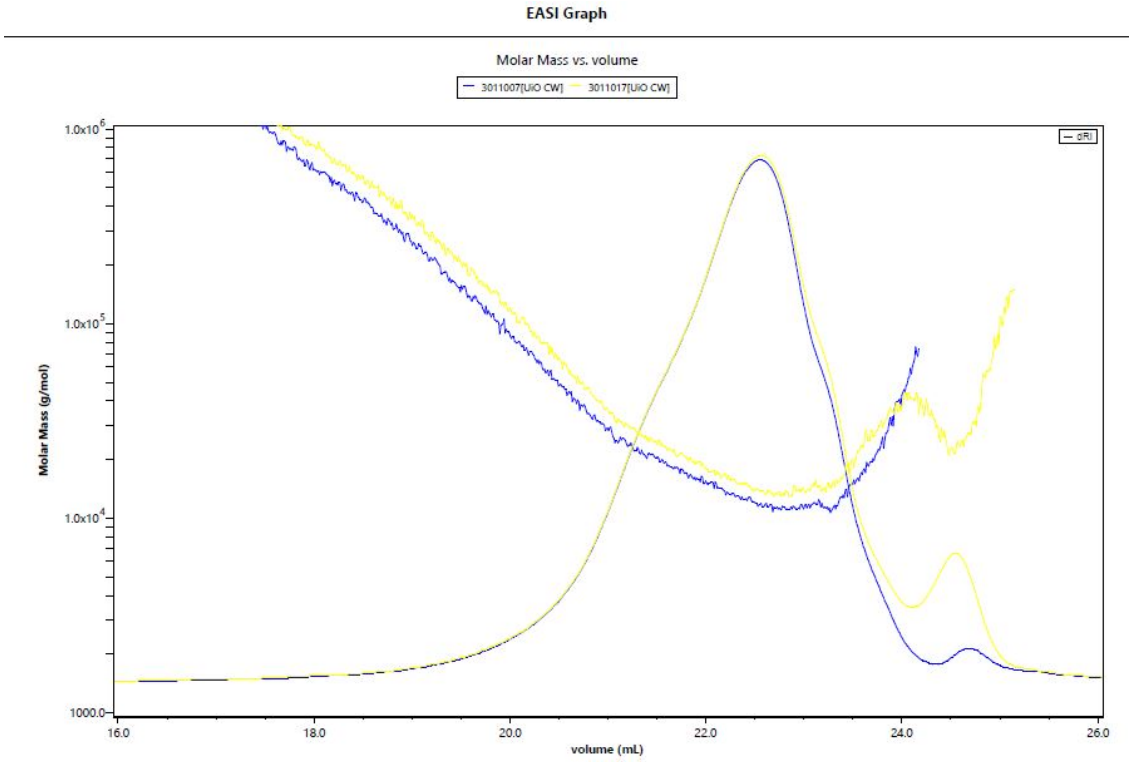


Fig. S20. SEC-MALLS chromatogram of IOE-WN

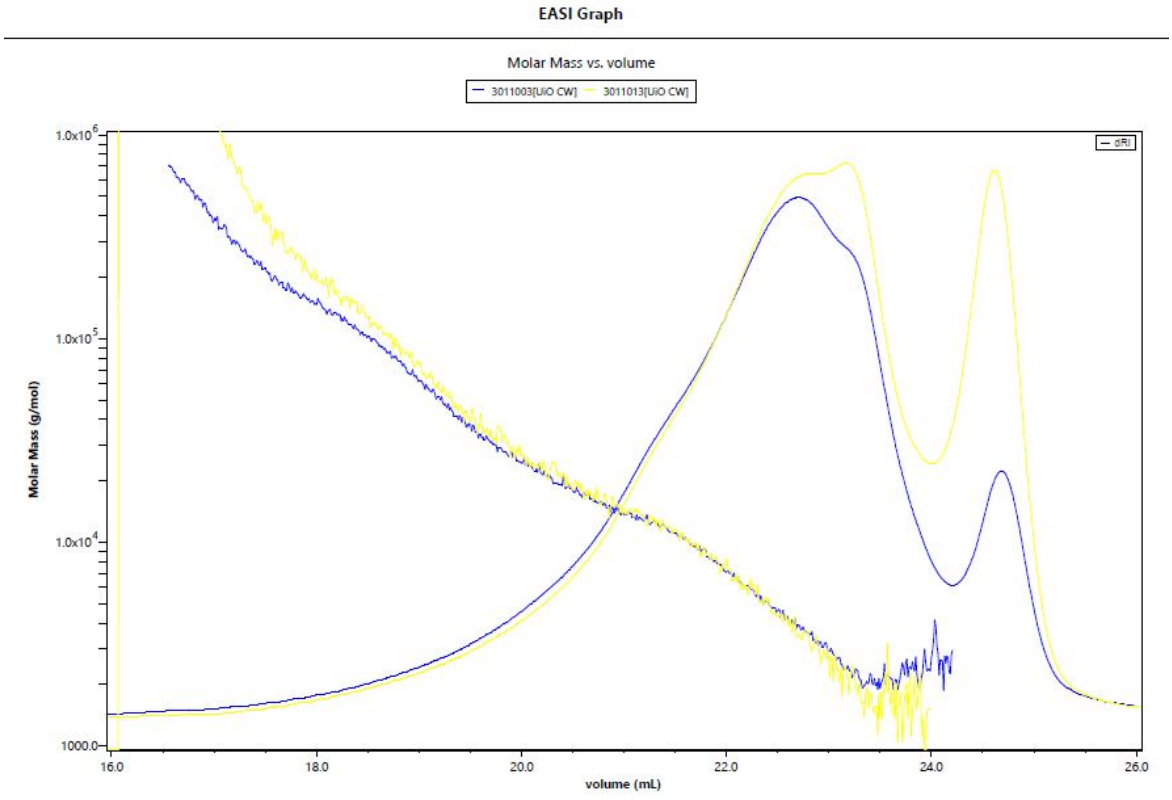


Fig. S21. SEC-MALLS chromatogram of IOI-WAc

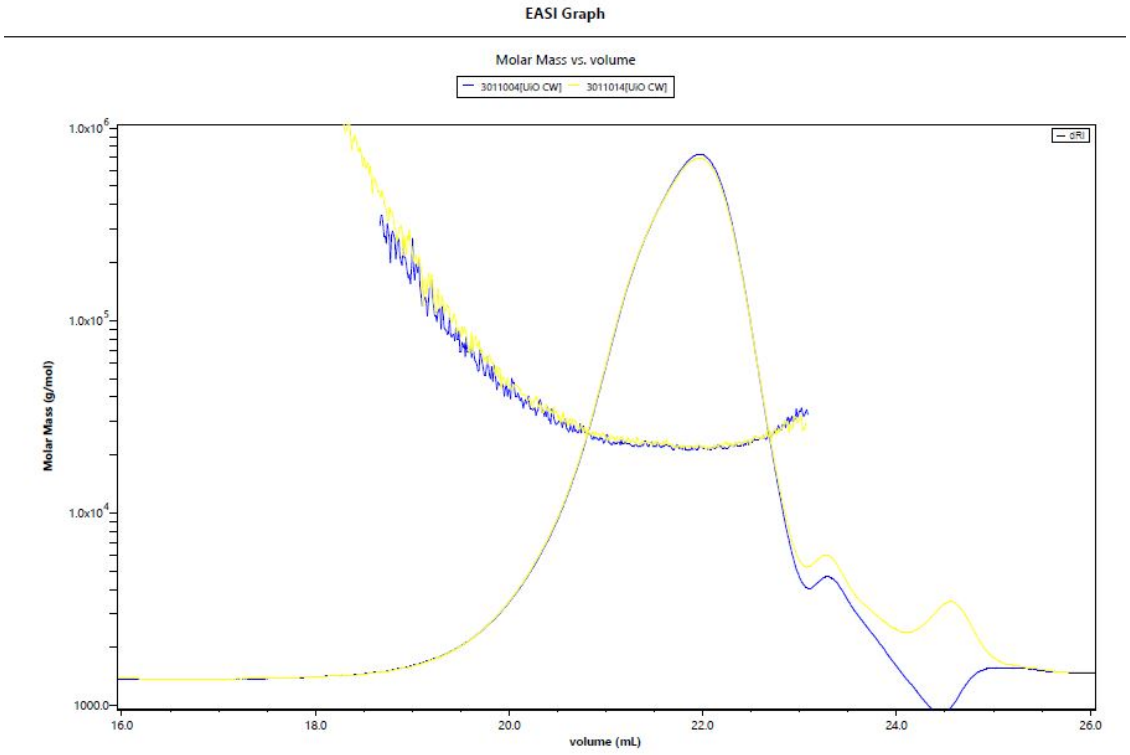


Fig. S22. SEC-MALLS chromatogram of IOI-WAcF1

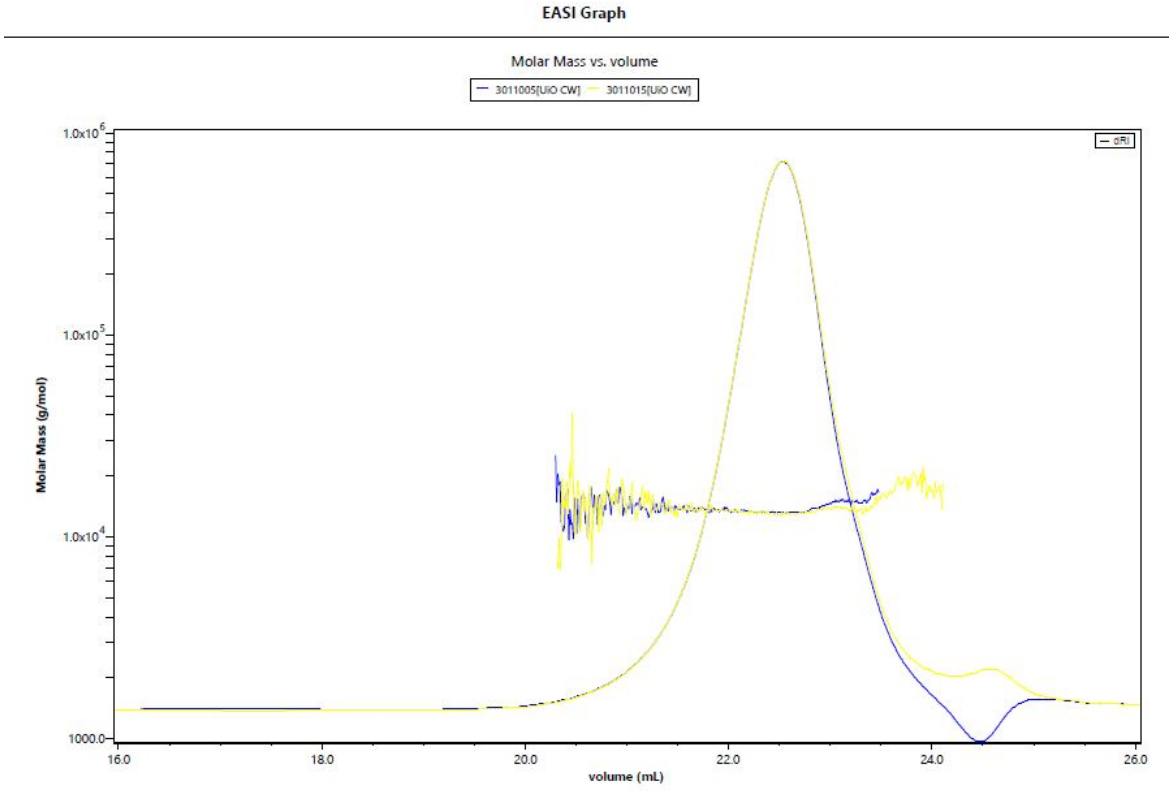


Fig. S23. SEC-MALLS chromatogram of IOI-WAcF2

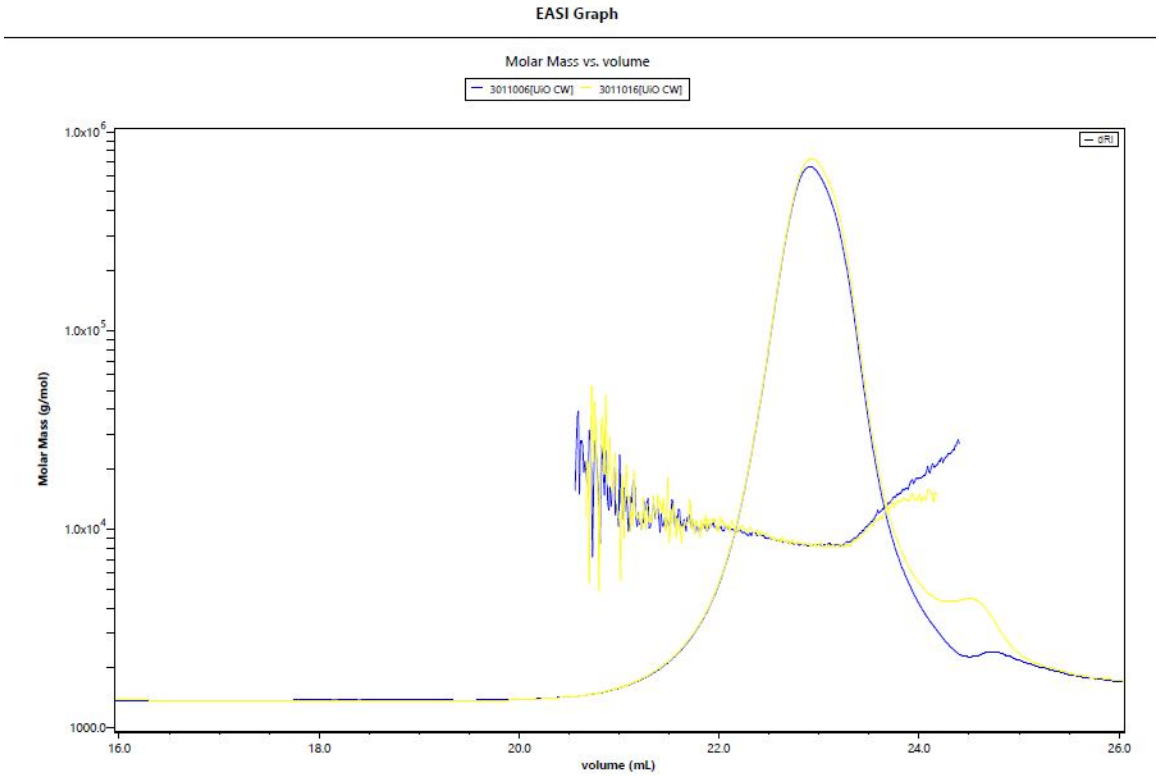


Fig. S24. SEC-MALLS chromatogram of IOI-WAcF3

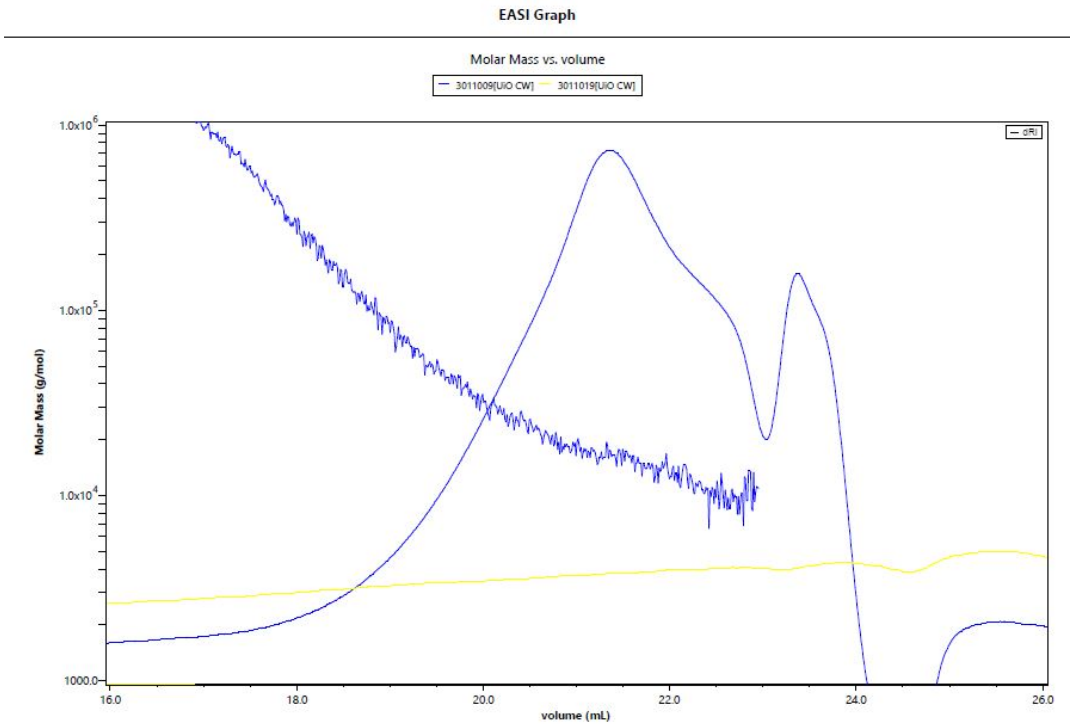


Fig. S25. SEC-MALLS chromatogram of IOE-WAc

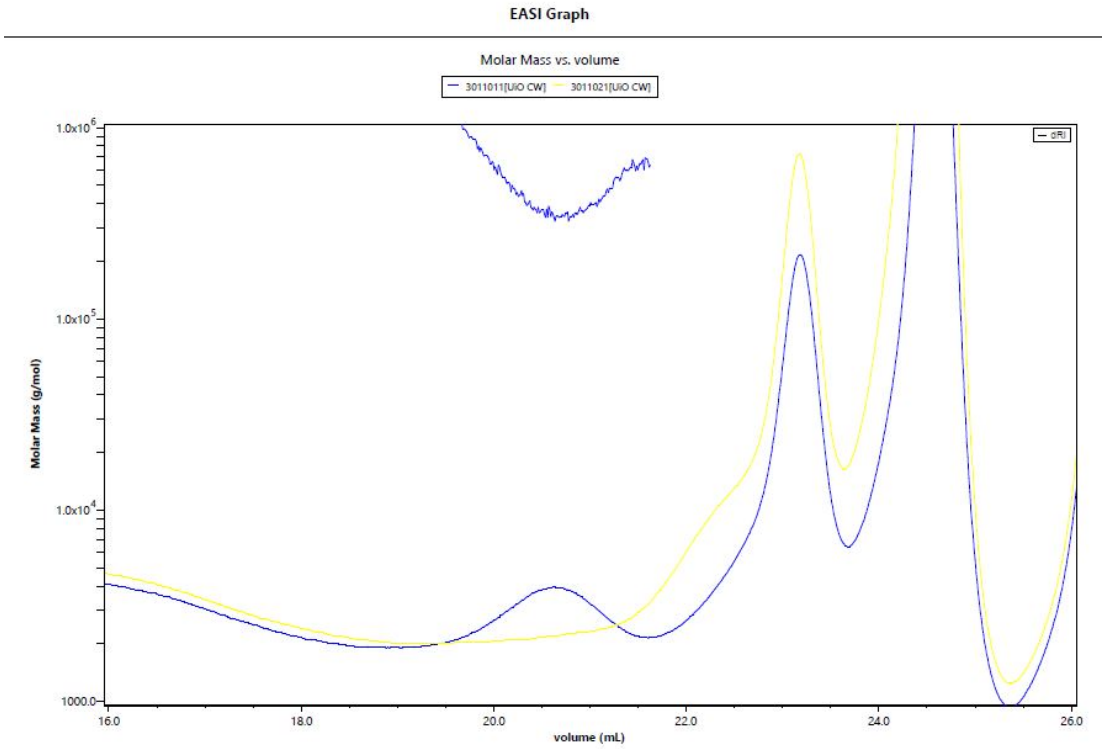


Fig. S26. SEC-MALLS chromatogram of IOI-A1

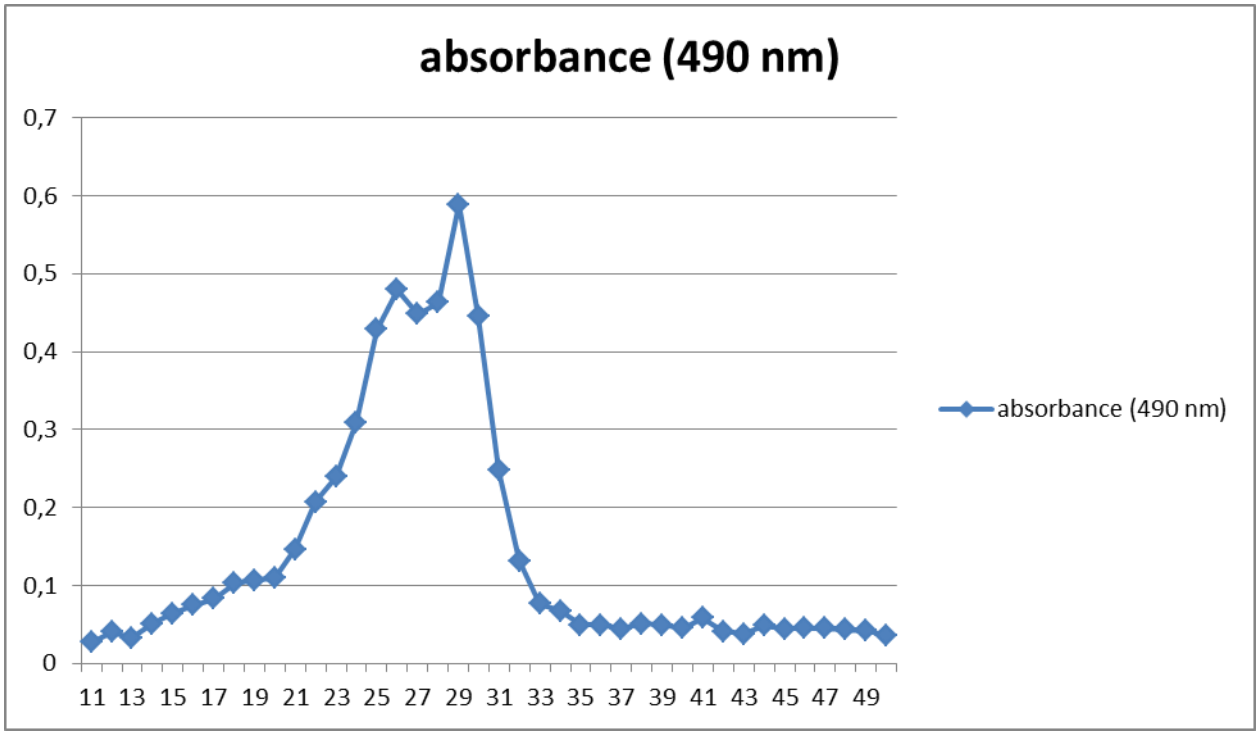


Fig. S27. The absorbance values of fraction 11-49 after phenol-sulphuric acid test of IOI-WN after Superose 6 SEC.

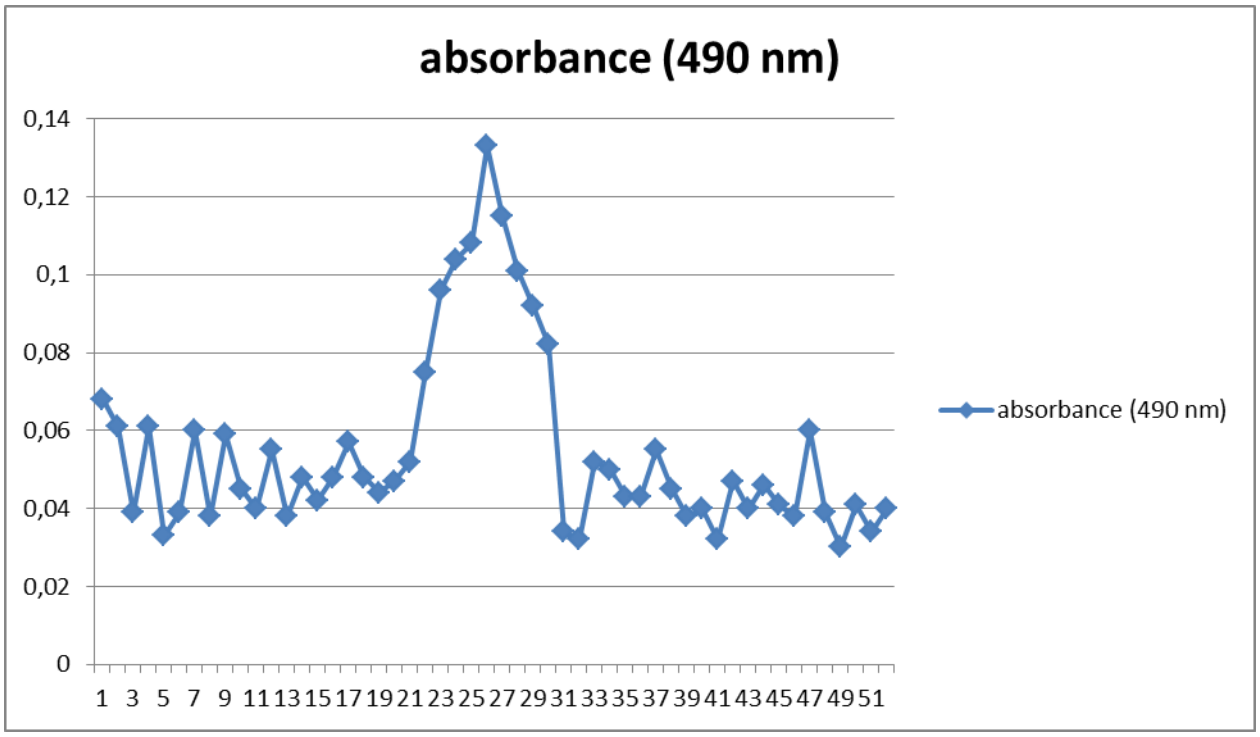


Fig. S28. The absorbance values of fraction 1-51 after phenol-sulphuric acid test of IOE-WN after Superose 6 SEC.

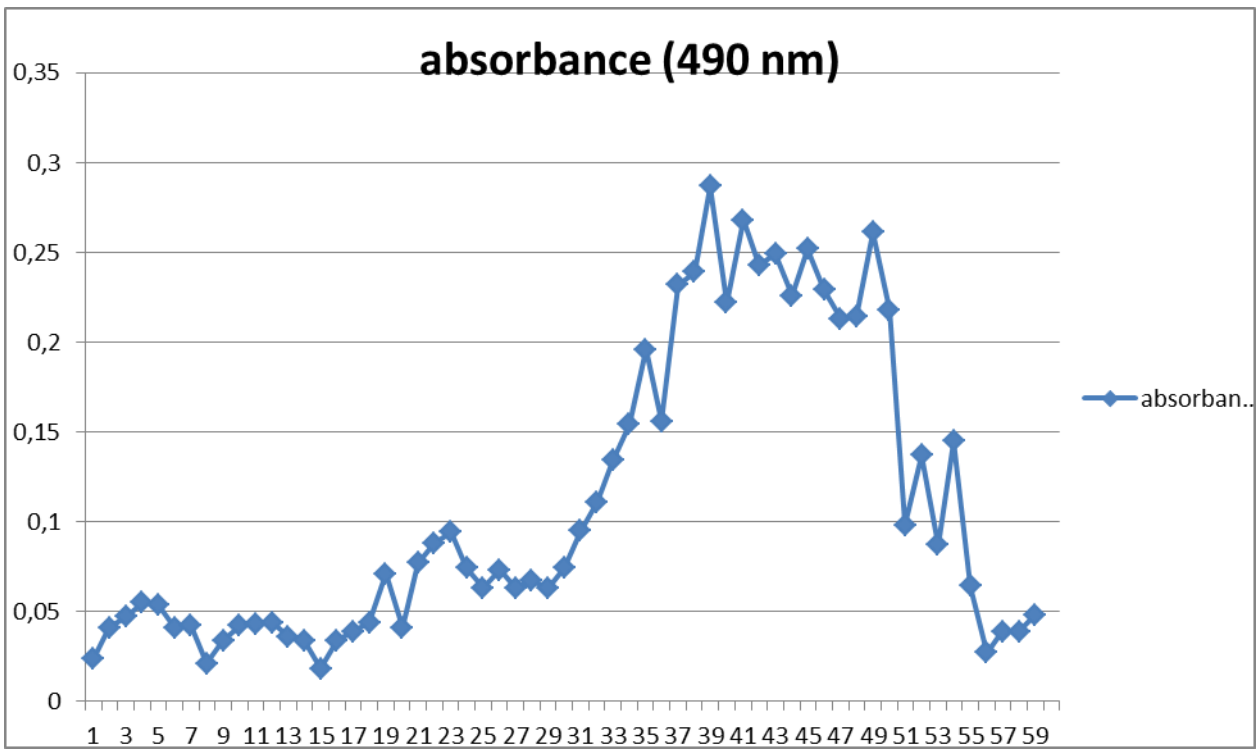


Fig. S29. The absorbance values of fraction 1-59 after phenol-sulphuric acid test of IOE-WN after HiLoad Supedex 200 SEC.

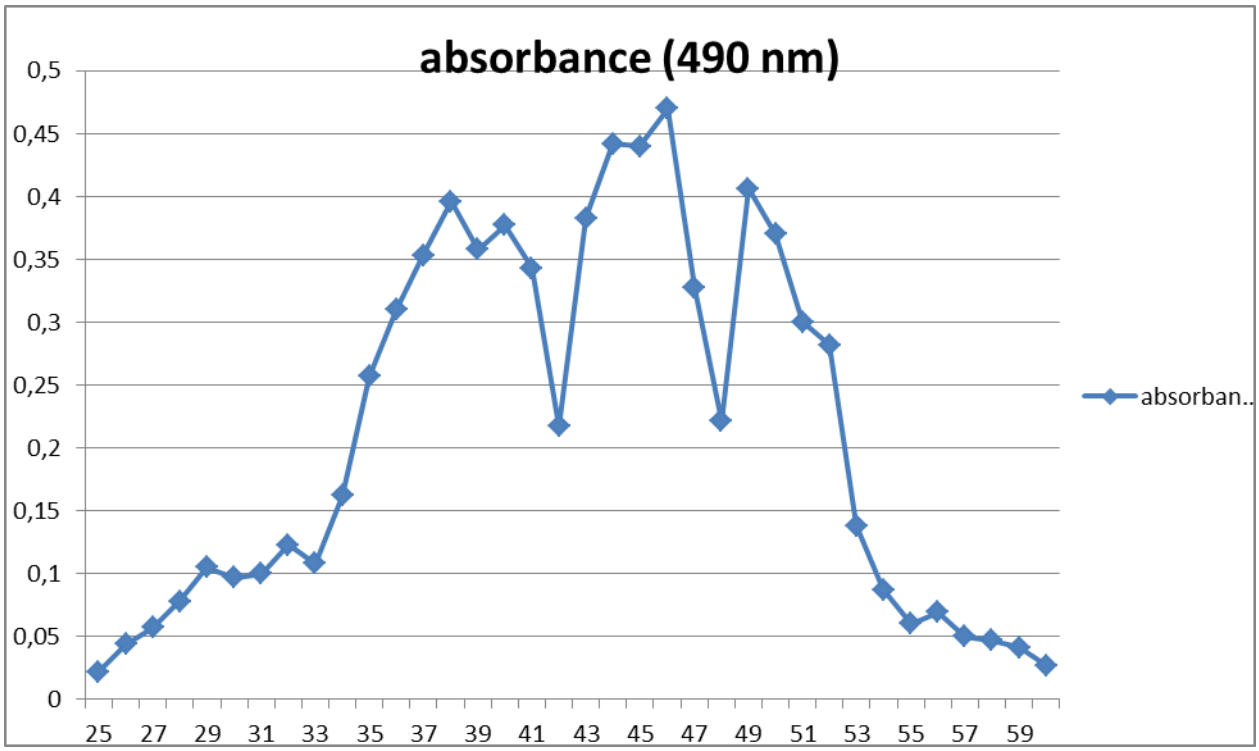


Fig. S30. The absorbance values of fraction 25-59 after phenol-sulphuric acid test of IOI-WAc after HiLoad Supedex 200 SEC. This sample was fractionated into three sub-frac tions, pooling fraction 30-41, 42-48 and 49-55.

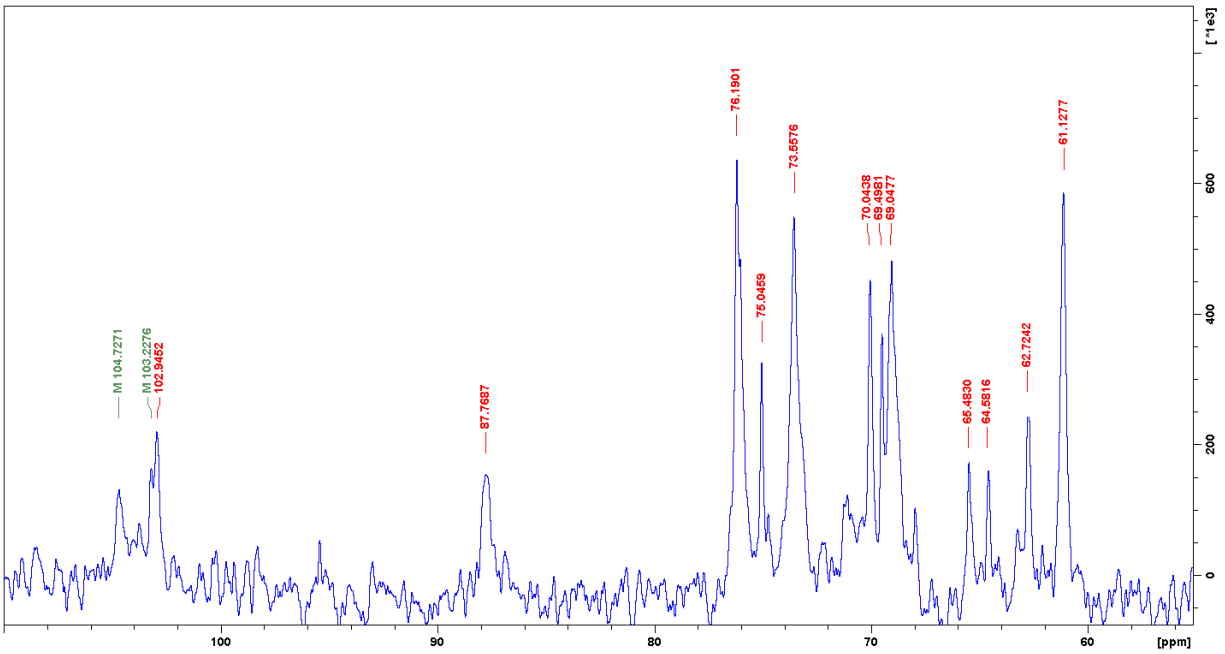


Fig. S31. ^{13}C NMR spectrum of the IOI-A1 fraction, showing the carbohydrate area from 55-110 ppm, with ppm values given for major peaks.

Paper 3

pH dependent binding patches of alginate trisaccharides on β-lactoglubolin investigated by HSQC-NMR

Stender, E. G. P.; Birch, J.; **Kjeldsen, C.**; Nielsen, L. D.; Duus, J. Ø.; Kragelund, B. B.; Svensson, B.

Manuscript in preparation

pH dependent binding patches of alginate trisaccharides on β-lactoglobulin investigated by HSQC-NMR

Emil G. P. Stender[1], Johnny Birch[1], Christian Kjeldsen[2], Lau D. Nielsen[3] Jens Ø. Duus[2], Birthe B. Kragelund[3], Birte Svensson[1*]

[1]Technical University of Denmark – Department of Biotechnology and Biomedicine

[2]Technical University of Denmark – Department of Chemistry

[3]University of Copenhagen – Department of Biology

*Corresponding author: Department of Biotechnology and Biomedicine - Technical University of Denmark. Søltofts Plads, building 224, DK-2800 Kgs. Lyngby. E-mail: bis@bio.dtu.dk; phone: +4545252740

Abstract

β-lactoglobulin is a promiscuous protein in terms of ligand interactions. It has been hypothesized that a specific binding patch on the protein mediates interaction with polyelectrolyte carbohydrates and controls selectivity for these. However, experimental evidence for such a patch on intact β-lactoglobulin remains to be reported. Here the binding patch on β-lactoglobulin of unsaturated alginate trisaccharides (AOS) is investigated at pH 2.65 and the dairy relevant pH 4.0 using ^{15}N ^{1}H-HSQC NMR. At pH 2.65 two hydrophilic surface-exposed binding patches are identified containing the following residues, which undergo significant chemical shift perturbations upon addition of AOS i) D11, I12, K14, and K75 and ii) K101, Y102, E127 D129 and D130. At pH 4.0 the latter site is not showing significant chemical shift perturbation, while the former is composed of K8, G9, D10, K75, F82, K83 and D85. The binding strength of AOS is very weak, likely in the mM range.

These results identify carbohydrate binding sites in BLG and can be helpful for design of polyelectrolyte-based BLG particles and coacervates with specific properties.

Introduction

β-lactoglobulin (BLG) is an extensively studied lipocalin from bovine milk [1,2] having a fold of nine β-strands (A-I), forming a calyx, and a couple of α-helixes (Fig. 1) [3]. Two disulfides (C66-C160 and C106-C119) and a free cysteine (C121) are buried in the hydrophobic core [4]. BLG is dimeric at neutral pH [4], exists in a monomer-dimer equilibrium at pH < 5 that at pH < 3 is completely dominated by the monomeric state [5]. It binds hydrophobic ligands, e.g. fatty acids and retinoids [6,7], in the central calyx [8]. Access to the calyx is controlled by the EF-loop adopting a closed conformation below pH 7.5 (Fig. 1) and an open confirmation at higher pH (the Tanford transition) [5]. Additionally, BLG has binding sites situated on the surface. Thus capryllic acid was shown to interact with the CD loop (residues 61-66) at the entrance to the calyx [7], while vitamin D_3 bound at a hydrophobic surface-exposed site containing D137, L140, K141, L143 M145, H146, I147 and R148 at the dimerization interface [6].

BLG interacts with a wide range of polysaccharides below its pI (4.7 – 5.2) [9–14] with formation of insoluble particles [9,13,15] or coacervates [12,14]. Several observations on these BLG complexes attributed their composition and physical properties to the polysaccharide characteristics [10,15–17]. This led to the hypothesis that one or more charge patches exist on the surface of BLG which mediate selective interactions with poly-anionic polysaccharides [10,12,13]. However, due to the large size of these complexes no details of their molecular structures have provided from experiments on intact BLG.

Alginate is a linear, acidic polysaccharide made of α-L-guluronic acid (G) and β-D-mannuronic acid (M) residues [18] organized as a co-polymer containing G-, M-, and M/G-blocks [19]. We previously investigated pH dependent particle formation between alginate and BLG and found that a mixture alginate trisaccharides (AOS) bound to BLG at pH 3 (K_d 1.1 mM) and pH 4 (K_d 0.6 mM) without forming insoluble particles [15]. The structure of carbohydrate binding sites in BLG has not been identified previously and ^{15}N ^{1}H-HSQC-NMR is used to show residues interacting with AOS at pH 2.65 and 4.0 by.

Materials and Methods

Production, purification and NMR of alginate oligosaccharides

AOS was prepared as previously described [15]. Briefly, 10 mg mL^{-1} alginate ($\bar{M}_n$ = 40 kDa M/G ratio = 0.6; gift of Dupont Nutrition Biosciences) was incubated with endoacting alginate lyase from *Sphingomonas sp* (Megazymes, Ireland), at 42° C for 6.5 h, followed by heating (90° C, 10 min) to stop the reaction and centrifuged (20,000 g, 10 min). The supernatant was desalted (Hiprep Desalt 26/10; GE Healthcare, USA) and oligosaccharides monitored by absorbance at 235 nm [15]. The desalted AOS was diluted with acetonitrile to 50 % (v/v) and further purified by HPLC (TSKgel Amide-80 column 5-µm particle size; 4.6 by 250 mm with 4.6-by 10-mm guard column (Tosoh, Japan); Ultimate 3000 HPLC (Dionex, CA) equipped with RI-101 refractive index detector (Showa Denko, Japan)) with mobile phase 70 % (v/v) acetonitrile in water at 70° C at a flow rate of 1 mL min^{-1}. Oligosaccharides were quantified using the phenol sulfuric acid method [15,20] and the purity assessed by TLC (silica gel 60 F_{254}; Merck, USA). Performed by spotting 3 µL sample twice, developed twice in 50 % butanol : 25 % acetic acid : 25 % miliQ water. Spots were visualized by taring (300° C) in 10 % sulfuric acid, 80 % ethanol, 8 % H_2O and 2 % orcinol.

AOS structures were analysed by NMR spectroscopy (Bruker Avance III (799.90 MHz for ^{1}H and 201.14 MHz for ^{13}C) equipped with a 5 mm TCI $^{1}H/(^{13}C, ^{15}N)$ cryoprobe using 4,4-dimethyl-4-silapentane-1-sulfonic acid as reference (-0.093 ppm for ^{1}H and -2.76 ppm for ^{13}C). The chemical shifts were assigned using 1D ^{1}H with presaturation, 1D ^{13}C, 2D double quantum filter correlated spectroscopy (DQF-COSY), 2D Rotating frame nuclear Overhauser effect spectroscopy (ROESY), 2D ^{13}C heteronuclear single quantum coherence (HSQC) with multiplicity editing, 2D ^{13}C HSQC-$[^{1}H,^{1}H]$ total correlation spectroscopy (HSQC-TOSCY) with 60 ms mixing time, 2D ^{13}C heteronuclear multi-bond correlation (HMBC) optimized for 10 Hz long range coupling constants and CLIP-HSQC. All spectra were recorded at 298 K using TopSpin 3.5 and processed and analysed using TopSpin 3.5 software (Bruker).

BLG cloning, mutagenesis and *Pichia pastoris* transformation

The vector encoding the BLG isoform A gene was manually optimized for *Pichia pastoris* resulting in a recombinant BLG A variant with an N-terminal sequence extension of 1E-2A-3E-4A-5Y and a point mutation V108F mimicking BLG described in [3]. The optimized DNA sequence was cloned into pPICZαA with the *Saccharomyces cerevisiae* α-mating factor introduced in frame; the final pPICZαA-BLGA plasmid was purchased (GeneArt; Thermofisher). Moreover an Xhoi restriction site was introduced after the α-factor to avoid additional amino acids after α-factor cleavage. We shall refer to the product of the optimized gene as BLG or BLGA. Site-directed mutagenesis to BLG K8A, K75A and K101A were done (Quick Change lightning mutagenesis kit; Agilent, USA) using the following primers: a34c 5'cgtcacccagaccatgcagggtttggatatcca3'; a34c_anti 5'tggatatccaaaccctgcatggtctgggtgacg3'; a235g_a236c 5'gtgctcagaagaagatcattgcagaagcaaccaagatccctg3'; a235g_a236c_anti 5'cagggatcttggttgcttctgcaatgatcttcttctgagcac3', a313g_a314c 5'ggttttggacaccgactacaaagcgtacttgttgttctgcatg3', a313g_a314c_anti

5'catgcagaacaacaagtacgctttgtagtcggtgtccaaaacc3' (Eurofins Genomic; Germany). Wild type (WT) and mutant plasmids were transformed into *Escherichia coli DH5α* by heat shock treatment and selected for zeocin resistance (Novagen, UK). The point mutations were confirmed by sequencing (GATC Biotech, Germany). pPICZαA-BLGA and mutant plasmids was linearized using the PmeI restriction site and transfected into *P. pastoris X-33* by electroporation (EasySelect™ Pichia Expression Kit; Invitrogen, USA). Transformants were selected on Zeocin containing Yeast extract Peptone Dextrose (YDP) agar plates incubated 3 d at 30°C.

Expression of isotope labelled recombinant BLG isoform A

Eight clones from each transformation were restreaked on YDP agar Zeocin plates and selected for expression in 50 mL Buffered Minimal Methanol media (BMM: 0.1 M KH2PO4 pH 6.0, 0.34 % (w/v) yeast nitrogen base without amino acids and ammonium chloride, 4×10^{-5} % (w/v) biotin, 1% ammonium sulphate, 0.5 % (v/v) glycerol) 5 d at 22°C. Aliquots were collected every 24 h analyzed by SDS-PAGE. The clones secreting the highest amount of protein to the supernatant were selected. Large scale expression of BLG WT and mutant proteins were done in 25 ml Buffered Glycerol-complex Medium (BMGY: 1 % (w/v) yeast extract, 2 % (w/v) peptone, 0.1 M KH2PO4 pH 6.0, 1.34 % (w/v) yeast nitrogen base, 4×10^{-5} % (w/v) biotin, 1 % (v/v) glycerol) inoculated with the chosen *P. pastoris* transformant, grown (30°C, 150 rpm) until OD_{600} = 2–6 and inoculated into 1 L BMGY medium and incubated as above in baffled 3 L shake flasks until OD_{600} = 2–6. The cells were harvested by centrifugation (1500g, 5 min, 22°C), resuspended in 1 L BMM media and incubated for 110 h (22° C, 150 rpm). Methanol was added to a final concentration of 0.5 % (v/v) every 24 h. Cells were pelleted after incubation (10,000*g*, 30 min, 4°C), the supernatant was filtered (0.45 µm) and concentrated by cross flow-filtration (SARTOFLOW® Slice 200 Benchtop System, 5 kDa Hydrostart ultrafiltration cassette at 4° C; Sartorius, Germany). After the yield of purified proteins was assessed spectrophotometrically at 280 nm and the conformational integrity was

confirmed by circular dichroism a large scale production was made in BMM containing ^{15}N-ammonium sulphate (BMM: 0.1 M KH2PO4 pH 6.0, 0.34 % (w/v) yeast nitrogen base without amino acids and ammonium chloride, 4×10^{-5} % (w/v) biotin, 1 % ^{15}N-ammonium sulphate (Cambridge Isotope Laboratories Inc, Andover, USA), 0.5% (v/v) glycerol) as described above.

Purification of recombinant BLG WT and mutants

Concentrated culture supernatant (see above) was buffer exchanged to 50 mM K_2HPO_3/KH_2PO_3 (KPi) pH 6.0, 150 mM NaCl (Hiprep Desalt column 26/10; GE healthcare) at 4° C and flow rate of 2 ml min^{-1} and concentrated (Amicon Ultra 15 centrifugal filter 3 kDa cut off; Merck, USA) at 4° C. Recombinant BLG proteins were purified by size exclusion chromatography (Hiload Superdex 75 26/60; GE healthcare) at a flowrate of 1.4 ml min^{-1} at 4° C in 50 mM KPi pH 6.0, 150 mM NaCl. The purified proteins where dialyzed (miliQ water, 3 kDa cut off Spectra/Por membrane; Spectrumlabs, USA) at 4° C by 3 x 100 fold dilution, each for 4 h each. The dialyzed proteins were lyophilized; evaluated by SDS-PAGE and stored at –20° C until use. For NMR experiments lyophilized proteins were dissolved in 55 mM KPi pH 2.65, pH 3.2 or pH 4.0 and dialyzed against this buffer (as above). Protein concentration were determined spectrophotometrically using a predicted molar extinction coefficient of 18700 M^{-1} (Protparam) [21].

Circular dichroism spectroscopy

Secondary structure of unlabeled BLG WT and mutants was assessed by circular dichroism. The lyophilized proteins were dissolved in 10 mM sodium phosphate pH 7.0, centrifuged (20,000*g*, 20 min, 4° C) and dialyzed (3 kDa cut off, Spectra/Por membrane; Spectrumlabs, USA) against 10 mM sodium phosphate by 3 x 100 fold dilution for at least 4 h at 4°C. Far-UV CD spectra were recorded at 250–190 nm in a 1 mm quartz cuvette (50 nm min^{-1}, 1 nm bandwidth and 2 s response time at 25°

C; Jasco J810 Spectropolarimeter). Ten scans were averaged and a buffer scan was subtracted as background. The molar ellipticity was calculated by:

$$[\theta] = 100 \times \frac{\theta_\lambda}{m} \times d$$

where m is the molar protein concentration, θ_λ the ellipticity at wavelength λ, and d the path length in cm as described in detail in [22].

NMR of BLG WT and mutant proteins

^{15}N-BLG WT and mutant proteins were diluted to 50 mM KPi H_2O/D_2O (9/1 v/v), 0.1 mM 2,2-dimethyl-2-silapentanesulfonic acid (DSS) and analyzed by NMR spectroscopy in 5 mm Shigemi microtubes. At pH 2.65 BLG WT is monomeric [5] and stable for many days [3], while a significant amount of dimer is present at pH 3.2 and 4.0 [5]. Protein samples were centrifuged (12,000 g, 20 min, 4°C) prior to analysis. All ^{1}H/^{15}N – heteronuclear single quantum coherence (HSQC) spectra were recorded in a 800 MHz Varian INOVA spectrometer equipped with a 5 mm triple resonance room temperature probe with a Z-field gradient at 37° C with a mixing time of 150 ms using a Varian/Agilent BioPack sequence or a 600 MHz Bruker AVANCE system equipped with a cryoprobe. Recorded free induction decays were processed with nmrPipe [23]. Proton chemical shifts were referenced to internal DSS at 0.00 ppm.

NMR assignment and chemical shift perturbation analysis

Assignment of BLG backbone nuclei at pH 2.65 was done by manually picking and linking resonances using CcpNMR analysis [24] to the chemical shift list of a BLG A solution structure previously solved by NMR spectroscopy under identical conditions [3]. Spectra of 400 µM BLG WT recorded at pH 2.65, 3.2 and 4.0 were overlaid and used to assign the backbone nuclei at pH 4.0. Secondly, ^{1}H/^{15}N – HSQC spectra were recorded of sample containing 100 µM BLG WT or

mutant protein and 15.5 mM AOS at pH 2.65 or pH 4 prior to mixing in 4 steps with a BLG sample not containing AOS resulting in a total of 6 spectra recorded per sample at varying AOS concentration. Chemical shift perturbation was used to map the binding patch on the surface of the protein, size of the chemical shift changes was calculated by:

$$\Delta\delta = \sqrt{(\Delta\delta H)^2 + (\Delta\delta N/5)^2}$$

Where $\Delta\delta H = \Delta\delta H_{free} - \Delta\delta H_{obs}$ and $\Delta\delta N = \Delta\delta N_{free} - \Delta\delta N_{obs}$ are changes in the proton and nitrogen chemical shifts (in ppm).

Results

Yield, purity and composition of AOS

An AOS trisaccharide mixture was prepared from alginate by treatment with an endoacting alginate lyase yielding di- and trisaccharides as end products [15] and purification by HPLC (see Methods) in amounts of 12 mg (Supporting Fig S1).

In order to determine the composition of the purified AOS the structures were determined by NMR. The spectra revealing a mixture of two trisaccharides having only guluronic acid at the reducing end and 75 % guluronic acid and 25 % mannuronic acid at the central sugar residue, respectively. The lyase removes the 4-OH group and proton at the 5-position of a-glycon by β-elimination [25], thus the C5 epimers β-D-mannuronic acid and α-L-guluronic acid lead to the same product at the non-reducing end (Figure 2; Supporting table S1).

The ^{1}H, ^{13}C HSQC spectrum of the AOS sample revealed seven distinct peaks corresponding to anomeric positions, two of which are from a reducing end. The reducing end could not be of

mannose configuration as it had an 8.4 Hz $^{3}J_{H,H}$ coupling constant between the anomeric proton to the 2-position proton, and thus it had to be of gulose configuration.

The seven distinct monosaccharide units were labelled **A-G** in order of descending anomeric ^{13}C chemical shift. The assignment of the individual monosaccharide units were performed using DQF-COSY, ROESY, HSQC-TOCSY and HMBC. The anomeric configuration was determined using $^{1}J_{H1,C1}$ coupling constants [26] obtained using CLIP-HSQC [27]. Anomeric signals, **A** and **E**, were considerably lower in intensity compared to the other five and corresponded to the central mannuronic acid residue linked to reducing end guluronic acid in α/β configuration. Signal **B** was determined to be from the central guluronic acid; CLIP-HSQC showing the anomeric proton to be equatorial, and as the $^{3}J_{H3,H4}$ coupling constant was 3.6 Hz, this was in α-configuration. Signal **C** was from the non-reducing end containing a double bond between C4 and C5. Similarly, the fourth was from an almost identical unit, and was also a non-reducing end. The **C** and **D** originated from whether they were attached to a guluronic acid or a mannuronic acid, respectively [28]. The two remaining anomeric positions were from the α- and β-reducing ends. As mentioned above, the reducing end was determined to be of gulose configuration. The relative signal ratio between the β-reducing end, **F**, and the α-reducing end, **G**, was approximately 1.9:1. Using HMBC and ROESY experiments it was possible to determine the connections between each monosaccharide, which were all 1,4 connected.

Yield and purity of BLG WT and mutant proteins

The purity of BLG proteins was assessed by SDS-PAGE and concluded to be more than 95 % (data not shown). The yields were BLG WT 40.4 mg L^{-1}, K8A 20.6 mg L^{-1}, K75A 35.2 mg L^{-1} and K101A 6.3 mg L^{-1} where L refers to liters of BMM containing ^{15}N-ammonium sulphate.

Chemical shift assignment of BLG WT at pH 2.65

To make it possible to study the chemical shift perturbations upon addition of AOS, BLG WT was prepared having identical primary structure as the BLG A variant previously studied by NMR and the spectra were recorded in similar conditions [3]. The 162 residues including eight prolines in BLG should result in 154 resonance peaks. The recorded ^{1}H, ^{15}N-HSQC spectrum has 156 dispersed peaks of which 145 could be assigned to BLG WT residues (supporting table S2). The remaining nine residues were ambiguous either due to lack of assignment in the reported CS list (L22, E157, E158 and I162) or because of peak overlap (Y2, K8, N63, and C160) (supporting table S2). Assigned peaks are in good agreement with the chemical shifts reported for BLG WT [3] indicative of the recombinant BLG WT being folded correctly.

Interaction between BLG WT and mutant proteins and AOS at pH 2.65

Chemical shift perturbation analysis identified residues affected by addition of AOS. At pH 2.65 increasing amounts of AOS caused chemical shift positions of backbone amides to change (Fig 3A). Three surface exposed regions, D11-K14, K101-Y102 and E127-D137, thus showed significant chemical shift perturbations, but all perturbations were small and fitting to a classical binding model was not possible (data not shown). At pH 2.65 also S30 and K75 showed significant chemical shift perturbations (Fig 4A, Table 1). Three BLG single mutant proteins K8A, K75A and K101A were produced to investigate the effect of removing positive charge in the affected regions. The argument for K8A was twofold i) alginate-BLG interactions are ionic and K8 is very close to the patch containing K75 in the tertiary structure of BLG. Since, three unassigned spin system underwent significant chemical shift perturbations by addition of AOS (Fig. 4A) K8 was hypothesized to be one of them and ii) K8A was originally unassigned. The structural integrity of the mutant proteins was assessed by far-UV CD on unlabeled recombinant BLG proteins. As expected for the β-sheet-dominated BLG [29] (Fig. 5) a global minimum at 216–218 nm was seen and small differences

were observed between WT and the mutant proteins. However, the ^{15}N, ^{1}H-HSQC spectrum of K101A BLG showed extensive peak overlaps suggesting a partially unfolded structure (supporting fig S2C). Hence, interaction studies were not pursued for K101A. At pH 2.65 K8 BLG showed no significant AOS elicited chemical shift perturbation (Fig 4A, table 1), and although K75 did show significant chemical shift perturbation at pH 2.65 by AOS the K75A mutant gave little difference in chemical shift perturbation compared to WT upon addition of AOS (Fig. 4, table 1).

Chemical shift assignment of BLG WT at pH 4

To assign the chemical shifts at pH 4, spectra recorded at pH 2.65, 3.2 and 4.0 were overlaid (Supporting fig S3; Supporting table S3) making it possible to follow differences in chemical shift positions between pH 2.65 and 4.0 thereby providing the peak assignment at pH 4.0. Several peaks, however, are broadened and as a consequence assignments were not possible (Supporting table S4). This peak broadening likely stems from BLG occurring in a monomer/dimer equilibrium at pH 4 [5], and accordingly residues with lost peaks were situated at the BLG dimerization interface (Supporting fig S4).

Comparison of chemical shift perturbation upon addition of AOS at pH 2.65 and 4.0

Chemical shift perturbations analysis was performed at the dairy relevant pH 4.0 (Fig 6B, 3B) which revealed several differences occur compared to the results at pH 2.65. Notably, Y102 no longer showed significant chemical shift perturbation, while K101 did but to a smaller extent than at pH 2.65 (Fig 6). By contrast, K75 retained significant chemical shift perturbation. Residues located in the C-terminal part of BLG showed no significant chemical shift perturbation and the signal for D129 was lost (Table 1; Supporting table S3). A few peaks, including those of the surface exposed

246 K8, K83, and D85 which appeared not affected at pH 2.65 changed significantly at pH 4. A80,
247 however, was affected as well although this residue is buried in the hydrophobic calyx and no
248 suggestion can be given for the structural basis of this effect.

249 Mapping the significant chemical shift perturbations at pH 2.65 and 4.0 to the surface of monomeric
250 BLG showed two patches (fig. 8), at pH 2.65 one contained D11, E12, K14 and K75 a second one
251 K101, Y102, E126, D128 and D129. At pH 4 residues surrounding K101 showed no significant
252 chemical shift and for K101 significantly smaller chemical shift perturbation than was seen that at
253 pH 2.65. This may be due to partial deprotonation of E126 (pK_a 4.07), D128 and D129 (pK_a 3.90)
254 thereby changing the electrostatic surface of the patch. At pH 4 the patch surrounding K75 also
255 changes. Thus K14 no longer contributes, while involvement of K8, K83, F82 and D53 appeared.

256 Since the chemical shifts change at pH 4.0 for both K8 and K75, turbidimetry was performed on
257 BLG WT, K8A and K75A at varying concentration of alginate (40 kDa) under the assumption that
258 the avidity effect of polysaccharide binding would increase potential differences. K8A and K75A
259 BLG were marginally less soluble than WT when mixed with alginate, but at high alginate
260 concentration BLG WT displays a higher turbidity than these two BLG mutant proteins (Supporting
261 fig S6). The reason for this may be the loss of a surface charge or decreased interaction with
262 alginate or both.

263 Discussion

264 As none of the chemical shift perturbations reached saturation they were not fitted to a binding
265 model. This is contrary to what we previously reported for binding of the AOS to BLG [15]. An
266 explanation may be found in the 5 fold higher ionic strength in the NMR experiments affecting
267 electrostatic interaction between alginate and BLG [13,15,30]. The buffer was chosen to mimic the

conditions reported in [3] and to ensure that of the pH did not change upon addition of AOS. BLG is known to interact with a wide range of ligands with different binding sites spanning from hydrophobic ligands accommodated in the calyx [8] to vitamins binding to the surface [6]. The binding patches described for AOS consistently differ from previously reported binding sites on the BLG surface [6,7]. Lysine is consistently present in the AOS binding patches and lysine residues undergo clear perturbation of chemical shifts. Still, only a three at pH 2.65 and four at pH 4.0 of the 15 surface exposed lysines [3] show significant chemical shift perturbation probably connected with local surface electrostatics of the immediate environment preventing interaction with AOS.

Surface binding of aroma compounds to BLG has previously been described at pH 2.0 by NMR. In that study two hydrophobic surface binding sites were identified consisting of K47, L57, K70, and I72 for γ-decalactone and K60, Y102, L104 and E129 for β-ionone [31]. Both sites are located in the hydrophobic part between β-strand G, the large α-helix and β-strand I. These are distinctly different from the AOS binding patches though Y102 is shared.

Charge patches responsible for interaction of BLG with acidic polysaccharides have been hypothesized in several papers [10,12,13,15,16]. The present work identifies binding patches of AOS on intact BLG at the molecular level, two patches are found at pH 2.65 one of which remains at pH 4.0. The hypothesized regions that could interact with acidic polysaccharides include residues 1-14, 41-60, 76-83 and 132-148 [12,13]. Here a patch consisting of K8, G9, D10, D53, K75, K83 and D85 is the area interacting with AOS at pH 4. Almost no surface exposed residues from the region 41-60 undergo significant chemical shift perturbation indicating lack of interaction. Excluding D53 exposed on the surface and V41 and L53 which are not (Fig 6) [3]. The observed interaction patch is consistent with hypothesized residues in regions 1-14 and 75-85 being involved in binding at both pH 4.0 and 2.65. Significant chemical shifts are also observed for region 132-148 at pH 2.65, but these were lost at pH 4.0 due to line broadening probably caused by formation of

BLG dimer [5]. At pH 2.65, a second binding patch was observed near K101 that has not previously been reported in the literature. This is consistent with our previous finding that at low pH one BLG monomer can bind two AOS while only one AOS binds at pH 4 [15]. Overall these results identify binding patches on BLG for anionic oligo- and polysaccharides which can aid in design of polysaccharides aimed for specific coacervate properties with BLG.

Conclusion

pH has prominent effect on the availability of functional AOS binding patches for on BLG. As determined using HSQC-NMR at pH 2.65 two surface exposed patches centered around K75 and K101 underwent significant chemical shift perturbations elicited by AOS. These patches are different from binding sites on BLG reported in literature for other compounds. At pH 4 most of the significant chemical shift perturbations around K101 seen at pH 2.65 did not occur upon addition of AOS indicative of loss of the patch. Moreover, the patch around K75 included other residues than at pH 2.65. Additional mutations in the binding patches and assays of macromolecular particle formation can to further describe the function of these binding sites. Overall, this work will help guide design for polyelectrolyte carbohydrates to achieve specific coacervate properties with BLG.

Acknowledgements

The project StrucSat is supported by a grant from the Danish Council of Strategic Research (Grant no. 1308-00011B) and a third of a PhD stipend (to EGPS) from Technical University of Denmark. Karina Jansen is thanked for technical assistance.

References

1. Flower DR. The lipocalin protein family: Structure and function. Biochem J. 1996;318: 1–14.

2. Bordin G, Raposo FC, de la Calle B, Rodriguez AR. Identification and quantification of major bovine milk proteins by liquid chromatography. J Chromatogr A. 2001;928: 63–76. doi:10.1016/S0021-9673(01)01097-4

3. Uhrínová S, Uhrín D, Denton H, Smith M, Sawyer L, Barlow PN. Complete assignment of 1H, 13C and 15N chemical shifts for bovine β-lactoglobulin: secondary structure and topology of the native state is retained in a partially unfolded form. J Biomol NMR. 1998;12: 89–107.

4. Brownlow S, Cabral JHM, Cooper R, Flower DR, Yewdall SJ, Polikarpov I, et al. Bovine β-lactoglobulin at 1.8 angstrom resolution - Still an enigmatic lipocalin. Structure. 1997;5: 481–495. doi:10.1016/S0969-2126(97)00205-0

5. Taulier N, Chalikian T V. Characterization of pH-induced transitions of β-lactoglobulin: Ultrasonic, densimetric, and spectroscopic studies. J Mol Biol. 2001;314: 873–889. doi:10.1006/jmbi.2001.5188

6. Yang MC, Guan HH, Liu MY, Lin YH, Yang JM, Chen WL, et al. Crystal structure of a secondary vitamin D3 binding site of milk β-lactoglobulin. Proteins Struct Funct Genet. 2008;71: 1197–1210. doi:10.1002/prot.21811

7. Loch J, Polit A, Gorecki A, Bonarek P, Kurpiewska K, Dziedzicka-Wasylewska M, et al. Two modes of fatty acid binding to bovine β-lactoglobulin-crystallographic and spectroscopic studies. J Mol Recognit. 2011;24: 341–349. doi:10.1002/jmr.1084

8. Wu SY, Pérez MD, Puyol P, Sawyer L. β-Lactoglobulin binds palmitate within its central cavity. J Biol Chem. 1999;274: 170–174. doi:10.1074/jbc.274.1.170

9. Birch J, Harðarson HK, Khan S, Van Calsteren M-R, Ipsen R, Garrigues C, et al. Effect of repeat unit structure and molecular mass of lactic acid bacteria hetero-exopolysaccharides on binding to milk proteins. Carbohydr Polym. 2017;177: 406–414. doi:10.1016/j.carbpol.2017.08.055

10. Du X, Dubin PL, Hoagland DA, Sun L. Protein-selective coacervation with hyaluronic acid. Biomacromolecules. 2014;15: 726–734. doi:10.1021/bm500041a

11. Aberkane L, Jasniewski J, Gaiani C, Scher J, Sanchez C. Thermodynamic characterization of acacia gum-β-lactoglobulin complex coacervation. Langmuir. 2010;26: 12523–12533. doi:10.1021/la100705d

12. Girard M, Turgeon SL, Gauthier SF. Quantification of the interactions between β-lactoglobulin and pectin through capillary electrophoresis analysis. J Agric Food Chem. 2003;51: 6043–6049. doi:10.1021/jf034266b

13. Hosseini SMH, Emam-Djomeh Z, Razavi SH, Moosavi-Movahedi AA, Saboury AA, Atri MS, et al. β-lactoglobuline-sodium alginate interaction as affected by polysaccharide depolymerization using high intensity ultrasound. Food Hydrocoll. 2013;32: 235–244. doi:10.1016/j.foodhyd.2013.01.002

14. Weinbreck F, de Vries R, Schrooyen P, de Kruif CG. Complex coacervation of whey proteins and gum arabic. Biomacromolecules. 2003;4: 293–303. doi:10.1021/bm025667n

15. Stender EGP, Khan S, Ipsen R, Madsen F, Hägglund P, Hachem MA, et al. Effect of alginate size, mannuronic/guluronic acid content and pH on particle size, thermodynamics and composition of complexes with β-lactoglobulin. Food Hydrocoll. 2017;75: 157–163. doi:10.1016/j.foodhyd.2017.09.001

16. Comert F, Malanowski AJ, Azarikia F, Dubin PL. Coacervation and precipitation in polysaccharide-protein systems. Soft Matter. 2016;12: 4154–4161. doi:10.1039/c6sm00044d

17. Fuenzalida JP, Nareddy PK, Moreno-Villoslada I, Moerschbacher BM, Swamy MJ, Pan S, et al. On the role of alginate structure in complexing with lysozyme and application for enzyme delivery. Food Hydrocoll. 2016;53: 239–248. doi:10.1016/j.foodhyd.2015.04.017

18. Haug A, Larsen B, Smodsrod O. A study of constitution of alginic acid by partial acid hydrolosis.

Acta Chem Scand. 1966;20: 183-. doi:10.3891/acta.chem.scand.20-0183

19. Haug A, Larsen B, Smodsrod O. Studies on sequence of uronic acid residues in alginic acid. Acta Chem Scand. 1967;21: 691-. doi:10.3891/acta.chem.scand.21-0691

20. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for determination of sugars and related substances. Anal Chem. 1956;28: 350–356. doi:10.1021/ac60111a017

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. Proteomics Protoc Handb. 2005; 571–607. doi:10.1385/1592598900

22. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta-Proteins Proteomics. 2005;1751: 119–139. doi:10.1016/j.bbapap.2005.06.005

23. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRpipe - a multidimentional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6: 277–293. doi:10.1007/BF00197809

24. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct Funct Genet. 2005;59: 687–696. doi:10.1002/prot.20449

25. Garron M-L, Cygler M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology. 2010;20: 1547–1573. doi:10.1093/glycob/cwq122

26. Bock K, Pedersen C. A study of 13CH coupling constants in hexopyranoses. J Chem Soc, Perkin Trans 2. 1974; 293–297. doi:10.1039/p29740000293

27. Enthart A, Freudenberger JC, Furrer J, Kessler H, Luy B. The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings. J Magn Reson. 2008;192: 314–322. doi:10.1016/j.jmr.2008.03.009

382 28. Campa C, Holtan S, Nilsen N, Bjerkan TM, Stokke BT, Skjåk-Braek G. Biochemical analysis of the
383 processive mechanism for epimerization of alginate by mannuronan C-5 epimerase AlgE4. Biochem
384 J. 2004;381: 155–164. doi:10.1042/BJ20031265

385 29. Corrêa D, Ramos CHI. The use of circular dichroism spectroscopy to study protein folding, form and
386 function. African J Biochem Res. 2009;3: 164–173.

387 30. Harnsilawat T, Pongsawatmanit R, McClements DJ. Characterization of beta-lactoglobulin-sodium
388 alginate interactions in aqueous solutions: A calorimetry, light scattering, electrophoretic mobility
389 and solubility study. Food Hydrocoll. 2006;20: 577–585. doi:10.1016/j.foodhyd.2005.05.005

390 31. Tavel L, Andriot I, Moreau C, Guichard E. Interactions between β-lactoglobulin and aroma
391 compounds: different binding behaviors as a function of ligand structure. J Agric Food Chem.
392 2008;56: 10208–10217. doi:10.1021/jf801841u

393

Figures

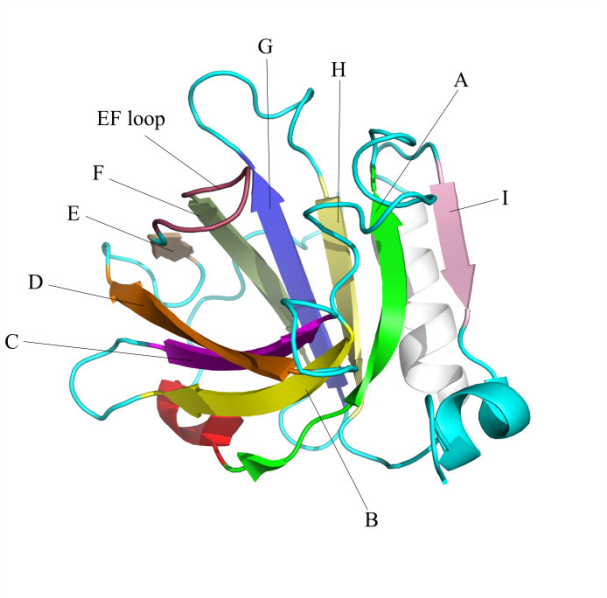


Figure 1) NMR structure of BLG by Uhrinova et al (PDB: 1DV9). β-strand A (green), β-strand B (yellow), β-strand C (purple), β-strand D (orange), β-strand E (grey), β-strand F (green gray), β-strand G (blue), β-strand H (light yellow), β-strand I (pink) and the EF loop in closed conformation (magenta).

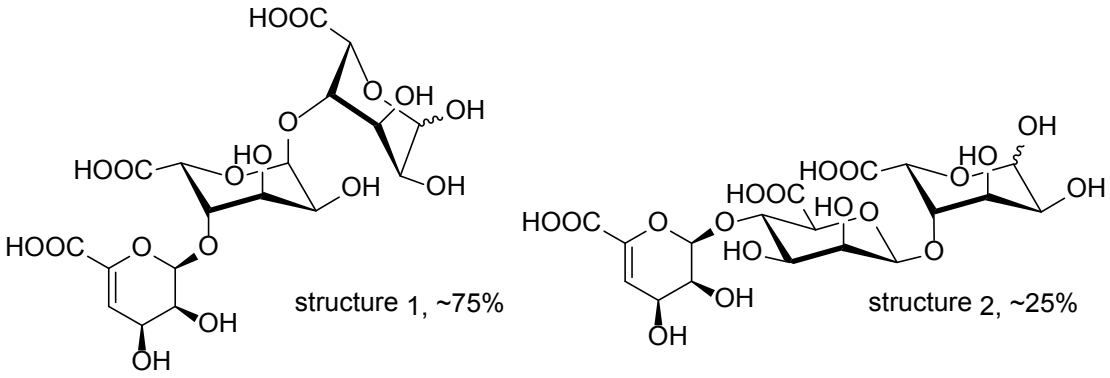


Figure 2) Structures of the alginate trisaccharides. The left structure is the major component in the mixture (~3/4)

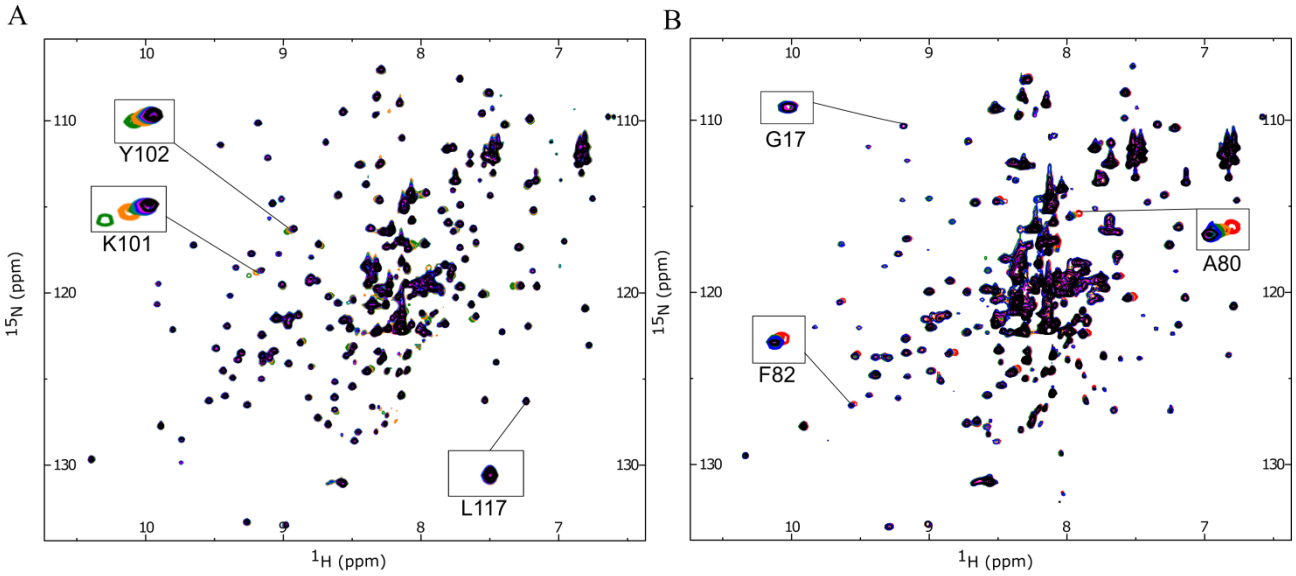


Figure 3) HSQC spectra of alginate titration of BLG WT at pH 2.65 with 0 mM AOS (black) and 15.5 mM AOS (green) (A) and pH 4.0 with 0 mM AOS (black) and 15.5 mM AOS (red) (B).

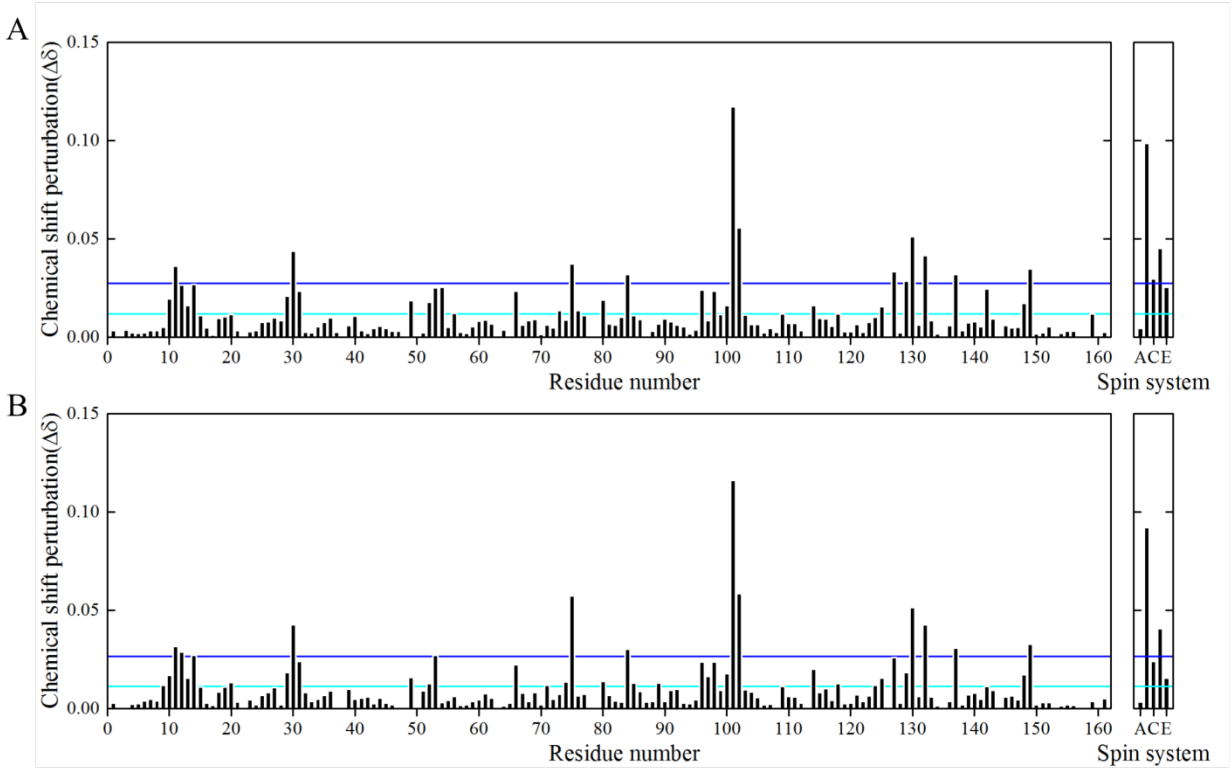


Figure 4) Chemical shift perturbation analysis of BLG WT (A) and BLG K75A (B) upon addition of ALG at pH 2.65. Cyan line is the average chemical shift and the blue line is the average chemical shift perturbation + one standard deviation.

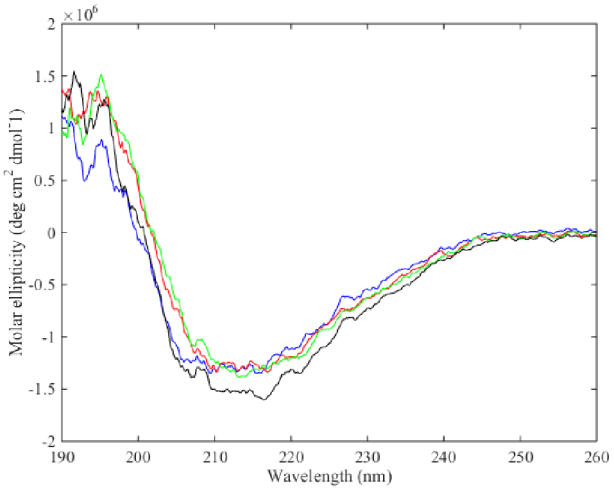


Figure 5) Far-UV CD spectra of BLG WT (black), BLG K75A (red), BLG K8A (green) and BLG K101A (Blue)

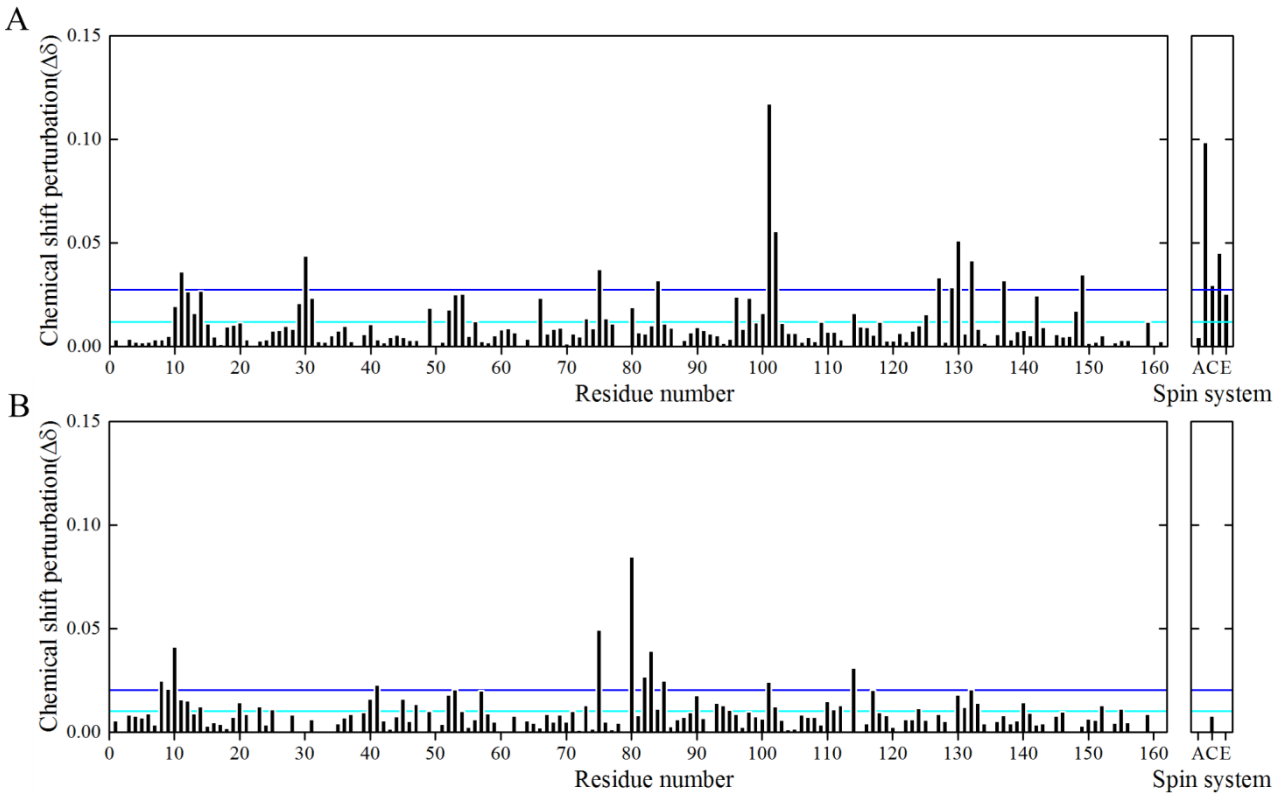


Figure 6) Chemical shift perturbation analysis of BLG WT-ALG interaction at pH 2.65 (A) and pH 4.0 (B). Cyan line is the average chemical shift and the blue line is the average chemical shift perturbation + one standard deviation.

Table 1 significant chemical shift perturbations in ^{1}H, ^{15}N-HSQC of BLG in the presence of 15.5 mM ALG. [x] selected peaks lost at pH 4.0.

	pH	K8	G9	L10	D11	I12	K14	S30	V41	D53	L57	K75	A80	F82	K83	D85	K101	Y102	E114	L117	E127	D129	D130	A132	D137	L149	{B}	{C}	{D}
BLG WT	2.65				x	x	x	x				x					x	x			x	x	x	x	x	x	x	x	x
BLG WT	4.0	x	x	x				[x]	x	x	x	x	x	x	x	x	x		x	x		[x]		x					
BLG K75A	2.65				x	x	x	x				x					x	x			x	x	x	x	x	x	x	x	x

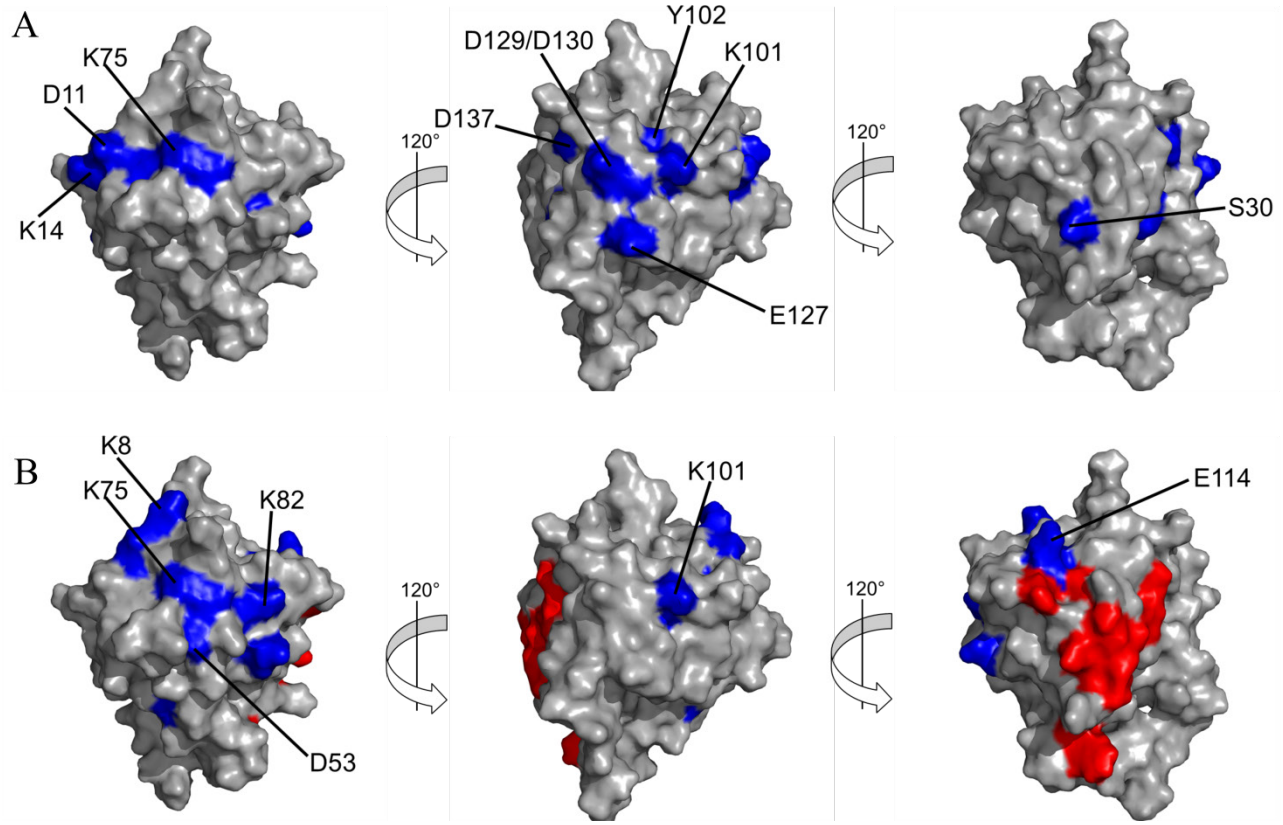
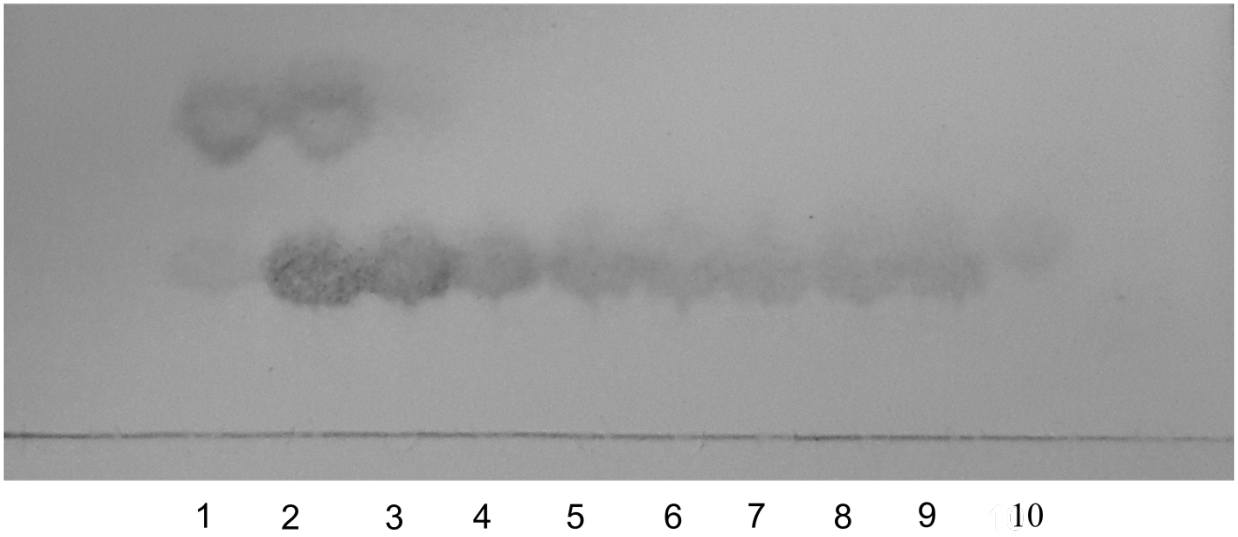


Figure 8) Mapping of significant chemical shift perturbations onto the surface of monomeric BLG WT at pH 2.65 and pH 4.0

Supplemental figures



Supporting figure S1) TLC of ALG purification (See methods). Numbers represent individual fractions from the HPLC purification. The upper band on the TLC represents the disaccharide product the lower the trisaccharides.

Supporting table S1) Assignment of the four trisaccharides

1β	$^{1}J_{H1,C1}$	1	2	3	4	5	6
Term **C**	171 hz	5.185 100.73	3.867 66.43	4.322 62.35	5.995 111.29	NA 141.73	NA 0
GulA mid **B**	171 hz	4.978 101.01	3.798 64.45	4.086 68.55	4.217 79.55	4.586 66.46	NA 0
b-GulA red **F**	164 hz	4.856 93.57	3.534 68.87	4.049 69.78	4.02 79.99	4.577 72.28	NA 0
1α	$^{1}J_{H1,C1}$	1	2	3	4	5	6
Term	171 hz	5.185	3.867	4.322	5.995	NA	NA 0

C		100.73	66.43	62.35	111.29	141.73	
GulA mid **B**	171 hz	4.978 101.01	3.798 64.45	4.086 68.55	4.217 79.55	4.586 66.46	NA 0
a-GulA red **G**	172 hz	5.18 93.03	3.823 69.39	0 0	4.123 79.97	4.794 65.86	NA 0
2β	$^{1}J_{H1,C1}$	1	2	3	4	5	6
Term **D**	171 hz	5.13 99.95	3.907 66.42	4.394 63.27	5.94 111.32	NA 141.73	NA 0
ManA mid **E**	159 hz	4.717 99.85	3.968 70.17	3.726 71.02	3.935 77.47	3.904 74.29	NA 0
b-GulA red **F**	164 hz	4.856 93.57	3.534 68.87	4.049 69.78	4.02 79.99	4.577 72.28	NA 0
2α	$^{1}J_{H1,C1}$	1	2	3	4	5	6
Term **D**	171 hz	5.13 99.95	3.907 66.42	4.394 63.27	5.94 111.32	NA 141.73	NA 0
ManA mid **A**	163 hz	4.696 101.54	3.972 69.05	3.726 71.02	3.935 77.47	3.87 74.36	NA 0
a-GulA red **G**	172 hz	5.18 93.03	3.823 69.39	0 0	4.123 79.97	4.794 65.86	NA 0

Supporting table S2) Assignment of chemical shifts at pH 2.65

Presented BLG assingment				Original peak list by Uhrínová, S et al.				Difference in chemical shift		
Position F1	Position F2	Assign F1	Assign F2	Position F1	Position F2	Assign F1	Assign F2	delta 1H	delta 15N	Delta-delta
8,714	127,01	-1AlaH	-1AlaN	8,66671	18,84152	-1AlaH	-1AlaN	0,04729	108,1685	21,63375
8,15643	124,5946	1AlaH	1AlaN	8,15671	124,796	1AlaH	1AlaN	-0,00028	-0,20139	0,040279
7,894	116,98	2TyrH	2TyrN	7,84671	116,836	2TyrH	2TyrN	0,04729	0,14404	0,055374
7,97976	120,0125	3ValH	3ValN	7,93671	120,166	3ValH	3ValN	0,04305	-0,15351	0,052876
8,06521	115,7838	4ThrH	4ThrN	8,01671	115,816	4ThrH	4ThrN	0,0485	-0,03218	0,048925
8,27049	123,7363	5GlnH	5GlnN	8,19671	123,486	5GlnH	5GlnN	0,07378	0,25033	0,089163
7,76571	111,5255	6ThrH	6ThrN	7,70671	111,026	6ThrH	6ThrN	0,059	0,49955	0,11603
7,74564	120,2697	7MetH	7MetN	7,74771	120,264	7MetH	7MetN	-0,00207	0,00578	0,002371
8,67298	127,6449	8LysH	8LysN	8,61671	127,316	8LysH	8LysN	0,05627	0,32896	0,086573
8,56641	109,4976	9GlyH	9GlyN	8,56871	109,51	9GlyH	9GlyN	-0,0023	-0,01233	0,003372
7,55358	120,1035	10LeuH	10LeuN	7,55271	120,04	10LeuH	10LeuN	0,00087	0,06356	0,012742
8,62551	126,7537	11AspH	11AspN	8,58671	126,786	11AspH	11AspN	0,0388	-0,03224	0,039332
8,87636	122,891	12IleH	12IleN	8,86971	122,854	12IleH	12IleN	0,00665	0,03701	0,00995
7,96968	117,1562	13GlnH	13GlnN	7,96671	117,145	13GlnH	13GlnN	0,00297	0,01122	0,003722
7,21671	113,686	14LysH	14LysN	7,21671	113,686	14LysH	14LysN	0	0	0
7,39204	110,2285	15Val	15Val	7,39071	110,21	15Val	15Val	0,00133	0,01851	0,003934

		H	N			H	N			
6,80901	120,9193	16AlaH	16AlaN	6,80871	120,877	16AlaH	16AlaN	0,0003	0,04237	0,008479
9,18302	110,1356	17GlyH	17GlyN	9,18271	110,115	17GlyH	17GlyN	0,00031	0,02062	0,004136
8,39327	118,4161	18ThrH	18ThrN	8,33671	118,116	18ThrH	18ThrN	0,05656	0,30016	0,08248
7,4529	124,024	19TrpH	19TrpN	7,45771	123,997	19TrpH	19TrpN	-0,00481	0,02701	0,007233
9,11053	112,1596	20TyrH	20TyrN	9,11071	112,1	20TyrH	20TyrN	-0,00018	0,05959	0,011919
9,90568	119,4652	21SerH	21SerN	9,90271	119,456	21SerH	21SerN	0,00297	0,00922	0,003496
7,7138	114,8999	23AlaH	23AlaN	7,70871	114,873	23AlaH	23AlaN	0,00509	0,02689	0,007405
9,07795	114,8068	24MetH	24MetN	9,07771	114,776	24MetH	24MetN	0,00024	0,03081	0,006167
9,30156	123,4667	25AlaH	25AlaN	9,29571	123,439	25AlaH	25AlaN	0,00585	0,02778	0,008068
8,31501	119,7417	26AlaH	26AlaN	8,30971	119,687	26AlaH	26AlaN	0,0053	0,05475	0,012165
8,36407	116,245	27SerH	27SerN	8,36071	116,204	27SerH	27SerN	0,00336	0,04101	0,008864
7,86704	119,7309	28AspH	28AspN	7,83671	119,656	28AspH	28AspN	0,03033	0,0749	0,033828
8,80004	124,2856	29IleH	29IleN	8,79271	124,22	29IleH	29IleN	0,00733	0,06567	0,015041
7,95915	114,1799	30SerH	30SerN	7,95671	114,158	30SerH	30SerN	0,00244	0,02194	0,005021
7,59857	117,0407	31LeuH	31LeuN	7,59471	117,005	31LeuH	31LeuN	0,00386	0,03576	0,008127
7,05274	111,2054	32LeuH	32LeuN	7,05271	111,156	32LeuH	32LeuN	3E-05	0,04947	0,009894

7,82273	118,3684	33AspH	33AspN	7,77671	118,306	33AspH	33AspN	0,04602	0,06242	0,047683
7,6322	120,6785	34AlaH	34AlaN	7,63271	120,642	34AlaH	34AlaN	-0,00051	0,03657	0,007332
9,05349	121,4988	35GlnH	35GlnN	9,01671	121,376	35GlnH	35GlnN	0,03678	0,12283	0,04423
7,71685	107,569	36SerH	36SerN	7,70971	107,522	36SerH	36SerN	0,00714	0,04704	0,011811
7,53554	126,2455	37AlaH	37AlaN	7,53471	126,215	37AlaH	37AlaN	0,00083	0,03056	0,006168
7,76613	114,0162	39LeuH	39LeuN	7,75571	113,949	39LeuH	39LeuN	0,01042	0,06724	0,017012
7,41082	122,2755	40ArgH	40ArgN	7,40971	122,227	40ArgH	40ArgN	0,00111	0,04855	0,009773
6,95828	117	41ValH	41ValN	6,95271	116,975	41ValH	41ValN	0,00557	0,02508	0,007496
8,60354	120,5741	42TyrH	42TyrN	8,59971	120,572	42TyrH	42TyrN	0,00383	0,00218	0,003855
8,62631	125,1627	43ValH	43ValN	8,62271	125,123	43ValH	43ValN	0,0036	0,03976	0,008729
8,96265	121,7778	44GluH	44GluN	8,91671	121,806	44GluH	44GluN	0,04594	-0,02815	0,046284
7,6186	116,2857	45GluH	45GluN	7,62071	116,223	45GluH	45GluN	-0,00211	0,06272	0,01272
8,66171	122,0424	46LeuH	46LeuN	8,66171	122,004	46LeuH	46LeuN	0	0,03846	0,007692
9,15635	122,4511	47LysH	47LysN	9,15771	122,424	47LysH	47LysN	-0,00136	0,02713	0,005594
8,59743	114,3018	49ThrH	49ThrN	8,60071	114,288	49ThrH	49ThrN	-0,00328	0,01386	0,004294
7,52999	112,0727	51GluH	51GluN	7,49671	112,086	51GluH	51GluN	0,03328	-0,0133	0,033386

8,28883	107,0215	52GlyH	52GlyN	8,28071	106,994	52GlyH	52GlyN	0,00812	0,02751	0,009808
7,18097	115,1938	53AspH	53AspN	7,17571	115,13	53AspH	53AspN	0,00526	0,06381	0,013803
8,80694	119,4394	54LeuH	54LeuN	8,79671	119,287	54LeuH	54LeuN	0,01023	0,15239	0,032149
8,62626	126,0635	55GluH	55GluN	8,62671	125,876	55GluH	55GluN	-0,00045	0,18757	0,037517
9,26173	126,5153	56IleH	56IleN	9,25971	126,498	56IleH	56IleN	0,00202	0,01737	0,004019
8,48296	128,6041	57LeuH	57LeuN	8,48571	128,546	57LeuH	57LeuN	-0,00275	0,05813	0,011947
9,06596	123,3919	58LeuH	58LeuN	9,06671	123,367	58LeuH	58LeuN	-0,00075	0,02498	0,005052
9,23144	117,7148	59GlnH	59GlnN	9,22371	117,655	59GlnH	59GlnN	0,00773	0,05983	0,014246
8,95487	121,5045	60LysH	60LysN	8,91671	121,496	60LysH	60LysN	0,03816	0,0085	0,038198
9,74021	129,856	61TrpH	61TrpN	9,73471	129,762	61TrpH	61TrpN	0,0055	0,09404	0,019596
8,74944	127,2512	62GluH	62GluN	8,67671	127,376	62GluH	62GluN	0,07273	-0,12476	0,076891
8,91417	121,6213	63AsnH	63AsnN	8,86671	121,476	63AsnH	63AsnN	0,04746	0,14538	0,055658
8,13732	110,9992	64AspH	64AspN	8,13871	110,966	64AspH	64AspN	-0,00139	0,03326	0,006796
6,75262	114,5227	65GluH	65GluN	6,75071	114,489	65GluH	65GluN	0,00191	0,03375	0,007015
8,67315	120,2401	66CysH	66CysN	8,66571	120,173	66CysH	66CysN	0,00744	0,06715	0,015353
8,97902	133,4899	67AlaH	67AlaN	8,97871	133,464	67AlaH	67AlaN	0,00031	0,0259	0,005189
8,38211	118,635			8,34671	118,636			0,035	-	0,03540

		68GlnH	68GlnN			68GlnH	68GlnN	4	0,00096	1
9,15043	123,6905	69LysH	69LysN	9,14771	123,65	69LysH	69LysN	0,00272	0,04051	0,008546
8,54061	123,6847	70LysH	70LysN	8,53471	123,623	70LysH	70LysN	0,0059	0,06175	0,013687
8,95827	124,1015	71IleH	71IleN	8,95371	124,008	71IleH	71IleN	0,00456	0,09353	0,019254
8,49789	125,4454	72IleH	72IleN	8,49371	125,402	72IleH	72IleN	0,00418	0,04339	0,009632
9,26399	133,2975	73AlaH	73AlaN	9,25771	133,246	73AlaH	73AlaN	0,00628	0,05153	0,012069
9,43616	124,5123	74GluH	74GluN	9,42871	124,432	74GluH	74GluN	0,00745	0,08031	0,017706
8,4843	123,1493	75LysH	75LysN	8,48471	123,032	75LysH	75LysN	-0,00041	0,11733	0,02347
8,44455	112,5842	76ThrH	76ThrN	8,43871	112,528	76ThrH	76ThrN	0,00584	0,05627	0,012679
8,27246	116,4517	77LysH	77LysN	8,21671	116,306	77LysH	77LysN	0,05575	0,1457	0,062906
7,97976	122,2947	78IleH	78IleN	7,97671	122,156	78IleH	78IleN	0,00305	0,1387	0,027907
7,85116	115,217	80AlaH	80AlaN	7,83971	115,155	80AlaH	80AlaN	0,01145	0,06201	0,016879
7,28916	117,3683	81ValH	81ValN	7,29071	117,327	81ValH	81ValN	-0,00155	0,04136	0,008416
9,53946	126,2751	82PheH	82PheN	9,53171	126,222	82PheH	82PheN	0,00775	0,05314	0,013154
9,48594	123,2371	83LysH	83LysN	9,48371	123,137	83LysH	83LysN	0,00223	0,10012	0,020148
7,7646	116,4437	84IleH	84IleN	7,72671	116,283	84IleH	84IleN	0,03789	0,16071	0,049687
8,79959	119,195	85AspH	85AspN	8,79471	119,212	85AspH	85AspN	0,00488	-0,01692	0,005939
7,22122	121,9317	86Ala	86Ala	7,21771	121,909	86Ala	86Ala	0,00351	0,02274	0,005745

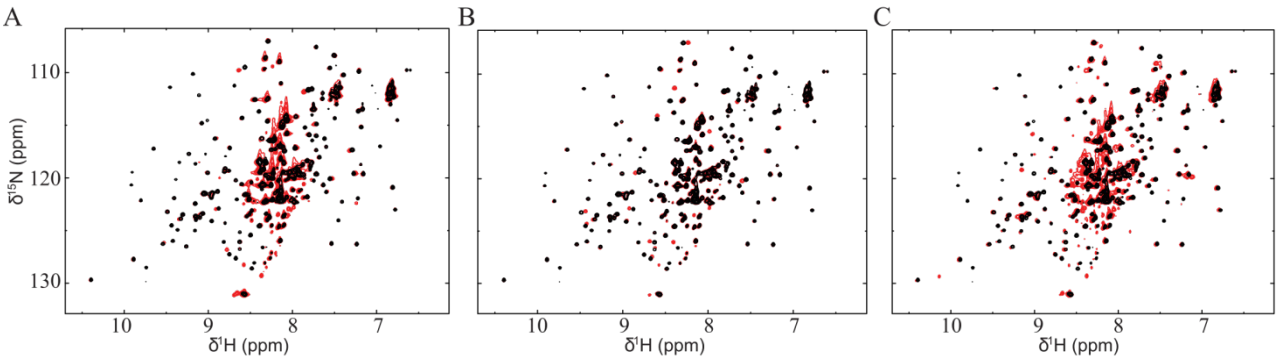
		H	N			H	N			
9,01196	114,5393	87LeuH	87LeuN	9,00271	114,586	87LeuH	87LeuN	0,00925	-0,04667	0,013141
8,68998	111,244	88AsnH	88AsnN	8,68071	111,222	88AsnH	88AsnN	0,00927	0,02205	0,010266
8,24907	117,3076	89GluH	89GluN	8,22671	117,226	89GluH	89GluN	0,02236	0,08164	0,027687
9,10042	115,6569	90AsnH	90AsnN	9,09471	115,64	90AsnH	90AsnN	0,00571	0,01696	0,006642
8,55857	121,6032	91LysH	91LysN	8,55671	121,486	91LysH	91LysN	0,00186	0,11721	0,023516
9,05227	121,2037	92ValH	92ValN	9,00671	121,296	92ValH	92ValN	0,04556	-0,09229	0,049157
9,42034	125,9924	93LeuH	93LeuN	9,42171	125,949	93LeuH	93LeuN	-0,00137	0,04346	0,008799
9,32618	123,8768	94ValH	94ValN	9,32071	123,826	94ValH	94ValN	0,00547	0,05085	0,011548
8,45414	128,07	95LeuH	95LeuN	8,45471	128,055	95LeuH	95LeuN	-0,00057	0,01507	0,003067
7,20926	109,8899	96AspH	96AspN	7,20571	109,887	96AspH	96AspN	0,00355	0,00298	0,0036
7,78325	118,8771	97ThrH	97ThrN	7,75671	118,816	97ThrH	97ThrN	0,02654	0,06118	0,029225
6,77772	123,0158	98AspH	98AspN	6,78371	123,026	98AspH	98AspN	-0,00599	-0,01015	0,006325
9,0333	117,7347	99TyrH	99TyrN	9,02871	117,67	99TyrH	99TyrN	0,00459	0,06475	0,013739
9,00755	119,5293	100LysH	100LysN	9,00471	119,499	100LysH	100LysN	0,00284	0,03034	0,0067
9,15464	118,6572	101LysH	101LysN	9,17471	118,672	101LysH	101LysN	-0,02007	-0,01474	0,020285
8,92234	116,313	102Tyr	102Tyr	8,93171	116,264	102Tyr	102Tyr	-0,009	0,04904	0,013564

		H	N			H	N	37		
9,15925	125,0007	103LeuH	103LeuN	9,15671	124,966	103LeuH	103LeuN	0,00254	0,03475	0,0074
9,36536	124,9315	104LeuH	104LeuN	9,36271	124,899	104LeuH	104LeuN	0,00265	0,0325	0,007019
8,8884	121,2799	105PheH	105PheN	8,86671	121,216	105PheH	105PheN	0,02169	0,06391	0,025176
9,65116	117,241	106CysH	106CysN	9,64471	117,185	106CysH	106CysN	0,00645	0,056	0,012924
9,91503	120,6674	107MetH	107MetN	9,91471	120,639	107MetH	107MetN	0,00032	0,02846	0,005701
8,39827	113,7012	108GluH	108GluN	8,39871	113,701	108GluH	108GluN	-0,00044	0,00023	0,000442
9,38562	120,1509	109AsnH	109AsnN	9,37971	120,102	109AsnH	109AsnN	0,00591	0,04891	0,011429
9,80254	122,1423	110SerH	110SerN	9,79871	122,111	110SerH	110SerN	0,00383	0,03133	0,007344
8,31275	123,4047	111AlaH	111AlaN	8,28671	123,226	111AlaH	111AlaN	0,02604	0,17874	0,044227
7,74809	113,4934	112GluH	112GluN	7,74071	113,434	112GluH	112GluN	0,00738	0,05941	0,013987
8,29799	114,6368	114GluH	114GluN	8,29471	114,629	114GluH	114GluN	0,00328	0,00783	0,003635
7,80129	115,9036	115GlnH	115GlnN	7,75671	115,736	115GlnH	115GlnN	0,04458	0,16766	0,055783
7,55463	109,5774	116SerH	116SerN	7,55771	109,554	116SerH	116SerN	-0,00308	0,02342	0,005606
7,23519	126,2904	117LeuH	117LeuN	7,23071	126,251	117LeuH	117LeuN	0,00448	0,03939	0,009063
8,81571	126,0732	118ValH	118ValN	8,80571	126,07	118ValH	118ValN	0,01	0,00327	0,010021
9,40553	121,9189	119Cy	119Cy	9,40371	121,903	119Cy	119Cy	0,00182	0,01591	0,003666

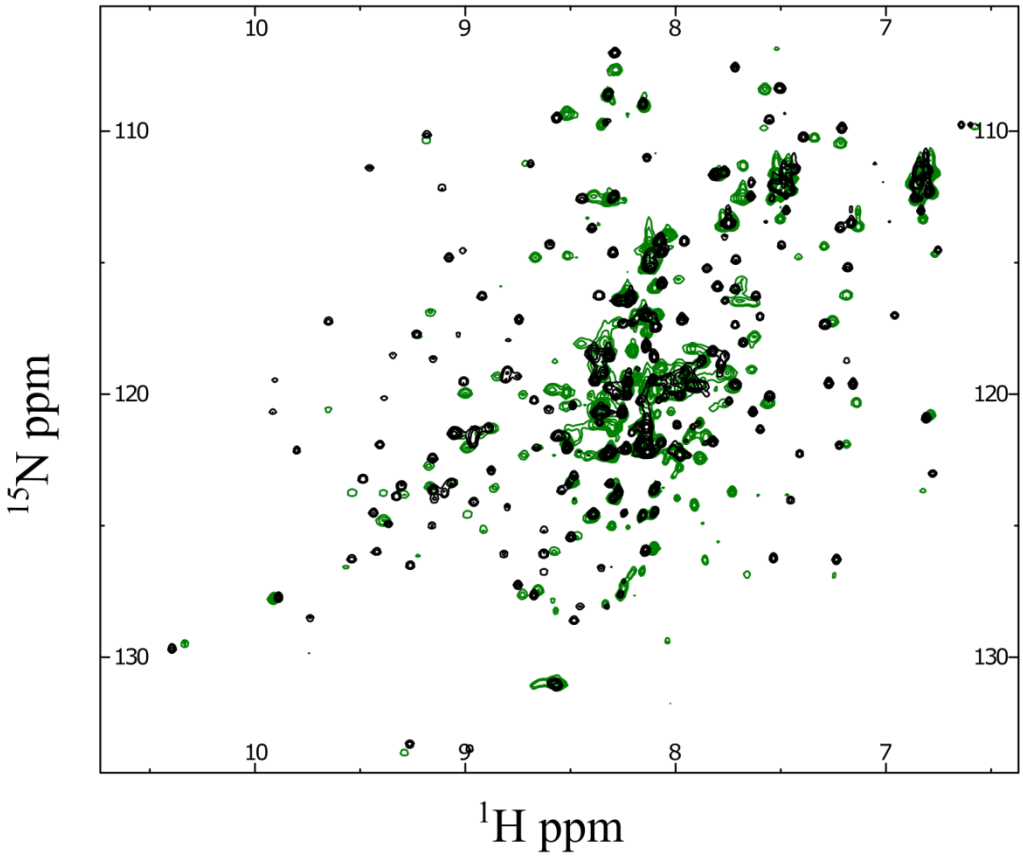
		sH	sN			sH	sN			
9,34365	118,5203	120GlnH	120GlnN	9,32671	118,446	120GlnH	120GlnN	0,01694	0,07436	0,022542
7,85199	119,9971	121CysH	121CysN	7,84671	120,076	121CysH	121CysN	0,00528	-0,07889	0,016638
9,73782	128,505	122LeuH	122LeuN	9,73771	128,487	122LeuH	122LeuN	0,00011	0,01802	0,003606
9,45469	111,3953	123ValH	123ValN	9,45371	111,378	123ValH	123ValN	0,00098	0,01733	0,003602
7,92677	119,6724	124ArgH	124ArgN	7,90671	119,556	124ArgH	124ArgN	0,02006	0,11644	0,030737
7,50602	108,3734	125ThrH	125ThrN	7,50771	108,354	125ThrH	125ThrN	-0,00169	0,01948	0,004247
8,28294	119,9774	127GluH	127GluN	8,25671	119,856	127GluH	127GluN	0,02623	0,12142	0,035745
8,34333	119,1833	128ValH	128ValN	8,30671	118,956	128ValH	128ValN	0,03662	0,22731	0,058377
6,54905	123,9292	129AspH	129AspN	6,54971	123,869	129AspH	129AspN	-0,00066	0,06024	0,012066
9,099	123,7389	130AspH	130AspN	9,08971	123,691	130AspH	130AspN	0,00929	0,04794	0,01335
8,36184	121,0666	131GluH	131GluN	8,32671	120,876	131GluH	131GluN	0,03513	0,19064	0,051845
7,59799	121,3508	132AlaH	132AlaN	7,59871	121,317	132AlaH	132AlaN	-0,00072	0,03383	0,006804
7,71521	116,0186	133LeuH	133LeuN	7,72671	116,256	133LeuH	133LeuN	-0,0115	-0,23733	0,048839
8,03054	119,4525	134GluH	134GluN	8,01671	119,626	134GluH	134GluN	0,01383	-0,17346	0,037347
7,96565	119,7573	135LysH	135LysN	7,91671	119,616	135LysH	135LysN	0,04894	0,14129	0,056512
8,51648	121,6446	136Phe	136Phe	8,51671	121,446	136Phe	136Phe	-0,000	0,19859	0,039719

		H	N			H	N	23		
8,74399	117,1587	137AspH	137AspN	8,74171	117,114	137AspH	137AspN	0,00228	0,04477	0,00924
7,95303	119,1884	138LysH	138LysN	7,90671	119,366	138LysH	138LysN	0,04632	-0,17753	0,058363
7,82219	121,8021	139AlaH	139AlaN	7,81871	121,75	139AlaH	139AlaN	0,00348	0,05216	0,010997
7,49697	114,3457	140LeuH	140LeuN	7,49171	114,282	140LeuH	140LeuN	0,00526	0,0637	0,013783
7,15578	119,6065	141LysH	141LysN	7,15371	119,582	141LysH	141LysN	0,00207	0,02458	0,005334
7,76552	118,5511	142AlaH	142AlaN	7,77671	118,306	142AlaH	142AlaN	-0,01119	0,24512	0,050285
7,71446	119,6259	143LeuH	143LeuN	7,71171	119,606	143LeuH	143LeuN	0,00275	0,01989	0,004836
7,67849	118,0502	145MetH	145MetN	7,67171	118,024	145MetH	145MetN	0,00678	0,02625	0,008575
8,79564	117,935	146HisH	146HisN	8,78971	117,917	146HisH	146HisN	0,00593	0,01806	0,006943
7,71735	117,3738	147IleH	147IleN	7,71471	117,335	147IleH	147IleN	0,00264	0,03883	0,008202
8,35305	126,5955	148ArgH	148ArgN	8,34471	126,557	148ArgH	148ArgN	0,00834	0,03853	0,011355
9,14659	123,9575	149LeuH	149LeuN	9,14471	123,935	149LeuH	149LeuN	0,00188	0,02257	0,00489
8,21158	116,3462	150SerH	150SerN	8,16671	116,306	150SerH	150SerN	0,04487	0,04019	0,045584
8,10716	119,4649	151PheH	151PheN	8,05671	119,806	151PheH	151PheN	0,05045	-0,34104	0,084838
8,75075	119,3184	152AsnH	152AsnN	8,74671	119,306	152AsnH	152AsnN	0,00404	0,01245	0,004746
8,12013	115,1838	154Thr	154Thr	8,11671	115,566	154Thr	154Thr	0,00342	-0,3821	0,076512

		H	N			H	N		8	
8,32129	122,2797	155GlnH	155GlnN	8,31171	122,264	155GlnH	155GlnN	0,00958	0,01578	0,010086
8,13395	122,2049	156LeuH	156LeuN	8,17671	122,466	156LeuH	156LeuN	-0,04276	-0,26103	0,067482
8,24512	124,5139	159GlnH	159GlnN	8,23671	124,826	159GlnH	159GlnN	0,00841	-0,31209	0,062982
8,27406	121,4658	160CysH	160CysN	8,22671	121,306	160CysH	160CysN	0,04735	0,15979	0,057126
8,16857	120,2484	161HisH	161HisN	8,13671	119,896	161HisH	161HisN	0,03186	0,35247	0,077359



Supporting figure S2) ^{15}N-HSQC of BLG WT (black) overlaid with ^{15}N-HSQC of mutant protein K8A (A) K75A (B) K101A (C) (in red)



Supporting figure S3) Overlaid HSQC spectra of BLG WT recorded at pH 2.65 (black) and pH 4.0 (green)

Supporting table S3) Peak assignment of BLG WT at pH 4.

Position F1	Position F2	Assign F1	Assign F2
8,714	127,01	-1AlaH	-1AlaN
8,15249	124,60326	None	1AlaN
7,894	116,97934	2TyrH	2TyrN
7,98107	120,03856	3ValH	3ValN
8,09869	115,95741	4ThrH	4ThrN
8,28391	123,95411	5GlnH	5GlnN
7,67919	111,31498	6ThrH	6ThrN
7,76694	120,3676	7MetH	7MetN
8,6534	127,46774	8LysH	8LysN

8,5181	109,36035	9GlyH	9GlyN
7,56091	120,36403	10LeuH	10LeuN
8,58382	127,84335	11AspH	11AspN
8,72169	122,32619	12IleH	12IleN
8,09266	117,00496	13GlnH	13GlnN
7,29469	114,37819	14LysH	14LysN
7,34103	110,25788	15ValH	15ValN
6,78999	120,80588	16AlaH	16AlaN
9,18423	110,34642	17GlyH	17GlyN
8,335	118,06956	18ThrH	18ThrN
7,47022	123,8326	19TrpH	19TrpN
9,16544	112,35702	20TyrH	20TyrN
9,91494	119,60535	21SerH	21SerN
7,79469	115,09779	23AlaH	23AlaN
9,06504	114,67118	24MetH	24MetN
9,323	123,69839	25AlaH	25AlaN
8,30617	119,70007	26AlaH	26AlaN
8,40461	116,21134	27SerH	27SerN
7,87439	119,58716	28AspH	28AspN
8,75492	124,25854	29IleH	29IleN
7,95915	114,1799	30SerH	30SerN
7,6256	117,82315	31LeuH	31LeuN
7,05274	111,20543	32LeuH	32LeuN
7,86727	118,36782	33AspH	33AspN
7,64566	120,76635	34AlaH	34AlaN
9,02323	121,55195	35GlnH	35GlnN
7,52043	106,86304	36SerH	36SerN
7,55159	126,31237	37AlaH	37AlaN
7,71813	114,00216	39LeuH	39LeuN
7,33993	122,0839	40ArgH	40ArgN
6,93893	117,0467	41ValH	41ValN
8,61541	120,42029	42TyrH	42TyrN
8,62408	125,05276	43ValH	43ValN
8,99272	122,04012	44GluH	44GluN
7,63043	116,53426	45GluH	45GluN
8,64507	122,0205	46LeuH	46LeuN
9,17446	122,7257	47LysH	47LysN
8,66702	114,80277	49ThrH	49ThrN
7,51061	112,13326	51GluH	51GluN
8,2869	107,66962	52GlyH	52GlyN
7,18685	116,24506	53AspH	53AspN
8,84793	119,32414	54LeuH	54LeuN
8,57733	125,9837	55GluH	55GluN

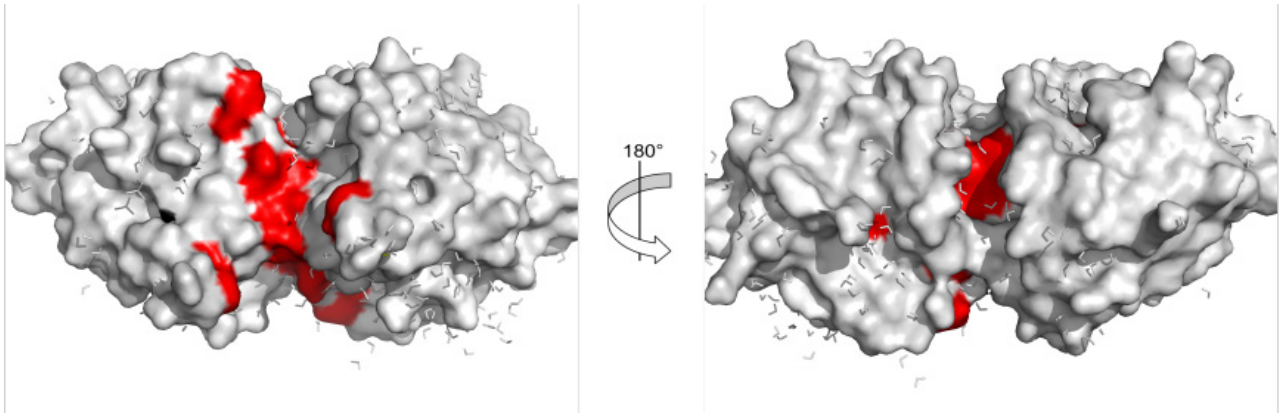
9,22524	126,13213	56IleH	56IleN
8,5193	128,69465	57LeuH	57LeuN
9,05892	123,36032	58LeuH	58LeuN
9,22422	117,78442	59GlnH	59GlnN
8,92294	121,5005	60LysH	60LysN
9,7433	129,80822	61TrpH	61TrpN
8,72787	127,63268	62GluH	62GluN
8,91417	121,62134	63AsnH	63AsnN
8,11994	110,97099	64AspH	64AspN
6,76569	114,6642	65GluH	65GluN
8,58824	119,84009	66CysH	66CysN
9,00822	133,41585	67AlaH	67AlaN
8,37675	118,73442	68GlnH	68GlnN
9,16777	123,52064	69LysH	69LysN
8,49975	123,38297	70LysH	70LysN
8,98812	124,57064	71IleH	71IleN
8,4856	125,42674	72IleH	72IleN
9,29059	133,63675	73AlaH	73AlaN
9,40332	124,79317	74GluH	74GluN
8,86128	123,56929	75LysH	75LysN
8,38966	112,48585	76ThrH	76ThrN
8,27008	116,36791	77LysH	77LysN
7,97952	122,31686	78IleH	78IleN
7,98791	115,6413	80AlaH	80AlaN
7,2541	117,25123	81ValH	81ValN
9,56952	126,57629	82PheH	82PheN
9,53814	123,74872	83LysH	83LysN
7,75112	116,38482	84IleH	84IleN
8,72482	120,01021	85AspH	85AspN
7,18809	121,90603	86AlaH	86AlaN
9,00004	114,55674	87LeuH	87LeuN
8,71409	111,23332	88AsnH	88AsnN
8,27401	117,27395	89GluH	89GluN
9,08647	115,52173	90AsnH	90AsnN
8,5147	121,59359	91LysH	91LysN
9,03815	121,21461	92ValH	92ValN
9,44071	125,98654	93LeuH	93LeuN
9,28846	123,81946	94ValH	94ValN
8,48537	128,06782	95LeuH	95LeuN
7,21482	110,46173	96AspH	96AspN
7,75403	119,1947	97ThrH	97ThrN
6,82303	123,66726	98AspH	98AspN
9,00943	117,5086	99TyrH	99TyrN
9,00086	119,97522	100LysH	100LysN
9,6492	120,5814	101LysH	101LysN

9,1665	116,88936	102TyrH	102TyrN
9,15791	124,91022	103LeuH	103LeuN
9,38556	124,80012	104LeuH	104LeuN
8,86883	121,31019	105PheH	105PheN
9,6324	117,19327	106CysH	106CysN
9,94944	120,6503	107MetH	107MetN
8,36756	113,5401	108GluH	108GluN
9,31585	120,13682	109AsnH	109AsnN
9,82524	122,02011	110SerH	110SerN
8,28885	123,42175	111AlaH	111AlaN
7,77428	113,62941	112GluH	112GluN
8,51444	114,7389	114GluH	114GluN
7,78218	116,04151	115GlnH	115GlnN
7,57996	109,8805	116SerH	116SerN
7,24807	126,90872	117LeuH	117LeuN
8,8264	125,78253	118ValH	118ValN
9,5114	122,05082	119CysH	119CysN
9,32355	118,48955	120GlnH	120GlnN
7,81454	120,03302	121CysH	121CysN
9,73866	128,56131	122LeuH	122LeuN
9,4416	111,54708	123ValH	123ValN
7,93195	119,76318	124ArgH	124ArgN
7,57508	108,4092	125ThrH	125ThrN
8,34175	120,41194	127GluH	127GluN
8,34697	119,25082	128ValH	128ValN
6,54905	123,9292	129AspH	129AspN
8,91334	125,16318	130AspH	130AspN
8,32864	121,14207	131GluH	131GluN
7,8827	122,45506	132AlaH	132AlaN
7,69738	116,45833	133LeuH	133LeuN
8,05092	119,54641	134GluH	134GluN
7,96565	119,75725	135LysH	135LysN
8,51634	121,87362	136PheH	136PheN
8,57401	118,74543	137AspH	137AspN
7,95684	119,32023	138LysH	138LysN
7,85513	121,68615	139AlaH	139AlaN
7,49057	114,77317	140LeuH	140LeuN
7,16841	119,78887	141LysH	141LysN
7,755	118,53939	142AlaH	142AlaN
7,72371	119,72597	143LeuH	143LeuN
7,63753	118,12747	145MetH	145MetN
8,81493	117,8165	146HisH	146HisN
7,70757	117,38947	147IleH	147IleN
8,34201	126,81059	148ArgH	148ArgN
9,20937	125,01044	149LeuH	149LeuN

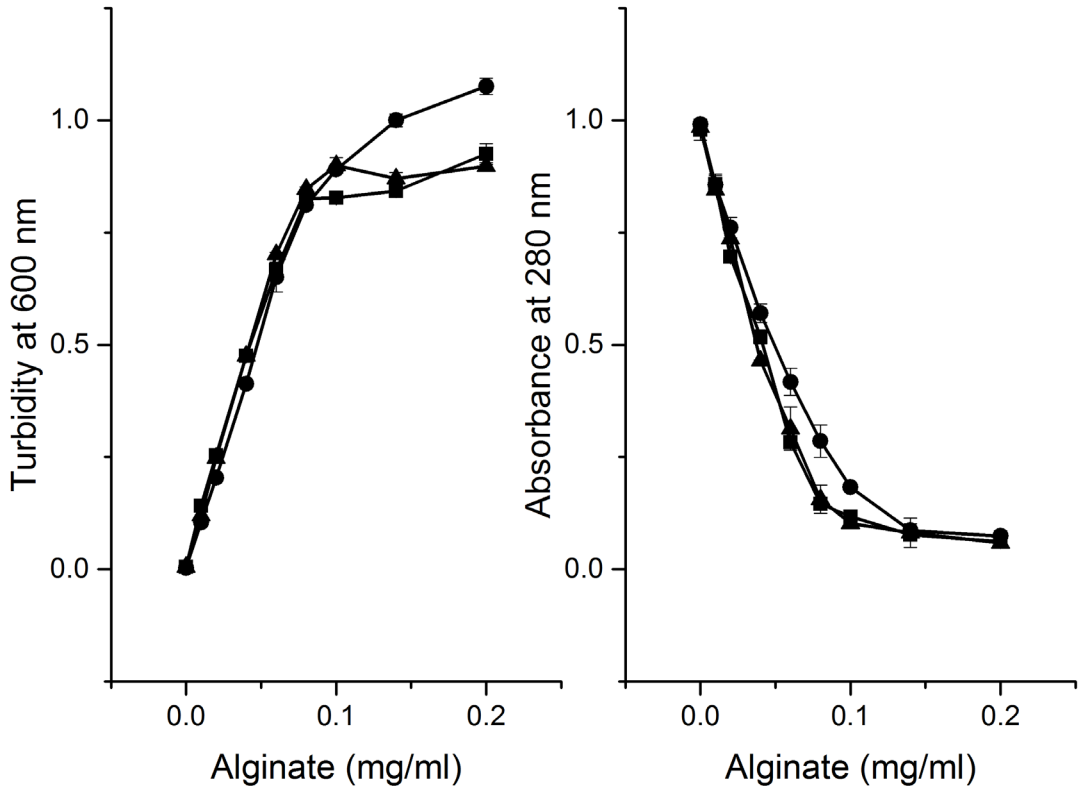
8,22909	116,33363	150SerH	150SerN
8,14169	119,74648	151PheH	151PheN
8,7258	119,25046	152AsnH	152AsnN
8,11869	115,19077	154ThrH	154ThrN
8,30967	122,34969	155GlnH	155GlnN
8,12967	122,20482	156LeuH	156LeuN
8,30288	125,02645	159GlnH	159GlnN
8,27406	121,46575	160CysH	160CysN
8,17986	120,29359	161HisH	161HisN
8,15198	109,00202	{153}H[306]	{153}N[305]
7,99379	121,15781	{159}H[318]	{159}N[317]
8,38869	124,6487	{167}[340]	{167}N[338]
8,10204	117,50465	{168}H[341]	{168}N[342]
8,13905	121,24688	{169}H[343]	{169}N[344]

Supporting table S3) Unassigned residues and peaks disappearing in the pH titration.

Unassigned peaks		Disappearing peaks during pH titration
Y2	38Pro	A26
L22	48Pro	S27
N63	50Pro	I29
D129	79Pro	S30
K135	113Pro	L32
E157	126Pro	D33
E158	144Pro	A34
C160	153Pro	W61
I162		V92
		Q115
		C121
		D129
		I147
		R148
		H161



Supporting figure S4) peaks disappearing (red) during the pH titration mapped to dimeric BLG WT (PDB: 1BEB).



Supporting figure S5) Turbidity and absorbance of supernatant of BLG WT and mutant proteins mixed with varying concentration of LMW. BLG WT (circle), K75A (square) and K8A (triangle).

Paper 4

Discovery of novel intermediates of *lacZ* β-galactosidase catalyzed by dDNP NMR

Kjeldsen, C.; Ardenkjær-Larsen, J. H.; Duus, J. Ø.

Manuscript in preparation

Discovery of novel intermediates of *lacZ* β-galactosidase catalyzed hydrolysis using dDNP NMR

Christian Kjeldsen[a], Jan H. Ardenkjær-Larsen[b] and Jens Ø. Duus*[a]

Abstract: Using dissolution dynamic nuclear polarization, the sensitivity of single scan solution state ^{13}C NMR can be improved up to four order of magnitude. In this study, the enzyme *lacZ* β-galactosidase from *E. coli.* was subjected to hyperpolarized substrate, and previously unknown reaction intermediates were observed, including 1,1-linked disaccharide. The enzyme is known for making 1,6-transglycosylation producing products like allolactose, that are also substrates. To analyze the kinetics, a simple kinetic model was developed and used to determine relative transglycosylation and hydrolysis rates of each of the intermediates, and the novel transglycosylation intermediates were determined as better substrates than the 1,6-linked one, explaining their transient nature. These findings suggest that hydrolysis and transglycosylation might be more complex than previously described.

As intermediates in enzymatic reactions are often low-populated and short lived they can be quite difficult to detect and even more difficult to quantify or identify. For carbohydrates that is often even more so the case, as they do not have any fluorophores or chromophores, unless added synthetically, which yields little to no information on intermediates, but can be used to determine kinetics. However, while NMR spectroscopy is an excellent tool for identifying carbohydrate structures, it suffers from low sensitivity compared to most other analytical methods, resulting in long acquisition times. Using dissolution Dynamic Nuclear Polarization (dDNP) the sensitivity of single scan solution NMR can be increased by up to four orders of magnitude.[1]

Conventional NMR spectroscopy has been used to study enzymatic reactions at equilibrium or for competition studies,[2] and for glycosidases it is possible to determine whether the enzyme has a retaining or inverting mechanism.[3] Using dDNP it is possible to observe low-populated reaction intermediates of enzymatic reactions,[4] and applying this for glycosidases represents the possibility to further understand the mechanisms and transglycosylation abilities of the enzymes.

However, most carbohydrates possess poor properties for hyperpolarization, as they have few or no quaternary positions with long T_1 relaxation. Exceptions to this are ketoses such as fructose, in which the anomeric position, which is also usually the position of interest, is quaternary with a long T_1, and consequently fructose has been used for *in vivo* murine studies.[5] For aldoses it is necessary to utilize ^{2}H labelling to elongate the T_1 of the positions of interest, and [U-^{2}H;U-^{13}C]glucose has been used for several studies,[6–10] but only one of these focused on the mechanisms of a single enzyme, namely phosphorylation of glucose by hexokinase.[9]

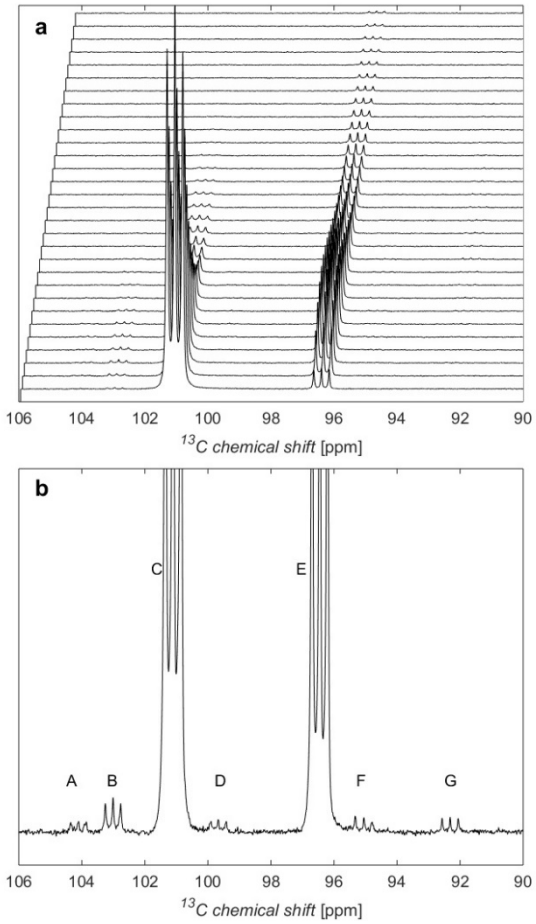


Figure 1. a: The first 30 spectra of doubly labelled Galp-onp enzymatic hydrolysis. The repetition time is 2 s and each spectrum is shifted 0.06 ppm upfield. **b**: The sum of spectra 5 to 10. A: β-Galp-1,3/4-Gal, B: β-Galp-1,6-Gal, C: β-Galp-onp, D: β-Galp-1,1-β-Galp, E: β-Galp, F: α-Galf and G: α-Galp.

Here we demonstrate the utility of dDNP NMR for increasing the understanding of carbohydrate converting enzymes, and the enzyme chosen as the model enzyme is the *lacZ* β-galactosidase, as it is well characterized.[11] The natural substrate is probably lactose,[12] but *ortho*- or *para*-nitrophenyl β-D-galactopyranosides are often used in biochemical studies as they are good substrates and release nitrophenols detectable at low concentrations using UV/VIS spectroscopy. This also suits our method quite nicely, as the nitrophenyl contains quaternary carbons enabling experiment optimization with natural

[a] C. Kjeldsen, Prof. J. Ø. Duus
Department of Chemistry
Technical University of Denmark
Kemitorvet, Bldg. 207, room 212, 2800 Kgs. Lyngby (DK)
E-mail: jduus@kemi.dtu.dk
[b] Prof. J. H. Ardenkjær-Larsen
Department of Electrical Engineering, Center for Hyperpolarization in Magnetic Resonance
Technical University of Denmark
Ørsteds Plads, Bldg. 349, room 126, 2800 Kgs. Lyngby (DK)

Supporting information for this article is given via a link at the end of the document.

abundance before using labelled substrate. The enzyme is known to do transglycosylation to form allolactose from lactose, and the reported transglycosylation products are only 1,6-linked.[11,13,14] Theoretically this was found to be the only viable product with glucose as acceptor.[15]

Using natural abundance *o*-nitrophenyl β-D-galactopyranoside it was possible to optimize both the hyperpolarization conditions for dDNP and the subsequent enzymatic measurements. Following this, the use of doubly labelled *o*-nitrophenyl β-D-[1-^{13}C;1-^{2}H]galactopyranoside (Galp-onp) as substrate allowed the monitoring of the anomeric position of the galactosides, as shown in figure 1. The labelled substrate was synthesized in four steps from labelled galactose (Omicron).

At these conditions, shown in figure 1, the enzyme is primarily hydrolysing the substrate, as there is limited access to acceptor. To promote the transglycosylation activity of the enzyme, an excess of galactose was added to act as acceptor, resulting in a large increase in the signals corresponding to non-reducing β-galactopyranosides, see figure 2.

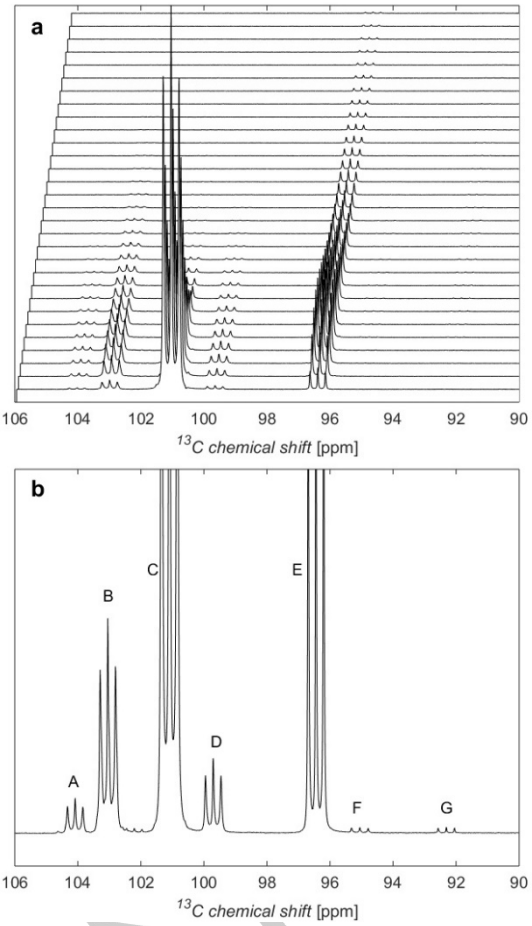


Figure 2. a: The first 30 spectra of doubly labelled Galp-onp enzymatic hydrolysis with added acceptor. The repetition time is 2 s and each spectrum is shifted 0.06 ppm upfield. **b**: The sum of spectra 1 to 10. The labels are the same as those in figure 1b.

The three signals with increased intensity, labelled as **A**, **B** and **D,** were assigned based on chemical shift comparison with literature. The largest of the three signals, **B**, was determined to be from β-Galp-1,6-Gal,[16] which was also the expected product. Slightly downfield of this signal is **A**, which was determined to arise from either β-Galp-1,3-Gal, β-Galp-1,4-Gal or both, as the anomeric position from both linkages have very similar chemical shift,[17] and is thus referred to as a 1,3/4-linkage. The last of the three major transglycosylation product peaks, **D**, was quite far upfield for a β-Galp, and was determined as having a 1,1-linkage.[18] To confirm this, 1-*O*-methylated galactopyranoside was used as acceptor, which caused the signal corresponding to the 1,1-linkage to almost disappear from the spectra, as shown in the supporting information figure S1.

Whether free galactose or one of the β-galactosides acts as the acceptor should have very limited effect upon the chemical shift of the donor or acceptor anomeric position, with the exception of the 1,1-linkage. Furthermore, the anomeric configuration of the acceptor should have little to no impact on the chemical shift of the donor anomeric position, again with the 1,1-linkage being the exception, where a β-anomeric glycopyranoside would be expected to have quite a large difference in chemical shift depending on the anomeric configuration of the aglycon of the 1,1-linkage.[18] Following this, it stands to reason that the 1,1-linked β-Galp labelled as **D** in figure 1 and 2 would indeed be a β-Galp-1,1-β-Galp. While the experiment with the added acceptor could have either α- or β-Galp as acceptor, the experiment performed with no added acceptor only have minute amounts of α-Galp until the mutarotation have had the time to equilibrate, and thus only β-Galp-1,1-β-Galp can be formed on the time-scale of the experiment.

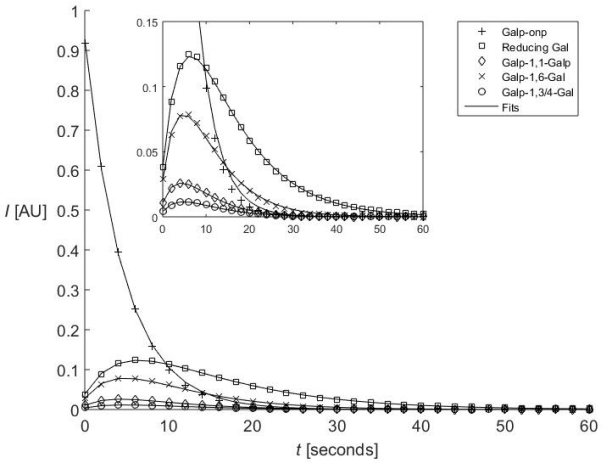


Figure 3. Peak integrals (in arbitrary units) of the five major species plotted over time with the models described in scheme 1 and 2 added. The insert is a zoom.

By plotting the integrals of each peak as a function of time, shown in figure 3, the relative transglycosylation rates of the three intermediate products can be determined by comparing the maximum intensity of each, which leads to the approximation that the 1,6-linkage is formed 3.0 times faster than the 1,1-linkage and 6.9 times faster than the 1,3/4-linkage. However, the hydrolysis rates are a bit more difficult to determine, as the

signal loss due to hydrolysis and polarization loss are both exponential decay. To extract approximate hydrolysis rates from the data, a kinetic model of the enzyme reaction was developed. Based on the seven reactions in scheme 1 the five equations in scheme 2 was devised. It is only the hyperpolarized doubly labelled anomeric positions that can be observed, and as such the results does not differentiate between di-, tri-, or oligosaccharides, as they would overlap in chemical shift, and thus whenever a reducing end is present in the model, it might be from free galactose or from a larger sugar.

*Galp-onp + $H_2O \xrightarrow{k1}$ *Gal + onp;

*Galp-onp + Gal $\xrightarrow{k2}$ *Galp-1,1-Galp + onp;

*Galp-onp + Gal $\xrightarrow{k3}$*Galp-1,6-Gal + onp;

*Galp-onp + Gal $\xrightarrow{k4}$ *Galp-1,3/4-Gal + onp;

*Galp-1,1-Galp + $H_2O \xrightarrow{k5}$ *Gal + Gal;

*Galp-1,6-Gal + $H_2O \xrightarrow{k6}$ *Gal + Gal;

*Galp-1,3/4-Gal + $H_2O \xrightarrow{k7}$ *Gal + Gal

Scheme 1. The seven reactions used for the kinetics model. The * marks the doubly labelled anomeric position used for monitoring. All non-reducing galactosides are β.

d[Galp-onp]/dt = - [Galp-onp]·(1/T_{1a}1+k1+k2+k3+k4)

d[Galp-1,1-Galp]/dt = [Galp-onp]·k2 - [Galp-1,1-Galp]·(1/T_{1a}2+k5)

d[Galp-1,6-Gal]/dt = [Galp-onp]·k3 - [Galp-1,6-Gal]·(1/T_{1a}3+k6)

d[Galp-1,3/4-Gal]/dt = [Galp-onp]·k4 - [Galp-1,3/4-Gal]·(1/T_{1a}4+k7)

d[Gal]/dt = [Galp-onp]·k1+[Galp-1,1-Gal]·k5+[Galp-1,6-Gal]·k6+[Galp-1,3/4-Gal]·k7 - [Gal]·1/T_{1a}5

Scheme 2. The five equations used for the model. T_{1a} is the apparent T_1 relaxations, which includes the polarization loss from NMR flip angles.

From these equations, the best fit to each of the five dataseries was found. The apparent T_1 relaxation for the reducing Gal was constrained by measuring it on the last few data points after the signal strength of the different substrates are gone, which gave values between 10.5 and 11.4 s across the repetitions. Similarly, the apparent T_1 relaxations for the four substrates are constrained to 7.5-8.5 s, based on measurements on Galp-onp without enzyme. Finally, any transglycosylation forming one disaccharide from another has been disregarded in the model. In figure 3 the fits are plotted alongside the data, and in figure 4 the fits are plotted with the contributions from the decay of polarization removed, meaning only the enzymatic reactions are left. Using the fit it was possible to verify the relative transglycosylation rates and to determine that the β-Galp-1,1-β-Galp was a far better substrate compared to β-Galp-1,6-Gal or β-Galp-1,3/4-Gal, which was hydrolyzed at similar rates.

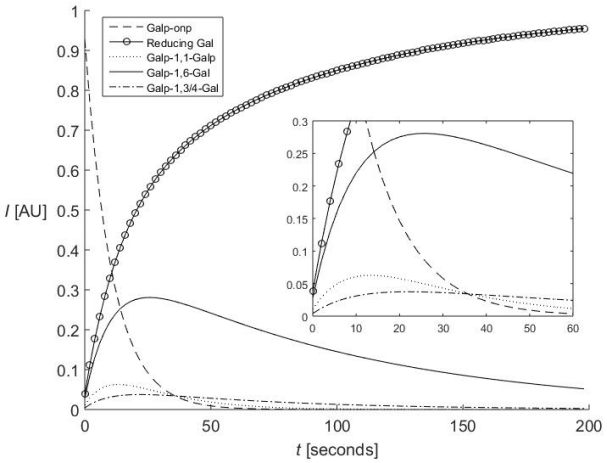


Figure 4. Plot of the modelled fits with the polarization decay removed. The insert is a zoom.

Compared to other methods of detection, dDNP NMR can achieve both high sensitivity and high temporal resolution. Even for though β-galactosidase is one of the most investigated carbohydrate converting enzymes, we discovered previously unknown intermediates of the reaction. It was possible to increase the knowledge of the enzyme, as it had not previously been reported to produce any non 1,6-linkages. Furthermore, the method could also be used to determine the relative transglycosylation rates producing each of the three observed intermediate products, and finally it was possible to approximate the relative rates of hydrolysis. To improve the transfer speed and initialization of experiment, a transfer system like the one described in Bowen and Hilty (2008) could be implemented.[19]

Achieving this level of detail, sensitivity and temporal resolution can help further the understanding of enzymatic reactions, as is shown herein with one of the best described and most investigated carbohydrate converting enzymes. In this case, low-populated and short lived transglycosylation intermediates were observed, and especially the Galp-1,1-Galp would be difficult to detect using other conditions, as it is a much better substrate than the other transglycosylation products. This data also suggests that the hydrolysis and transglycosylation mechanism is more complex than previously described, as it goes through multiple low-populated short lived intermediates.

Experimental Section

The DNP was carried out on samples containing Galp-onp (1.4 M) in DMSO with OX063 trityl radical (18 mM). The samples, 35-45 µL, were polarized for at least 120 min using a HyperSense (Oxford Instruments) operating at 3.35 T and <1.4 K with a 94 GHz microwave source. The dissolution was performed using 5.0 mL of 175° C aqueous pH 7.4 phosphate buffer (50 mM) and approximately 0.5 mL was manually injected into the NMR tube containing the enzyme, approximately 9 u per mM substrate, and the acquisition started. The final concentration of substrate was approximately 10 mM after mixing with the enzyme. Any addition of acceptor was done prior to mixing with enzyme to a final concentration of approximately 175 mM. The NMR acquisition was

carried out on a Varian Inova 400 MHz using 100 scans of 20° flip angles using 32768 points with a repetition time of 2 s. In the first scan, a signal to noise increase corresponding to 8-10% polarization was achieved. The NMR data was acquired with VnmrJ 4.2 (Agilent), processed using TopSpin 3.5 (Bruker) and data analysis was performed using Matlab R2015a (MathWorks). The fitting was performed by minimizing the squared difference between the integrals and the model using the Matlab functions fmincon and ode45. Synthesis of isotope labelled substrate is described in the supporting information.

Acknowledgements

C.K. acknowledges the Novo Nordisk foundation for funding: Biotechnology-based synthesis and production research, program grant no 5371. The Danish National Research Foundation is acknowledged for funding the Center for Hyperpolarization in Magnetic Resonance (DNRF124).

Keywords: Enzyme catalysis • Hyperpolarization • Kinetic modelling • NMR spectroscopy • Transglycosylation

[1] J. H. Ardenkjær-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 10158–10163.

[2] G. Robillard, E. Shaw, R. G. Shulman, *Proc Natl Acad Sci USA* **1974**, *71*, 2623–2626.

[3] S. G. Withers, D. Dombroski, L. A. Berven, D. G. Kilburn, R. C. Miller, R. A. J. Warren, N. R. Gilkes, *Biochem. Biophys. Res. Commun.* **1986**, *139*, 487–494.

[4] P. R. Jensen, S. Meier, J. H. Ardenkjær-Larsen, J. Ø. Duus, M. Karlsson, M. H. Lerche, *Chem. Commun. (Camb).* **2009**, 5168–5170.

[5] K. R. Keshari, D. M. Wilson, A. P. Chen, R. Bok, P. E. Z. Larson, S. Hu, M. Van Criekinge, J. M. Macdonald, D. B. Vigneron, J. Kurhanewicz, *J. Am. Chem. Soc.* **2009**, *131*, 17591–17596.

[6] S. Meier, M. Karlsson, P. R. Jensen, M. H. Lerche, J. Ø. Duus, *Mol. Biosyst.* **2011**, *7*, 2834–2836.

[7] S. Meier, P. R. Jensen, J. Ø. Duus, *ChemBioChem* **2012**, *13*, 308–310.

[8] T. B. Rodrigues, E. M. Serrao, B. W. C. Kennedy, D.-E. Hu, M. I. Kettunen, K. M. Brindle, *Nat. Med.* **2013**, *20*, 93–97.

[9] E. Miclet, D. Abergel, A. Bornet, J. Milani, S. Jannin, G. Bodenhausen, *J. Phys. Chem. Lett.* **2014**, *5*, 3290–3295.

[10] S. Meier, M. Karlsson, P. R. Jensen, *ACS Sustain. Chem. Eng.* **2017**, *5*, 5571–5577.

[11] D. H. Juers, B. W. Matthews, R. E. Huber, *Protein Sci.* **2012**, *21*, 1792–1807.

[12] R. Egel, *J. Theor. Biol.* **1979**, *79*, 117–119.

[13] R. E. Huber, G. Kurz, K. Wallenfels, *Biochemistry* **1976**, *15*, 1994–2001.

[14] C. W. Chen, C. C. Ou-Yang, C. W. Yeh, *Enzyme Microb. Technol.* **2003**, *33*, 497–507.

[15] N. F. Bras, P. A. Fernandes, M. J. Ramos, *J. Chem. Theory Comput.* **2010**, *6*, 421–433.

[16] B. Rodriguez-Colinas, M. A. De Abreu, L. Fernandez-Arrojo, R. De Beer, A. Poveda, J. Jimenez-Barbero, D. Haltrich, A. O. Ballesteros Olmo, M. Fernandez-Lobato, F. J. Plou, *J. Agric. Food Chem.* **2011**, *59*, 10477–10484.

[17] S. W. A. Hinz, R. Verhoef, H. A. Schols, J. P. Vincken, A. G. J. Voragen, *Carbohydr. Res.* **2005**, *340*, 2135–2143.

[18] S. Koto, S. Inada, S. Zen, *Chem. Lett.* **1980**, 403–406.

[19] S. Bowen, C. Hilty, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 5235–7.

WILEY-VCH

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

Text for Table of Contents

((Insert TOC Graphic here))

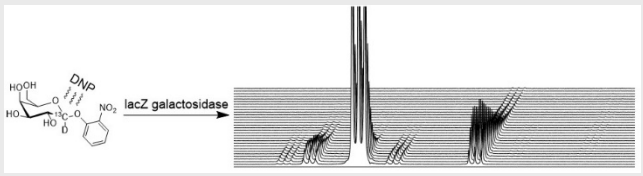
*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Layout 2:

COMMUNICATION



*Christian Kjeldsen, Jan H. Ardenkjær-Larsen, Jens Ø. Duus**

Page No. – Page No.
Discovery of novel intermediates of *lacZ* β-galactosidase catalyzed hydrolysis using dDNP NMR

Using dissolution dynamic nuclear polarization, the mechanisms of *lacZ* β-galactosidase was investigated, and novel intermediates were observed. These intermediates are made by transglycosylation, but as they are also substrates, they are quickly hydrolysed again. These results suggests that hydrolysis and transglycosylation might be more complex than previously described.

WILEY-VCH

Supporting Information

Discovery of novel intermediates of *lacZ* β-galactosidase catalyzed hydrolysis using dDNP NMR

Christian Kjeldsen, Jan H. Ardenkjær-Larsen and Jens Ø. Duus*

Abstract: Using dissolution dynamic nuclear polarization, the sensitivity of single scan solution state ^{13}C NMR can be improved up to four order of magnitude. In this study, the enzyme *lacZ* β-galactosidase from *E. coli.* was subjected to hyperpolarized substrate, and previously unknown reaction intermediates were observed, including 1,1-linked disaccharide. The enzyme is known for making 1,6-transglycosylation producing products like allolactose, that are also substrates. To analyze the kinetics, a simple kinetic model was developed and used to determine relative transglycosylation and hydrolysis rates of each of the intermediates, and the novel transglycosylation intermediates were determined as better substrates than the 1,6-linked one, explaining their transient nature. These findings suggest that hydrolysis and transglycosylation might be more complex than previously described.

DOI: 10.1002/chem.2017xxxxx

Table of Contents

Experimental Procedures

D-[1-^{13}C;1-^{2}H]galactose was purchased from Omicron Biochemicals, DMSO from Merck ox063 from GE Healthcare and all other chemicals was purchased from Sigma Aldrich. NMR spectra were acquired on a Bruker 400 MHz AvanceIIIHD with a 5 mm prodigy probe.

Synthesis of β-D-[1-^{13}C;1-^{2}H]galactopyranoside pentaacetate

Adapted from litterature[1], 492 mg (2.7 mmol) of D-[1-^{13}C;1-^{2}H]galactose was suspended in 10 mL dry pyridine and cooled to 0° C in an ice bath. Subsequently, 2.85 mL (25.8 mmol) of acetic anhydride was added slowly followed by 20 mg (0.16 mmol) 4-diaminomethylpyridine (DMAP). The reaction mixture was stirred at 0° C for 2 hours, allowed to reach room temperature and then stirred 16 hours at room temperature. The reaction mixture was concentrated on rotary evaporator, redissolved in 2 mL of ethyl acetate, washed with water (3x3 mL), and saturated aqueous potassium hydrogen sulfide (3x3 mL). The organic layer was concentrated on rotary evaporator, and the resulting product was used without any further purification.

Synthesis of α-bromo-D-[1-^{13}C;1-^{2}H]galactopyranoside tetraacetate

Adapted from literature[2], the crude β-D-[1-^{13}C;1-^{2}H]galactopyranoside pentaacetate was dissolved in 4.8 mL dry CH_2Cl_2 and cooled to 0° C. Subsequently, 2 mL (34 mmol) 33% HBr in acetic acid was added slowly, and the mixture was stirred for 30 minutes at 0° C, after which it was allowed to reach room temperature and stirred for an additional 5 hours. Following this, 5 mL of CH_2Cl_2 was added and the mixture was washed with saturated aqueous $NaHCO_3$ (2x15 mL) and water (15 mL). The organic layer was concentrated on rotary evaporator and was used without any further purification.

Synthesis of *o*-nitrophenyl β-D-[1-^{13}C;1-^{2}H]galactopyranoside tetraacetate

Adapted from litterature[3], the acetate protected α-bromo-galactopyranoside was dissolved in 15 mL dry MeCN and transferred to a separate mixture containing 0.95g (6.83 mmol) of *o*-nitrophenol, 1.80g (6.53 mmol) of silver carbonate, 1.0 mL (8.40 mmol) of 2,6-dimethylpyridine and ~5g activated 4Å molecular sieves in 40 mL dry MeCN. Subsequently, the reaction mixture was stirred for one hour in the dark, filtered through a plug of Celite, which was subsequently rinsed with 10 mL of CH_2Cl_2, and the solvents were removed on rotary evaporator. The concentrate was redissolved in 50 mL of CH_2Cl_2 and washed with water (30 mL), 5% NaOH (3x30 mL), 1M HCl (30 mL) and water (2x30 mL), following which the organic layer was concentrated on rotary evaporator and used without any further purification.

Synthesis of *o*-nitrophenyl β-D-[1-^{13}C;1-^{2}H]galactopyranoside

Adapted from litterature[3], the acetylated *o*-nitrophenyl β-D-galactopyranoside was dissolved in 40 mL of dry MeOH and 15 mL of dry CH_2Cl_2. Subsequently, 40 mg (7.38 mmol) was added and the reaction mixture was stirred for 5 hours at room temperature. The solvent was removed on a rotary evaporator and the concentrate was redissolved in 40 mL MeOH and acidic cation-exchange resin was added and the mixture was stirred for an hour at room temperature. Following this, the resin was removed using Buchner funnel and the product was recrystallized from the filtrate in several batches.

Results and Discussion

1-O-methylated galactopyranoside as acceptor

In figure S1 the reaction of the hyperpolarized doubly labelled Galp-onp substrate with the enzyme β-galactosidase with added 1-*O*-methyl galactopyranoside as acceptor. Compared to figure 1 and figure 2, the peak at 99.7, corresponding to the 1,1-linked product, has disappeared from the spectra.

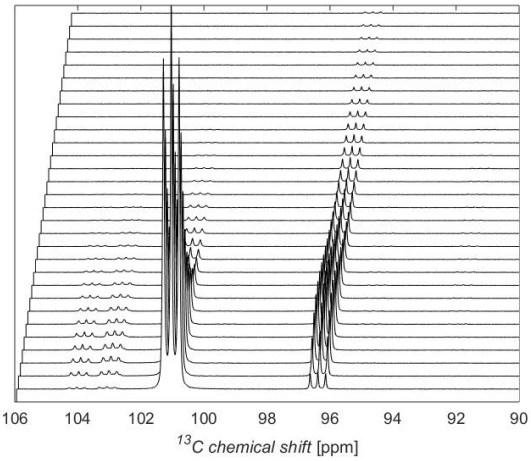


Figure S1. The first 30 spectra of doubly labelled Galp-onp enzymatic hydrolysis with 1-O-Me Galp as acceptor. The repetition time is 2 s and each spectrum is shifted 0.06 ppm upfield.

References

[1] A. Thomas, A. Shukla, S. Sivakumar, S. Verma, *Chem. Commun.* **2014**, *50*, 15752–15755.

[2] V. Percec, P. Leowanawat, H. J. Sun, O. Kulikov, C. D. Nusbaum, T. M. Tran, A. Bertin, D. A. Wilson, M. Peterca, S. Zhang, et al., *J. Am. Chem. Soc.* **2013**, *135*, 9055–9077.

[3] D. E. Green, C. L. Ferreira, R. V. Stick, B. O. Patrick, M. J. Adam, C. Orvig, *Bionconjugate Chem.* **2005**, *16*, 1597–1609.

Paper 5

GH20 β-*N*-acetyl hexosaminidase from psychrotrophic marine bacterium *Paraglaciecola hydrolytica*

Visnapuu, T.; Teze, D.; **Kjeldsen, C.**; Duus, J. Ø.; Lie, A.; André-Miral, C.; Pedersen, L. H.; Stougaard, T.; Svensson, B.

Manuscript in preparation

GH20 β-*N*-acetyl hexosaminidase from psychrotrophic marine bacterium *Paraglaciecola hydrolytica*

Triinu Visnapuu,[a]*# David Teze,[a,b] Christian Kjeldsen,[c] Jens Øllgaard Duus,[c] Aleksander Lie,[d] Corinne André-Miral,[b] Lars Haastrup Pedersen,[d] Peter Stougaard,[e] Birte Svensson[a]#

Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark[a]; Unité de Fonctionnalité et Ingénierie des Protéines (UFIP), Université de Nantes, Nantes, France[b]; Department of Chemistry, Technical University of Denmark, Kongens Lyngby, Denmark[c]; Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark[d]; Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg C, Denmark[e]

Running head: *N*-acetyl hexosaminidase of *Paraglaciecola hydrolytica*

#Corresponding authors:

Dr. T. Visnapuu; *Present address: Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia; e-mail: triinu.visnapuu@ut.ee

Prof. B. Svensson; Department of Biotechnology and Biomedicine, Building 224, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark; e-mail: bis@bio.dtu.dk

ABSTRACT

Genomes of terrestrial bacteria have been investigated intensively to find a new sources of bioactive compounds or novel enzymes which could have some applications in biotechnology. Marine bacteria in contrast received far less attention even though a high bioactive potential of the bacterial genomes have been predicted. This especially applies to the polysaccharide degradation capacity of water-originated cold-tolerant bacteria from which many glycoside hydrolases with new or altered activities could be isolated.

β-*N*-acetyl hexosaminidases (EC 3.2.1.52) are glycoside hydrolases (GHs) found in GH families 3, 20, 84 and 116 acting on *N*-acetylglucosides and *N*-acetylgalactosides with release of non-reducing end *N*-acetyl-D-hexosamine residues. Whereas such enzymes have been associated with chitin metabolism and changes of *N*-acetylhexosamine-containing compounds, e.g. glycosphingolipids and human milk oligosaccharides. In the present study the potential β-*N*-acetyl hexosaminidases of marine bacterium *Paraglaciecola* hydrolytica were identified by genome analysis, which according to the corresponding protein sequence were very different from previously characterized β-*N*-acetyl hexosaminidases. Recombinant PhNah20A, which was produced in *Escherichia coli* in yields of around 2 mg mL^{-1}, hydrolyzed Glc*N*Ac- and Gal*N*Ac-containing sugars with similar activity. PhNah20A was of low stability by fully active after 4 d at 37°C in the presence of 0.5% bovine serum albumin or 0.5% Triton X-100. PhNah20A moreover is able to catalyse transglycosylation activity producing and β-Galp-1,4-β-Glcp-1,1-β-GlcpNAc and β-GlcpNAc-1,3-Galp- β-1,4-Glcp when lactose was used as an acceptor.

IMPORTANCE

Glycoside hydrolases from terrestrial and marine environments in general are very different. Indeed, the biotopes are physically distinct and the main polysaccharide structures offer a stark contrast from the terrestrial ubiquitous crystalline cellulose and hemicelluloses to chitosan, alginate, agar, carrageenans, etc. found in marine habitats. Despite considerable efforts to unveil their properties, marine glycoside hydrolases remain understudied. Here we aimed at characterizing the hexosaminidases of a newly discovered and sequenced bacterial strain, *Paraglaciecola hydrolytica*, which has been found to be proficient for marine polysaccharide hydrolysis.

INTRODUCTION

Approximatively half of the carbohydrates on earth are produced by marine algae (1), and thus marine bacteria have developed myriads of carbohydrate degrading enzymes. As the marine and terrestrial polysaccharides differ greatly, it is expected that marine bacterial carbohydrate active enzymes (CAZymes) have distinct characteristics compared to more well-known counterparts, which are overwhelmingly from land-dwelling organisms (1).

Polysaccharide degradation is a complex process and participation of several hydrolyzing enzymes and their regulators might be mandatory. Chitin, a biopolymer with β-1,4 linked *N*-acetyl-D-hexosamine (GlcNAc) residues, is one of the most abundant carbohydrates in nature. The chitinolytic bacterial cells in marine environments is estimated to amount to 5% (2). Chitin-degrading genes and high chitinolytic potential are seen for species from marine Gammaproteobacteria of the genera *Vibrio* and *Pseudoalteromonas* which commonly live in tight association with eukaryotic organisms like crustaceans and seaweed (3).

One group of enzymes involved in chitin degradation are β-*N*-acetyl hexosaminidases (β-NAHAs) (EC 3.2.1.52) found in GH families 3, 20, 84 and 116 (2)(www.cazy.org). However, despite a large number of identified sequences in genomes and metagenomes, only a limited number of β-NAHAs of marine origin have been characterized. Chitinases Chi-I and II isolated from the temperature and halotolerant bacterium *Salinivibrio costicola* (4), NagA and CbsA from the hyperthermophiles *Thermotoga maritima* and *T. neapolitiana* (5), Nag20A from the widespread *Aeromonas hydrophila* (6) and ExoI protein from *Vibrio* species (7) have been investigated in detail after cloning, overexpression and purification.

Recently, a new species of agar-degrading marine bacterium was isolated from eelgrass (*Zostera sp.*) in Denmark and identified as a *Paraglaciecola hydrolytica* (previously named *Paraglaciecola* sp. strain S66). Analysis of the whole-genome sequence of *P. hydrolytica* showed a vast amount of genes encoding algal polysaccharide-degrading enzymes, i.e. agarases, alginases, pectinases, xylanases, and carrageenanases. All together 280 CAZymes were predicted, including 113 GHs (8). Notably, *P. hydrolytica* was able to use various carbohydrates, e.g. agar, agarose, porphyran, *κ*-carrageenan, alginate and laminarin, as carbon sources (9). Therefore, this bacterium could be considered as a potent source for new saccharide-acting enzymes with a high biotechnological potential to utilize marine biomass. As the growth temperature for *P. hydrolytica* is 10–25°C (9), probably cold-acting enzymes could be discovered.

Hexosaminidases due to the presence of equatorial C2 *N*-acetyl group in substrate residues act through an anchimeric substrate assisted mechanism through an oxazolinium-ion intermediate (10), rather than the typical double-displacement mechanism including a glycosyl-enzyme intermediate that is applied by most retaining glycoside hydrolases. Recently GH3 hexosaminidases were proven to be glycoside phosphorylases rather than glycoside hydrolases (11).

Three putative genes were encoding *N*-acetyl hexosaminidases from *P. hydrolytica,* belonging to both GH3 and GH20 families were identified and cloned and one *Ph*Nah20A was successfully produced and characterized with regard to stability and enzymatic activity including transglycosylation products using lactose as acceptor and identified by NMR spectroscopy.

RESULTS

Identification of putative *N*-acetyl hexosaminidases from *P. hydrolytica*

The draft genome sequence from the newly isolated Gammaproteobacterium *P. hydrolytica* (8) was automatically annotated using the Rapid Annotation using Subsystem Technology (RAST) (12) revealing three putative *N*-acetyl hexosaminidases (EC 3.2.1.52), one from family GH3 – now considered hexosamine phosphorylases (11) – and two belonging to the GH20 family. All three putative NAHA genes are located in the contig 11 of the *P. hydrolytica* whole genome

shotgun sequence (NCBI ref. NZ_LSNE01000003.1).

The hypothetical protein from *Paraglaciecola* sp. S66 (NCBI accession: WP_068373836.1; 807 residues, calculated Mw 91938 Da; see Fig.1B) belongs to the GH20 hexosaminidase superfamily. The closest hits (protein BLAST) are GH20 β-*N*-hexosaminidases from phylogenetically close water and soil bacteria. None of those genes from *Paraglaciecola* or related bacterial species have been previously heterologously expressed for characterization of the corresponding proteins. The highest protein sequencex identity was in the 45–52% range for NANAs from *Lacimicrobium alkaliphilum* (WP_099034787.1; identity 52%), *Buttiauxella ferragutiae* (WP_064543196.1; identity 47%), *Shewanella woodyi* (WP_012326528.1; 45%), hypothetical protein from *Pseudoalteromonas rubra* (WP_046003581.1; identity 45%) and NAHA from *Serratia plymuthica* (WP_062869400.1; identity 45%). Based on the sequence analysis the above-mentioned proteins contain a GH20 catalytic domain with an $(\alpha/\beta)_8$ barrel fold and N-terminal domain with possible zincin-like fold (PF02838; http://pfam.xfam.org/family/PF02838). No signal peptide sequence was predicted for WP_068373836.1 by SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/)

The hypothetical protein (WP_082768773.1; 879 residues, calculated Mw 98142 Da; see Fig1C) is predicted to consist of an N-terminal 28 residues signal peptide and three domains; a putative carbohydrate binding domain of CHB_HEX superfamily (PF03173; http://pfam.xfam.org/family/PF03173); domain 2 of GH20 (PF02838; http://pfam.xfam.org/family/PF02838); and in the C-terminal part a catalytic domain of GH20 (PF00728; http://pfam.xfam.org/family/PF00728) . WP_082768773.1 resembles chitobiase of *Aliiglaciecola lipolytica* (WP_008846397.1; identity 49%), NAHA of *Shewanella halifaxensis* (WP_012276260.1; identity 47%), NAHA *Lacimicrobium alkaliphilum* (WP_099036016.1; identity 46%), and hexosaminidase of *Paraglaciecola arctica* (WP_007619251.1; identity 45%).

Glycoside hydrolase (WP_068373761.1; 628 aa, calculated Mw 68366 Da; Fig. 1A) is predicted homologous to GH3 of *Paraglaciecola arctica* (WP_039989320.1; identity 69%), NAHA of *Paraglaciecola psychrophila* (WP_007639257.1; identity 68%), GH3 from *Pseudoalteromonas atlantica* (WP_011576853.1; identity 57%) and NAHA A precursor from *Paraglaciecola agarilytica* (WP_008305385.1; identity 57%). The conserved domain analysis revealed the presence of a periplasmic β-glucosidase BglX domain (PF00933) belonging to GH3 family.

Genomic regions around *N*-acetyl hexosaminidase genes of *P. hydrolytica*

Utilization of chitobiose in *Escherichia coli* has been thoroughly studied and is shown to be mediated by the *chb*BCARFG genes of the *chb* operon. *chb*B, *chb*C and *chb*A thus encode the permease that transports chitobiose across the inner membrane and phosphorylates the substrate at the C-6 position of the nonreducing end residue. Chitobiose-6-P is further metabolized intracellularly. The *chb*F gene encodes a phospho-β-glucosidase and *chb*R an AraC-type repressor-activator. The operon is regulated by three transcription factors: NagC, a key regulator of the nag genes involved in amino sugar metabolism, ChbR operon-specific regulator and CRP-cyclic AMP (CRP-cAMP) (see (13) and references therein).
The chitin or chitobiose metabolism of *P. hydrolytica* or phylogenetically close organisms (from the order *Alteromonadales*) have not been studied nor have the responsible proteins been identified.

According to the analysis of flanking regions in the genome sequence of *P. hydrolytica* around the *nah* genes, settings typical to a *chb* operon were not recognized. Some putative genes encoding proteins with potential participation in degradation of acetylated compounds, transporting function and transcription regulation were disclosed

(Fig. 1). MFS (membrane transport proteins for small molecules) transporters WP_068373856.1 and WP_068374723.1, were identified as well as several types of transcriptional regulators: LacI family (WP_068374732.1), TetR/AcrR family (WP_068374736.1), MurR/RpiR family (WP_068373766.1). Also TonB-dependent receptor (WP_068374714.1), which type of receptors is shown to be involved in carbohydrate scavenging by aquatic bacteria (14), was identified downstream of WP_082768773.1 (Fig. 1C). Notably, many sequences resulting in hypothetical or proteins/domains with unknown function were disclosed in those regions.

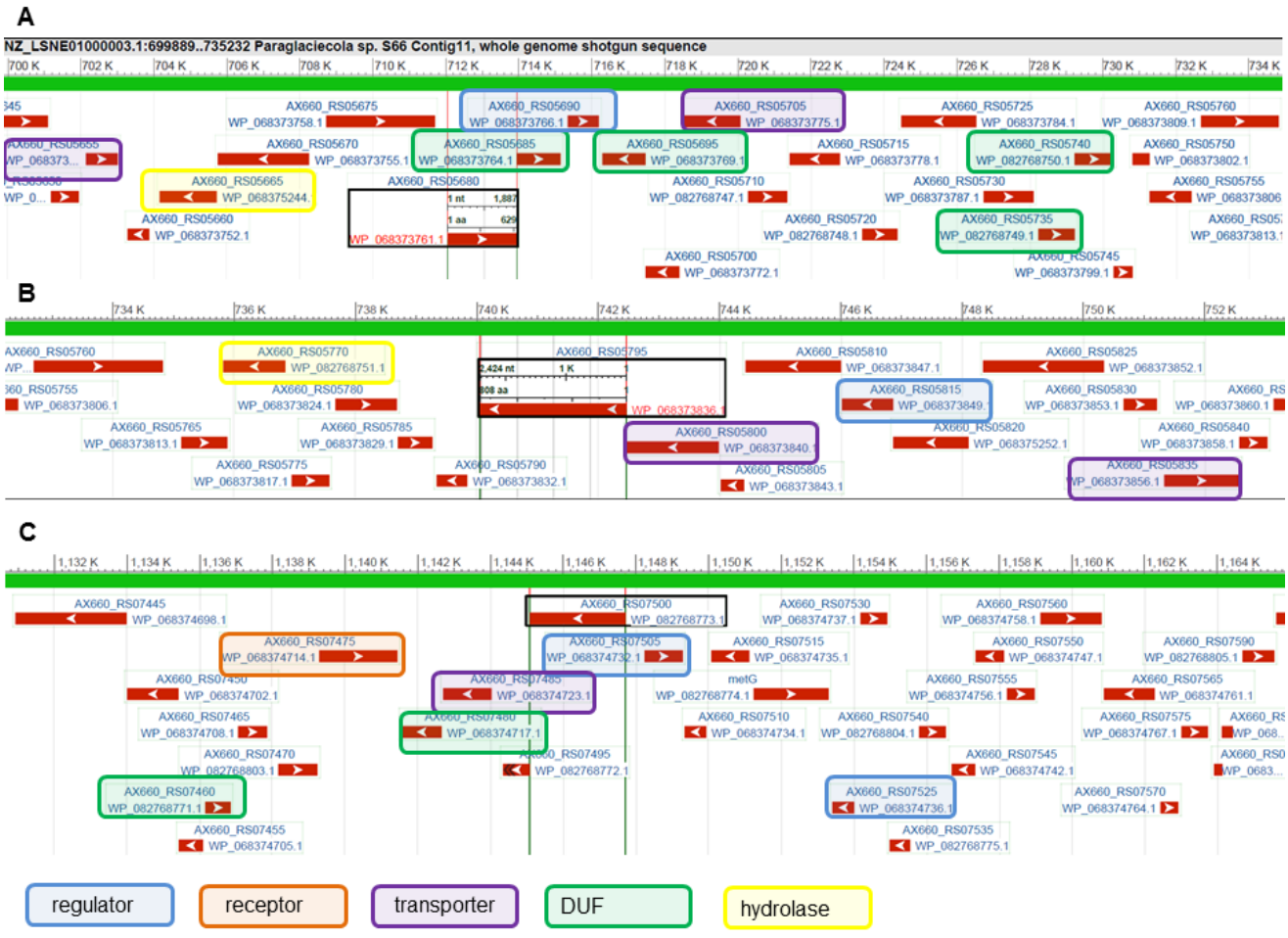


FIG 1 Genomic regions (A, B, C) and proteins flanking the identified NAHAs (black frame). Predicted functions of the proteins were classified and marked with frames of different color. The information was retrieved from NCBI (Ref: NZ_LSNE01000003.1) and Pfam databases. DUF – domain of unknown function.

Cloning and expression studies of NAHAs

The three candidate genes were cloned from genomic DNA and inserted into the pURI3TEV expression vector. WP_082768773.1 was cloned without the N-terminal signal peptide sequence. DNA sequencing confirmed the cloned sequences were identical to those of the *P. hydrolytica* genome.

Despite several attempts to express the genes only one (WP_068373836.1) resulted in a protein product in *E. coli*. The GH20 NAHA was termed *Ph*Nah20A. No other respective candidate proteins were found in the insoluble fraction or in whole cells from

IPTG-induced culture analyzed on SDS-PAGE. As the yield of *Ph*Nah20A was modest (up to 6 U per mg of protein), different stains and induction strategies were tried. Using an auto-induction medium ZYM-5052 (16) did not give higher expression rates (Fig. 2) as opposed Nyffenegger *et al*. (15)

E. coli strain BL21(DE3) was used for the production of *Ph*Nah20A yielding at most 2 mg from 1 L of induced culture due to difficulty in getting sufficient purity and several NAHA activity containing fractions were discarded during purification steps.

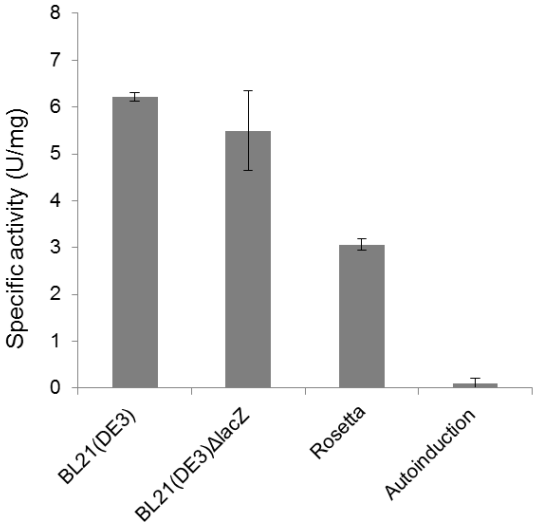


FIG 2 *Ph*Nah20A production in different *E. coli* strains grown in LB and induced by IPTG or in the auto-induction medium ZYM-5052 (30 h incubation) in BL21(DE3) background.

Characterisation of *Ph*Nah20A

Enzyme stability

Enzymes from *P. hydrolytica* could be moderately cold-acting or not stable at higher temperatures as the bacterium has optimum growth at 20–25°C (9). Indeed, *Ph*Nah20A was poorly stable in buffer (Fig. 3), protein

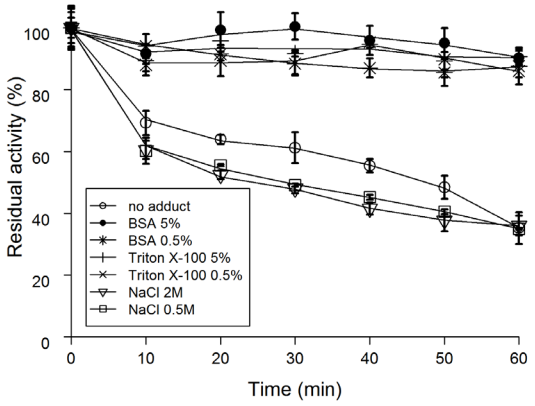


FIG 3 *P*hNah20A stability and effect of stabilizers detected using diluted protein kept on ice for certain time and the specific activity detected at 37°Cat pH 6.

(0.5% BSA) or detergent (0.5% Triton X-100) considerably increased the stability at pH 6.0 (Fig. 3). NaCl, however, was slightly detrimental to the stability. This behavior together with the absence of a signal peptide suggests that the enzyme is cytoplasmic, where a high protein concentration can ensure its stability.

Without stabilizing agents, the activity was completely lost at 50°C within 5 min. By contrast, in the presence of 0.5% BSA 50% and 3% residual activity was present after 20 min and 4 h incubation, respectively, (Fig. 4). Importantly, when incubated with BSA at 37°C the enzyme activity was intact after 4 d.

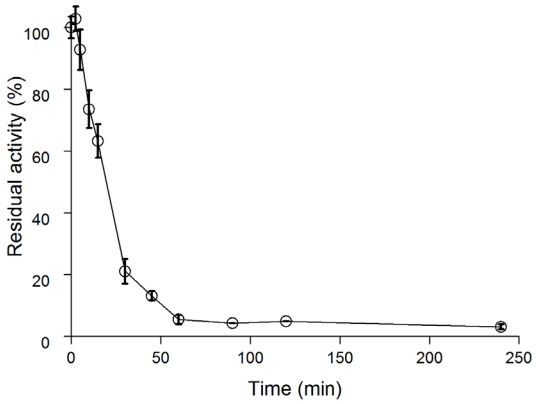


FIG 4 *Ph*Nan20A thermal inactivation at 50°C, pH 6 in the presence of 0.5% BSA.

pH and temperature optima therefore were determined in 0.5% BSA (Fig. 5). Noticeably after a rapid initial inactivation, the enzyme retains a few percent of the starting activity after several hours at 50°C.

pH and temperature optima

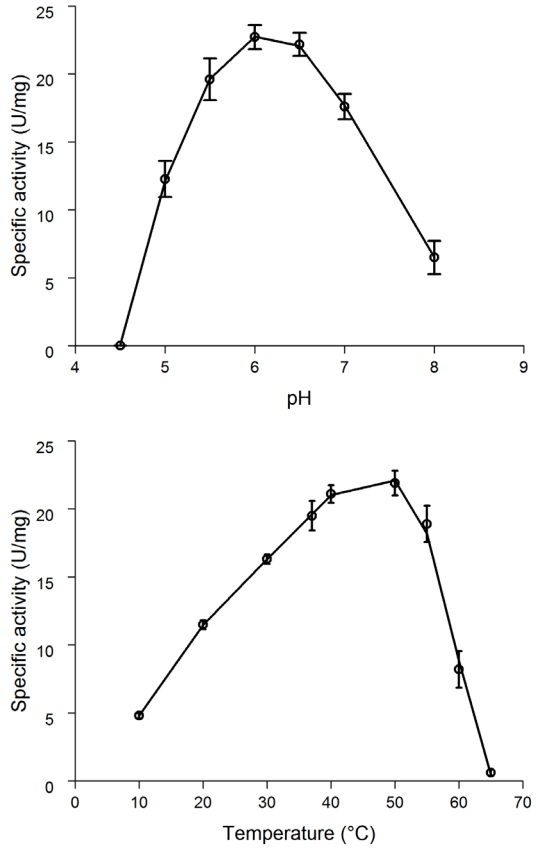


FIG 5 *Ph*Nah20A pH and temperature optima. The effect of pH (top panel) and temperature (bottom panel) on initial rates of hydrolysis of 2 mM *p*NPGlc*N*Ac.

It is worth noting that *Ph*Nah20A was found to be most active in the pH range 5.0-7.5 centered at pH 6.0, while *P. hydrolytica* shows optimal growth at pH 7.5 (9), where the enzyme activity is about half the maximum value. This supports an intracellular location of the enzyme (most probably in the periplasma) in the native host whereas the conditions are more favourable for this enzyme activity there.

Kinetic parameters of *Ph*Nah20A.

*Ph*Nah20A catalyses hydrolysis of pNPGlcNAc and pNPGalNAc and displayed a twofold lower K_M value towards the former substrate (Table 1), whereas k_{cat} was essentially the same leading to half the catalytic efficiency constant for *p*NPGalNAc.

Substrate	*p*NP-GlcNAc	*p*NP-GalNAc
K_M (mM)	0.12 ± 0.02	0.23 ± 0.03
V_{max} (U·mg^{-1})	23.6 ± 1	24.8 ± 1.5
k_{cat} (s^{-1})	36.9 ± 1.5	38.9 ± 2.4
k_{cat}/K_M (mM^{-1} ·s^{-1})	308 ± 13	169 ± 10

Table 1 Kinetic parameters of Ph-NAHA_1.

*Ph*Nah20A is therefore identified as a hexosaminidase rather than gluco- or galactosaminidase.

Transglycosylation study

In addition to p-nitrophenyl hexosaminides *Ph*Nah20A catalysed hydrolysis of chitooligosaccharides and lacto-*N*-triose II (LNT2), a core structure of human milk oligosaccharides. Recently the GH20 BbhI enzyme from *Bifidobacterium bifidum*, was found to efficiently catalysis hydrolysis of LNT2 (17) as well as its synthesis by transglycosylation in 44.9% yield from 20 mM *p*NP-GlcNAc and 400 mM lactose (18). This motivated, investigation of the ability of *Ph*Nah20A to form LNT2 by transglycosylation using lactose as an acceptor and either *p*NP-GlcNAc or NAG-oxazoline as activated donor. This resulted in transglycosylation products which were identified by NMR (see Supplemental information Fig. S1-S4 and related discussion in SI) in a mixture containing, LNT2, a non-reducing trisaccharide, being predominant, and a third product. This formation of several transglysylation products makes *Ph*Nah20A less suitable for direct synthesis of LNT2 from lactose.

Since *Ph*Nah20A is not expected to act on LNT2, or other trisaccharides involving lactose and a β-GlcNAc residue in its natural habitat, it seems to be promiscuous with regard to the types of

bond it can act upon, which is not uncommon among GH20 NAHAs.

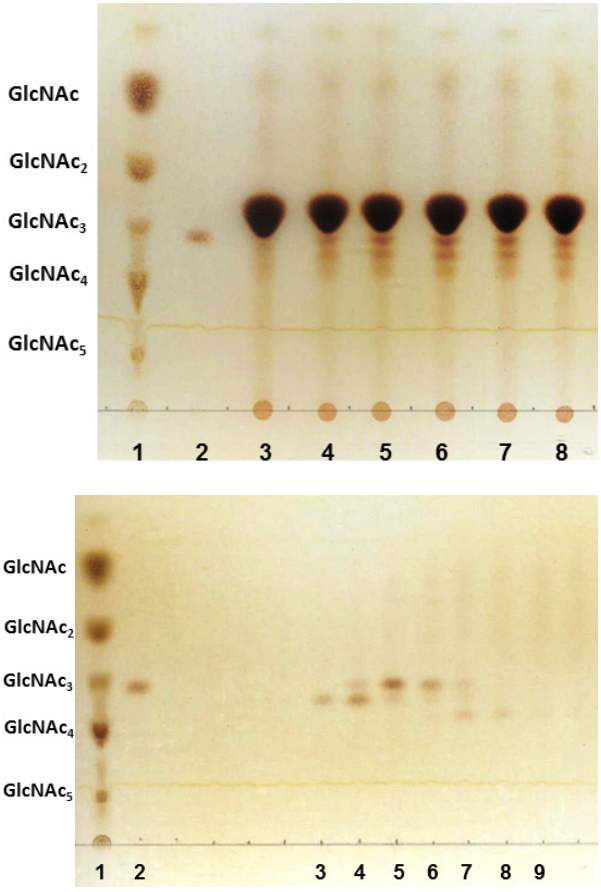


FIG 6. (A) Time course of transglycosylation by *Ph*Nah20A using NAG-oxazoline as donor and lactose as acceptor (see Materials and methods for details). COS (lane 1) and LNT2 (lane 2) were used as reference sugars. 3 – Control sample without *Ph*Nah20A; 4 – 0.03 h; 5 – 0.5 h; 6 – 1 h; 7 – 2 h; 8 – 24 h reaction time. (B) TLC analysis of the trisaccharide-containing fractions of the *Ph*Nah20A transglycosylation reaction mixture separated by gel-permeation chromatography. The enzymatic synthesis, separation and purification steps were conducted as described above. COS (lane 1) and LNT2 (lane 2) were used as reference sugars. 3 - fraction 49; 4 - fraction 50; 5 - fraction 51; 6 - fraction 52; 7 - fraction 53; 8 - fraction 54; 9 - fraction 55. Samples in bold were analyzed by NMR.

As *N*-acetyl hexosaminidases especially of fungal but also of bacterial origin have been shown to exhibit transglycosylating activity (Slamova et al., 2010 and additional references; (15, 18), it was decided to test this ability and characterise eventual products of *Ph*Nah20A. A few reports have described the formation of lacto-N-triose II (LNT2) when lactose was used as acceptor (Matsuo et al., 2003; (15, 18). In some studies formation of LNT2 isomers occurred but the structure of these was not identified (15). LNT2 is an initial building block in human milk oligosaccharides (HMO) (Kobata, 2010) and it therefore attracted biotechnological interest. In LNT2 β-GlcNAc*p* is substituting the 3-position of galactose in the non-reducing end of lactose.

The reaction of *Ph*Nah20A with NAG-oxazoline as a donor and lactose as acceptor was monitored first by using TLC (Fig. 6A). Based on the results, *Ph*Nah20A synthesizes at least two different trisaccharides by the transglycosylation. One of these migrating similar to LNT2 in TLC was confirmed by NMR spectroscopy (see Supplementary material) after purification (HPAEC-PAD).

GPC purification of reaction mixture resulted fractions containing transglycosylation products separated from the acceptor (Fig. 6B). Some of the fractions were analysed by NMR. The first of the three samples analysed by NMR, fraction 50, contained a mixture of LNT2 and a non-reducing trisaccharide in an approximately 3:7 ratio. The non-reducing trisaccharide was determined to β-Gal*p*-1,4-β-Glc*p*-1,1-β-GlcNAc*p*, reported before as a transglycosylation product of a β-N-acetylhexosaminidase of fungal origin, but only assigned by NMR as a peracetylate (19). Sample 51 consists of the above-mentioned two sugars, but in a 4:1 ratio. Fraction 53 was found to also contain LNT2 as well as a different trisaccharide, which due to the low abundance was not structure-determined.

However, based on chemical shifts it seemed unlikely that the galactose acted as acceptor, but that either the 2- or 3-position of the glucose had. HSCQ spectra of the fractions and NMR assignments of the two main products are shown in Supplementary information.

Materials and methods

Materials

Lacto-N-triose II was purchased from Elicityl Oligotech (France). Lactose *p*NPGal*N*Ac were from Carbosynth (USA), *N*,*N*'-diacetylchitobiose (GlcNAc$_2$) from Omicron (USA), *N*,*N*',*N*''-triacetylchitobiose (Glc*N*Ac$_3$) and *p*NPGlc*N*Ac from Megazyme (Ireland) and *N*-acetylated chitooligosaccharides (COS) from Koyo Chemicals (Japan). Other chemicals were all purchased from Sigma-Aldrich (Germany) and used without further purification.

Bacterial strains and media

Paraglaciecola hydrolytica (type strain S66^T) (8, 9) was grown on Difco Marine Broth 2216 (BD, USA) or Marine Broth supplemented with 15 g·L^{-1} agar at room temperature (23°C).

E. coli DH5α was used for cloning and *E. coli* BL21 (DE3) and Rosetta (Novagen, Merch, USA) for gene expression studies and BL21 (DE3) for protein production. *E. coli* was grown in Luria-Bertani (LB; MoBio, USA) broth or on LB agar plates. Media was supplemented with 100 mg·L^{-1} ampicillin for selection. Auto-induction medium ZYM-5052 was prepared as shown previously (16). Liquid cultures were aerated on a shaker (160 rpm).

Molecular cloning and plasmids

The genomic DNA of *P. hydrolytica* was purified using Gentra Puregene Yeast/Bact kit B (Qiagen, USA) and plasmid DNA was isolated using GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific, USA). The DNA content was evaluated on NanoDrop Lite (Themo Fisher Scientific, USA).

Three putative NAHA-encoding genes of *P. hydrolytica* were isolated from the genomic DNA using primers pURITEVFwGl1002 and pURITEVRevGl1002 for the GH3 enzyme, pURITEVFwGl1026 and pURITEVRevGl1026, pURITEVFwGl1378 and pURITEVRevGl1378 for the GH20 enzymes, respectively (Supplementary Table 1). The genes were cloned into pURI3TEV vector by PCR-based cloning (20). The expression plasmids contained wild-type genes under the control of the T7 polymerase Ø promoter and were termed pURITEV-Gl1002, pURITEV-Gl1026 and pURITEV-Gl1378.

Protein expression and purification

E. coli BL21 (DE3) expressing the PhNah20A gene cloned in the pURI3TEV vector was grown in 1 L of LB medium at 37°C until reaching OD_{600} ≈ 0.5. The culture was induced by addition of 0.5 mM isopropyl thio-β-D-galactoside (IPTG) and incubated 20 h at 22°C. For initial expression study,10 µL of 20 mL culture from different time points (0 h, 4 h, 20 h) was mixed with 4 µL of SDS-PAGE loading dye, heated 10 min at 80°C to lyse the cells and denaturate the proteins. Then the mixtures were centrifuged 1 min 12000 *g* at room temperature, separated by pre-cast SDS-PAGs and respective buffer according to the manufacturers' instructions (NuPAGE, Thermo Fischer Scientific, USA) in XCell SureLock mini-cell electrophoresis system (Thermo Fischer Scientific, USA) and gels were stained using Coomassie Brilliant Blue G-250 dye. Small-scale cell lysates were prepared from the cell pellet of IPTG-induced cells. The pellet was suspended in 0.4 mL 50 mM Na-phosphate buffer (ph 7) and cells were lysed using 0.4 mL BugBuster protein extraction reagent

(Merck, USA) and approx. 100 U Benzoase nuclease (Merck, USA) was added. Lysates were centrifuged 20 min at 4°C 12 000 *g*.

For purification, cell pellet was collected by centrifugation (10 000 *g*, 15 min, 4°C) and were resuspended in 20 mL of lysis buffer (0.1 M phosphate, pH 8, 0.5 M NaCl, 10 mM imidazole containing 1 µL of Benzonase Nuclease). Cells were disrupted using Cell Pressure Homogenizer (Stansted, UK). To remove the cell debris, centrifugation (25 000 *g*, 20 min, 4°C) was applied and the supernatant was filtered through 0.45 µm membrane filter. Recombinant *Ph*Nah20A was purified by Ni^{2+}-affinity chromatography (HisTrapHP column, GE Healthcare, Sweden) followed by size-exclusion chromatography on HiLoad 16/60 Superdex 200 pg column (GE Healthcare) on an ÄKTA Avant chromatography system (GE Healthcare). Size-exclusion chromatography was conducted in 50 mM sodium phosphate buffer (pH 7.0) containing 0.3 M NaCl at a flow rate of 2 mL·min^{-1}. The eluted protein was analyzed by SDS-PAGE. Enzyme was concentrated using by Amicon ultra-15 30K centrifugal filter device (Merck, USA). PhNah20A was stored in above-mentioned buffer supplemented in 0.05% BSA and 0.02% Na-azide.

Enzyme assays

Routinely, *Ph*Nah20A specific activity was determined using 2 mM *p*NPGlcNAc as a substrate in two times diluted McIlIvaine's buffer with pH 6 (0.063 M Na_2HPO_4; 0.018 M citric acid) containing 0.05% BSA at 37°C. Reaction was started by adding enzyme with suitable dilution in McIlIvaine's buffer (pH 6). Dilutions were prepared immediately before use and kept on ice for short periods. Reaction was stopped at suitable time point by 250 µL 1 M Na_2CO_3 (final conc. 0.3 M). OD_{400nm} were recorded by UVA/UVB spectrophotometer. The method was calibrated with *p*NP. One U of specific activity was defined as 1 µmol of *p*NP released by mg of protein per min.

Protein concentration was determined using Pierce Coomassie (Bradford) Protein Assay Kit (Thermo Fischer Scientific, USA) in cell lysates and using NanoDrop Lite (Themo Fisher Scientific, USA) for purified protein. Extinction coefficient of *Ph*Nah20A was calculated in ExPasy server (https://web.expasy.org/protparam/).

To determine pH optimum, McIlIvaine's buffers with different pH values (pH 4.0–8.0 were prepared. The specific activities were determined as shown above at 37°C using 2 mM *p*NPGlcNAc as a substrate.

Temperature optimum was determined by measuring initial velocity of *p*NP release (specific activity) at different temperatures 10–65°C. Suitable enzyme dilutions were used and final protein content in the reaction mixtures was 0.3 per 1 mL of the reaction mixture.

Kinetic studies

Initial velocities of hydrolysis of 0.05−10 mM pNPGlcNAc and 0.1–2 mM pNPGalNAc were determined in 500 µL 0.5x McIlIvaine's buffer (pH 6) with 0.05% BSA at 37°C. Protein concentration was 0.3–1.2 µg per 1 mL of the reaction mixture. Reaction was stopped by addition of 250 µL 1 M Na_2CO_3 at suitable time points. OD_{400nm} of all kinetic samples were measured with a UVA/UVB spectrophotometer. Initial velocities were calculated from the slope of the zero-order plot of product concentration (*p*NP) against reaction time, plotted against substrate concentration and the resulting curves were fitted to the Michaelis-Menten hyperbolic equation $\mathrm{v} = \frac{\mathrm{V_{max}}*[S]}{K_\mathrm{M}+[\mathrm{S}]}$ using OriginPro 2015 (OriginLab) to obtain k_{cat} and K_M. $\frac{k_{cat}}{K_M}$ was determined at

low substrate concentration, where reaction rates can be approximate to the equation:

$$v = \frac{k_{cat}}{K_M * [E] * [S]}.$$

Transglycosylation assay

Reaction mixtures for identification of the yields and structures of transglycosylation products contained 100 mM NAG-oxazoline (for details, see below) (from 1 M stock in 50 mM Na-borate buffer, pH 9.3), 200 mM lactose, 10 U/mL (116 µg mL^{-1}) *Ph*Nah20A, 0.5% BSA in 50 mM sodium phosphate pH 8, at 37°C. Reactions were stopped at suitable time points by heating (5 min, 90°C), cooled at room temperature and centrifuged (12 000 *g*, 1 min, 4°C). Samples were diluted 4 times in milliQ water for TLC and 150 times for HPAEC-PAD. The mixture without *Ph*Nah20A was used as a control. Reaction mixtures for subsequent gel permeation chromatography-aided (GPC) separation of transglycosylation products contained 10 U mL^{-1} *Ph*Nah20A, 100 mM NAG-oxazoline and 200 mM lactose in 50 mM sodium phosphate, pH 8, 0.5% BSA and incubated 2 h at 37°C. The reaction was stopped by heating (5 min, 90°C). The sample was diluted 4 times with sterile milliQ water and the enzyme was removed (Amicon Ultra 0.5 mL centrifugal device, Mw cut-off 30 kDa; Merck, USA). Prior to GPC, particles were removed (0.45 µm filters; Millex, Merck, USA).

Chromatographic methods

TLC of reaction products was performed on Silica Gel 60 F254 plates (Merck, Germany). Samples containing 0.2 mg of sugar (1 µL) were spotted onto TLC plates developed by two runs in chloroform:acetic acid:water (6:7:1; v:v:v) (20, 21) or *n*-butanol: ethanol: water (5:3:2; v:v:v) (22). Sugars were visualized by spraying with orcinol dye (0.5 % 5-methyl resorcinol and 10 % H_2SO_4 in ethanol) and subsequent heating of the dried plate.

HPAEC-PAD for saccharide analysis was performed on Dionex ICS-5000 (Thermo Fisher Scientific, USA) and Dionex CarboPacP1 column (4.6x250 mm; Thermo Fisher Scientific, USA).LNT2, D-glucose, D-galactose, lactose, GlcNAc, $GlcNAc_2$ and *N*-acetylated chitooligosaccharides were used as markers and for calibration. D-fucose was used as internal standard in HPAEC-PAD.

Compounds in reaction mixtures were separated using GPC on Bio-Gel P-2 (Bio-Rad, USA) and column dimensions: 16x900 mm (XK16/100, GE Healthcare, Sweden) mounted on an ÄKTAprime plus chromatography system (GE Healthcare, Sweden). Approximately 10 mg sugar was injected (0.5 mL reaction mixture containing 20 g L^{-1} saccharides) and degassed milliQ water was used as an eluent at flow rate of 0.1 $mL \cdot min^{-1}$ at room temperature. Pressure limit was set to 0.3 MPa. Fraction size was 2 mL and reducing sugars were detected and quantified by Nelson-Somogyi method (23) using glucose and *N*-acetyl glucosamine as standard. Reducing sugar-containing fractions were dried (SpeedVac; Thermo Fisher Scientific, USA) at 50°C, dissolved in 50 µL milliQ water. Fractions were assessed by TLC for degree of polymerization and preliminary identification of products. For nuclear magnetic resonance (NMR) analysis, trisaccharide-containing fractions of two identical runs were pooled, dried (SpeedVac evaporator) and dissolved in 0.5 mL D_2O (Sigma-Aldrich, USA).

Synthesis of NAG-oxazoline

NAG-oxazoline (2-methyl-(1,2-dideoxy-α-D-glucopyrano)-oxazoline) was synthesized and purified essentially as described (24). Briefly, 2 g GlcNAc (9

mmol) was dissolved in 20 mL acetic anhydride, then 10 mL pyridine was added and the solution was stirred overnight. After extraction by dichloromethane (DCM) and successive washings (Na_2CO_3, H_2O, H_2SO_4, H_2O), the organic layer was dried and evaporated. Trimethylsilyl trifluoromethanesulfonate (0.8 mL) was added to 1.5 g peracetylated glucosamine dissolved in 1,2-dichloroethane and stirred at 50°C until completion (4 h).
The reaction was stopped by adding 2 mL trimethylamine, diluted with 50 mL DCM, washed with cold water, dried and evaporated. The product was purified by flash chromatography (eluent: cyclohexane:1% triethylamine in ethyl acetate 100:0 to 40:60). To 300 mg of peracetylated oxazole in 10 mL anhydrous methanol at 0°C was added 15 µL 5.3 M sodium methanolate in methanol. The solution was stirred at room temperature until the reaction was completed (≈ 3 h), dried and the resulting NAG-oxazoline was used without further purification.

Nuclear magnetic resonance (NMR)

All spectra were recorded on an 800 MHz Bruker Avance III (799.75 MHz for ^{1}H and 201.10 MHz for ^{13}C) equipped with a 5 mm TCI cryoprobe. Acetone was used as internal reference (2.22 ppm and 30.89 ppm for ^{1}H and ^{13}C, respectively). The following experiments were used for the elucidations: ^{1}H with presaturation, DQF-COSY, ROESY, HSQC, HSQC-TOCSY and HMBC.

Acknowledgements

The Danish Council for Strategic Research is acknowledged for supporting the project grant 1308-00014B “OliGram. Design and gram scale enzymatic synthesis of human milk oligosaccharides.”
The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.
We thank Pernille K. Bech (University of Copenhagen, Denmark) for providing *P. hydrolytica* strain and Karina Jansen for general technical assistance.

References

1. Hehemann JH, Boraston AB, Czjzek M. 2014. A sweet new wave: Structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr Opin Struct Biol 28:77–86.

2. Beier S, Bertilsson S. 2013. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front Microbiol 4:1–12.

3. Machado H, Sonnenschein EC, Melchiorsen J, Gram L. 2015. Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. BMC Genomics 16:158.

4. Aunpad R, Rice DW, Sedelnikova S, Panbangred W. 2007. Biochemical characterisation of two forms of halo- and thermo-tolerant chitinase C ofSalinivibrio costicola expressed inEscherichia coli. Ann Microbiol 57:249–257.

5. Choi KH, Seo JY, Park KM, Park CS, Cha J. 2009. Characterization of glycosyl hydrolase family 3 ??-N-acetylglucosaminidases from Thermotoga maritima and Thermotoga neapolitana. J Biosci Bioeng 108:455–459.

6. Lan X, Ozawa N, Nishiwaki N, Kodaira R, Okazaki M, Shimosaka M. 2004. Purification, Cloning, and Sequence Analysis of β-N-Acetylglucosaminidase from the Chitinolytic Bacterium Aeromonas hydrophila Strain SUWA-9. Biosci Biotechnol Biochem 68:1082–1090.

7. Keyhani NO, Roseman S. 1996. The

Chitin Catabolic Cascade in the Marine Bacterium *Vibrio furnissii* . J Biol Chem 271:33425–33432.

8. Schultz-johansen M, Glaring MA, Bech PK, Stougaard P. 2016. Draft Genome Sequence of a Novel Marine Bacterium , Paraglaciecola sp . Strain S66 , with Hydrolytic Activity against Seaweed 4:4–5.

9. Bech PK, Schultz-Johansen M, Glaring MA, Barbeyron T, Czjzek M, Stougaard P. 2017. Paraglaciecola hydrolytica sp. Nov., a bacterium with hydrolytic activity against multiple seaweed-derived polysaccharides. Int J Syst Evol Microbiol 67:2242–2247.

10. Piszkiewicz D, Bruice TC. 1968. Glycoside Hydrolysis. 111. Intramolecular Acetamido Group Participation in the Specific Acid Catalyzed Hydrolysis of Methyl 2-Acetamido-2-deoxy-~-~-glucopyrano. J Am Chem Soc 378:5844–5848.

11. Macdonald SS, Blaukopf M, Withers SG. 2015. N-acetylglucosaminidases from CAZy family GH3 are really glycoside phosphorylases, thereby explaining their use of histidine as an acid/Base catalyst in place of glutamic acid. J Biol Chem 290:4887–4895.

12. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:206–214.

13. Verma SC, Mahadevan S. 2012. The ChbG gene of the chitobiose (chb) operon of Escherichia coli encodes a chitooligosaccharide deacetylase. J Bacteriol 194:4959–4971.

14. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse J, Lauber E, Arlat M. 2007. Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. PLoS One 2.

15. Nyffenegger C, Nordvang RT, Zeuner B, Łężyk M, Difilippo E, Logtenberg MJ, Schols H a., Meyer AS, Mikkelsen JD. 2015. Backbone structures in human milk oligosaccharides: trans-glycosylation by metagenomic β-N-acetylhexosaminidases. Appl Microbiol Biotechnol.

16. Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234.

17. Garrido D, Ruiz-Moyano S, Mills D a. 2012. Release and utilization of N-acetyl-d-glucosamine from human milk oligosaccharides by Bifidobacterium longum subsp. infantis. Anaerobe 18:430–435.

18. Chen X, Xu L, Jin L, Sun B, Gu G, Lu L, Xiao M. 2016. Efficient and Regioselective Synthesis of BETA-GalNAc/GlcNAc-Lactose by a Bifunctional Transglycosylating BETA-N-Acetylhexosaminidase from Bifidobacterium bifidum. Appl Environ Microbiol 82:5642–5652.

19. Rauvolfová J, Kuzma M, Weignerová L, Fialová P, Přikrylová V, Pišvejcová A, Macková M, Křen V. 2004. β-N-Acetylhexosaminidase-catalysed synthesis of non-reducing oligosaccharides. J Mol Catal B Enzym 29:233–239.

20. Stingele F, Newell JW, Neeser JR. 1999. Unraveling the function of glycosyltransferases in Streptococcusthermophilus Sfi6. J Bacteriol 181:0–6354.

21. Viigand K, Visnapuu T, Mardo K, Aasamets A, Alamäe T. 2016. Maltase protein of Ogataea (Hansenula) polymorpha is a counterpart to the resurrected ancestor protein ancMALS

of yeast maltases and isomaltases. Yeast 33:415–432.

22. Reiffová K, Nemcová R. 2006. Thin-layer chromatography analysis of fructooligosaccharides in biological samples. J Chromatogr A 1110:214–221.

23. McCleary B V., McGeough P. 2015. A Comparison of Polysaccharide Substrates and Reducing Sugar Methods for the Measurement of endo-1,4-β-Xylanase. Appl Biochem Biotechnol 177:1152–1163.

24. André-Miral C, Koné FM, Solleux C, Grandjean C, Dion M, Tran V, Tellier C. 2015. De novo design of a trans-β-N-acetylglucosaminidase activity from a GH1 β-glycosidase by mechanism engineering. Glycobiology 25:394–402.