NMR CHARACTERISATION OF GROUP B STREPTOCOCCUS CAPSULAR POLYSACCHARIDE REPEATING UNITS

Maximillian Ludwig Keresztesi

Supervisors: A. Prof. Neil Ravenscroft and A. Prof. Michelle Kuttel

A dissertation submitted to the University of Cape Town in fulfilment of the requirements for the degree of

Master of Science

Department of Chemistry

University of Cape Town

February 2022

Declaration

I hereby declare that ***NMR Characterisation of Group B Streptococcus Capsular Polysaccharide Repeating Units*** is my own work and has not been presented for the award of any degree at any university. I know the meaning of plagiarism and declare that all of the work in the dissertation, except for that which is properly acknowledged, is my own.

Signed by candidate

_____

February 2022

Abstract

Group B *Streptococcus* (*Streptococcus agalactiae*) is a Gram-positive β-haemolytic bacterium and the leading cause of neonatal mortality by sepsis, pneumonia and meningitis. To date, ten serotypes of Group B *Streptococcus* (GBS) have been recognised (Ia, Ib, II - IX), each identified and differentiated by their sialic acid-containing capsular polysaccharide. Capsular polysaccharides are the virulence factor for bacterial pathogens and the target for vaccine development, with multivalent polysaccharide-protein conjugate vaccines licenced against bacteria such as *Neisseria meningitidis* and *Streptococcus pneumoniae*.

Nuclear magnetic resonance (NMR) spectroscopy has been established as an extremely useful and robust method for tracking the manufacturing process of carbohydrate vaccines from polysaccharide antigen through to conjugate vaccines. The 1D proton profiles of most of the GBS antigens have been published, however, the identity spectra were recorded at 298 K, resulting in broad peaks and overlap of the large water signal with diagnostic GBS signals in the anomeric region. This study attempts to aid the development of GBS glycoconjugate vaccines by fully characterising the repeating units of the six most common GBS serotypes (Ia, Ib, II - V) by NMR recorded at a higher temperature of 343 K to serve as a database of reference GBS NMR spectra and chemical shift assignments.

Full NMR characterisation of the repeating unit of each serotype was achieved by use of an array of 1D and 2D NMR experiments including proton, carbon, proton-proton scalar and dipolar correlation experiments and proton-carbon heteronuclear single-quantum and multiple bond correlation experiments. The assignments of all six serotypes largely agree with NMR data published for these serotypes. The exception to this was GBS V, where data presented in this study shows that the assignments of the anomeric peaks of GlcNAc and the backbone β-Glucose are reversed relative to their assignments in the current literature.

The 1D and 2D NMR spectra presented in this study can be used for identity, integrity and purity testing of polysaccharide batches. They allow identification of each serotype by its diagnostic anomeric peaks, can confirm the structural integrity of the polysaccharide both before and after conjugation and can detect the presence of impurities such as residuals. Ultimately, they represent a powerful reference resource for use in the development, preparation and control testing of future GBS glycoconjugate vaccines.

Acknowledgements

I wish to thank and acknowledge the contributions of the people without whom this dissertation would not exist:

First and foremost, to my supervisors, Assoc. Prof. Neil Ravenscroft and Assoc. Prof Michelle Kuttel, for their wealth of advice, support and, most of all, patience.

To all the students of the PD Hahn Level 6 and 7 research groups that provided advice, feedback, comradery and motivation: Christopher Thurling, Larnelle Garnie, Roxanne Mohunlal, Jason Hlozek, Zaheer Timol, Nicole Richardson and Alexios Vicatos to name a few.

To my friends whose support was some stability during very unstable times.

The University of Cape Town for funding support.

PATH and vaccine partners for providing Group B *Streptococcus* polysaccharide and conjugate samples for analysis.

Table of Contents

Abstract

List of Figures

List of Tables

Abbreviations

COSY	Correlation spectroscopy
CPS	Capsular polysaccharide
DEPT	Distortionless enhancement by polarization transfer
DOSY	Diffusion ordered spectroscopy
EOD	Early-onset disease
Gal	Galactose
GBC	Group B carbohydrate
GBS	Group B *Streptococcus*
GC	Gas chromatography
Glc	Glucose
GlcNAc	N-Acetyl-D-glucosamine
Hib	*Haemophilus influenzae* type b
HMBC	Heteronuclear multiple bond correlation
HOD	Partially deuterated water
HPAEC-PAD	High-performance anion exchange chromatography with pulsed amperometric detection
HPLC	High-performance liquid chromatography
HSQC	Hetereonuclear single-quantum correlation
IAP	Intrapartum antibiotic prophylaxis
LOD	Late-onset disease
Neu5Ac	N-Acetyl neuraminic acid
Nm	*Neisseria meningitidis*
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
RU	Repeating unit
Sp	*Streptococcus pneumoniae*
TOCSY	Total correlation spectroscopy
TT	Tetanus toxoid

Chapter 1. Introduction

Streptococcus agalactiae, also known as Group B *Streptococcus*, is a Gram-positive β-haemolytic bacterium that colonises the gut and genital tracts of 20-30% of healthy adult humans [1,2] and is generally commensal, showing no pathogenicity. Nevertheless, it can cause invasive infections and is the leading cause of neonatal mortality by sepsis, pneumonia and meningitis. While less frequent, it can also cause infections in adults, especially pregnant women and other immunocompromised individuals such as people with comorbidities and the elderly [3,4].

Group B *Streptococcus* (GBS) has been divided into ten serotypes (Ia, Ib, and II - IX) based on the sialic acid-containing capsular polysaccharide [5,6], with the six serotypes covered in this study (Ia, Ib, and II - V) accounting for approximately 98% of global GBS disease cases [7]. This chapter provides a short introduction to the incidence and effects of GBS disease, as well as an overview of GBS vaccine development.

1.1 Early history of Group B *Streptococcus*

Group B *Streptococcus* was first differentiated from other haemolytic streptococci by Rebecca Lancefield in 1933 when she isolated it from the milk of cows suffering from bovine mastitis [8]. While it was originally only of interest as a veterinary pathogen, Lancefield did find GBS colonisation in the vaginal tracts of asymptomatic women. GBS pathogenicity in humans was first described in 1938 when reports of fatal infections in women shortly after they had given birth were published [9]. By the 1960s, GBS had been identified as a cause of severe perinatal infections [10]. In the following years GBS disease reports increased dramatically, either because of an increase in incidence or more accurate diagnosis, until, by the 1970s, it had emerged as a leading cause of neonatal morbidity and mortality in the United States [11], with some mortality rates reported as high as 55% [12]. The rate of neonatal GBS infections remained high throughout the 1980s [13], which prompted the first research into the risk factors for developing GBS disease as well as the susceptibility of the bacterium to antibiotics.

1.2 Group B Streptococcal disease

1.2.1 Group B *Streptococcus* disease in neonates

As mentioned, GBS was first implicated in neonatal infections in 1961 [10], wherein GBS colonisation of the mother was present in an abnormally high percentage of difficult pregnancies. The majority of infant infections, approximately 80% [13], occur only a few days after birth, known as early-onset disease (EOD). Conversely, infections that occur between 1 week and 3 months after birth are referred to as late-onset disease (LOD). Early-onset infections can occur during pregnancy as a result of the bacterium spreading up into the amniotic fluid in mothers with GBS colonisation in the genital tract. Aspiration of the contaminated amniotic fluid can then lead to invasive disease and even result in spontaneous abortions or stillbirth [14]. Another source of EOD infection occurs during birth as the infant passes through the vaginal tract. Late-onset infections are sometimes also thought to be caused during birth, but are generally attributed to horizontal transmission from external sources [15].

Early-onset disease typically manifests as bacteremia, pneumonia or meningitis, but other syndromes include bone, joint or soft tissue infections [16]. The mortality rates of EOD vary from approximately 5% in developed countries to as high as 27% in resource-poor areas such as Africa [7]. Premature infants and infants with low birth weight have also been shown to be at higher risk, displaying a higher mortality rate [16]. Late-onset disease presents as meningitis more commonly than EOD does, but it may similarly manifest as bacteremia or pneumonia as well as bone, joint, soft tissue or urinary tract infections [17]. Globally, the average mortality rate of LOD is approximately 7% with Africa again having the highest rate of 12% [7]. Additionally stillbirths due to GBS account for approximately 1% of stillbirths globally and approximately 4% of stillbirths in Africa [14].

In addition to the risk of death that comes with neonatal GBS disease, it can also be responsible for long-term disabilities. Chronic sequelae arising from neonatal GBS disease include mild, moderate or even severe neurological impairment, and can occur in up to 50% of GBS disease survivors [18,19].

1.2.2. Group B *Streptococcus* disease in adults

A large subsection of the adult population that suffers from invasive GBS disease is pregnant women, with pregnant women showing invasive disease incidence twice as high as that of non-pregnant women [20]. Invasive GBS infections in pregnant women are generally bacteremia, but meningitis, pneumonia and bone/joint infection are also possible. Non-invasive syndromes, on the other hand, commonly include urinary tract infection, amnionitis and wound infections. Additionally, premature labour and rupture of membranes caused by GBS infection presents a risk to both mother and foetus [20].

Non-pregnant adults do, however, also suffer from GBS infections, though the most common patients are elderly or are otherwise immunocompromised. In fact, the incidence of GBS disease among non-pregnant adults has seen a marked increase from the 1990s [4]. The syndromes associated with GBS disease in non-pregnant adults most commonly include bacteremia or skin/soft tissue infection, with pneumonia and meningitis occurring less frequently [21]. Most adults with invasive GBS disease suffer from some sort of comorbidity, including heart, kidney, neurologic or liver disease as well as alcohol abuse, smoking or obesity. Chronic diseases such as diabetes mellitus and cancer are also often present as comorbidities. Age is also an important factor since the average age of infected non-pregnant adults is 63 years [4].

1.3 Group B *Streptococcus* across the globe

Globally, the rate of GBS colonisation in pregnant women ranges between 7% and 30% [2], with serotype distribution varying quite drastically in different geographical regions. As mentioned there are ten currently recognised serotypes [5,6], of which Ia, Ib, II, III and V are the most common in Europe and the United States [22,23]. Colonisation in the United Arab Emirates and Egypt is dominated by serotype IV and V, respectively [24,25]. On the other hand, serotypes VI and VII are the most common in Japan [26], while the most recently identified serotype, IX, was identified in Denmark [6]. This geographically dependent serotype diversity is one of the reasons that effective GBS vaccines would need to be multivalent. When it comes to invasive infections, serotype III accounts for the majority of late-onset disease globally [27], while the majority of invasive disease in non-pregnant adults is caused by serotypes Ia and V [27].

With the introduction of intrapartum antibiotic prophylaxis (IAP) in the 1990s, the incidence rate of early-onset GBS disease in developed countries, such as the United States, has seen a significant decrease (87% decrease from 1.7 cases/1000 live births in 1993 to approximately 0.21 cases/1000 live births in 2019 [3]). Similar data for lower income countries, however, only shows a slight decrease from 1.21 to 1.12 cases/1000 live births in 2017 [7]. Additionally, IAP has been shown to have minimal impact on the incidence of LOD, since it primarily offers protection during and immediately after birth [28].

1.3.1 Prevalence of Group B *Streptococcus* in South Africa

While the majority of invasive GBS disease research has been performed in developed countries, GBS is of greater concern in third world and developing regions such as Africa where the incidence and fatality rates of GBS disease are approximately twice as high as those in Europe and the United States [29]. This is likely due, in part, to the fact that implementation of screening and intrapartum antibiotic prophylaxis has been significantly less successful in middle- to low-income countries [28,30]. The incidence rate in South Africa is even more dire, approximately twice that of the rest of the African continent [19,29].

South Africa has been the subject of some GBS disease research since the late 1970s, with the first report on neonatal GBS disease published in 1978 [31]. This was followed by studies into the disease burden of GBS on infants of African and Indian descent [32,33], wherein infection rates higher than those recorded in other studies in both developed and developing countries were observed. Another study, in 2003 [30], again noted a GBS disease burden higher than that of developed countries and showed that the risk-based strategy being employed at the time was either ineffective or was not being properly implemented. A number of studies since 2010 have also shown that the high incidence of maternal HIV in South Africa is also an aggravating factor, showing that HIV-infected neonates had a significantly increased risk of both EOD and LOD [19,34,35]. One of the recent studies evaluated the burden of neurological sequelae in infants, noting that approximately 13% of 6 month-old infants that had survived invasive GBS disease displayed potential neurological development problems [19]. A recent study has also shown that methods for disease detection and microbiological diagnosis in South Africa, especially in areas outside the major cities, are still not at the level where a screening-based approach is logistically feasible [36].

The picture these studies paint is that South Africa has a distressingly high GBS disease incidence rate, exacerbated by a high incidence of maternal HIV, and that said rate has not significantly decreased in the last three decades, even though a risk-based IAP strategy has been implemented. Since resources for disease detection and microbiological diagnosis are limited, a screening-based IAP approach is not viable, and the risk-based strategy is either ineffective or poorly implemented. IAP treatment of all newborns could reduce the incidence of EOD but would not have any effect on LOD or in utero infection incidence. Additionally, this method may contribute to the emergence of anti-microbial resistance, with clindamycin-resistant strains already causing approximately 40% of GBS infections, thus limiting treatment options for people with penicillin allergies [37,38]. The need for a better method of GBS disease prevention for South Africa, and the fact that that method is most likely vaccination, has been noted in a number of studies since the early 2000s [19,30,35,36,39].

1.4 Cell surface polysaccharides of Group B *Streptococcus*

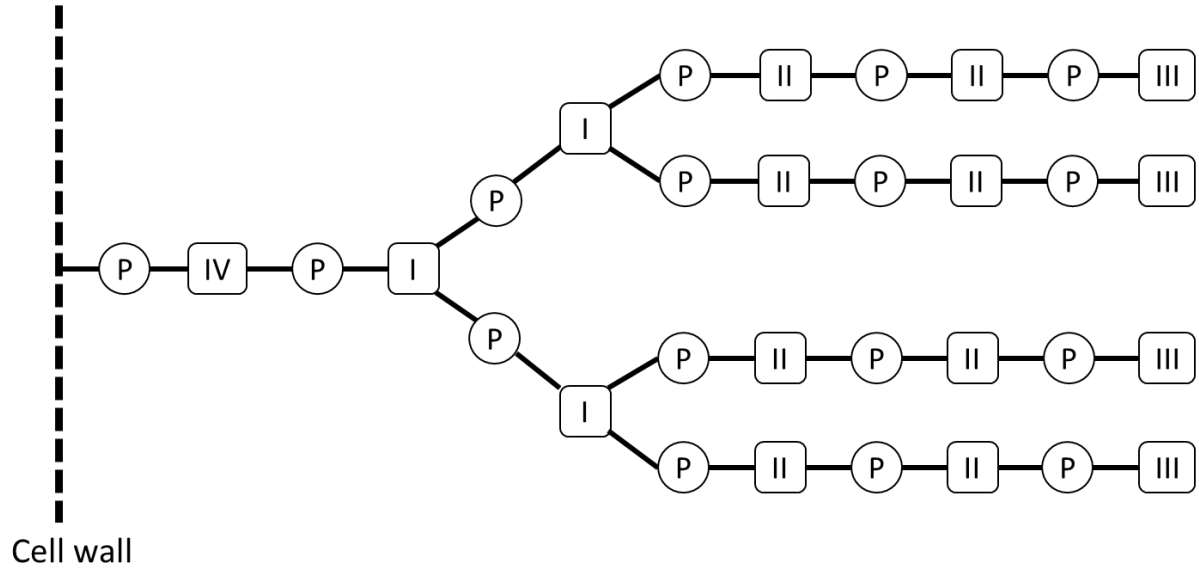
Bacterial cell surface polysaccharides are found covalently bonded to the external surface of the cell. They have been recognised as important factors in bacterial virulence for years [40]. They are polymers mostly composed of monosaccharides arranged in a repeating unit (RU) but can also include non-carbohydrate substituents such as acetate or phosphate. In addition to this, they can be heteropolymeric, constituting multiple different carbohydrate monomers, for example Group B carbohydrate described below (Chapter 1.4.1). Considering the great variety with which monosaccharide units can join to each other within the repeating unit, this presents an enormous structural diversity that bacteria have taken full advantage of.

Since the CPS often contains numerous negatively charged and polar groups, they are highly hydrated, containing more than 95% water. This helps the bacterium avoid desiccation, allowing it to survive longer in the environment, which aids in transmission [41]. Another potential function of the CPS is to aid in adhesion to surfaces both inside and outside a host organism [42], which helps with colonisation. Due to their position on the surface of the bacterium they are also the first antigen exposed to the immune system of the host during a pathogenic infection and thus play an important role in bacterial survival against the non-specific immune system [43]. Additionally, some CPSs can confer some resistance against the specific, or antibody-mediated, immune response, especially if those polysaccharides contain molecular motifs that mimic ones found in the host biology. As will be discussed later, GBS

capsular polysaccharides perform this function, since they all contain sialic acid (Neu5Ac) units, which is a sugar unit also found in human biology, making it difficult for the human immune system to identify the GBS cells as non-self.

1.4.1 Group B carbohydrate

The surface of the GBS bacterium is characterised by a number of these carbohydrate polymers, chief among which is the Group B Carbohydrate (GBC). It is present in all serotypes of GBS and is composed of glycosyl-phosphates of rhamnose, galactose, N-Acetyl-D-glucosamine and glucitol linked to each other by phosphodiester bonds [44]. It is linked to the peptidoglycan layer of the cell through a phosphodiester bond and, as can be seen in Figure 1.1, is a complex, branched structure. Because it is unique to the Group B streptococcal bacteria, and present in all GBS strains, presence of GBC is used as the diagnostic indicator in antigen assays to identify the presence of GBS [45].



I: α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→4)-α-L-Rha*p*-(1→2)-[α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→4)]-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→1')-[α-L-Rha*p*(1→3')]-D-Glucitol-**P**-

II: α-L-Rha*p*-(1→2)-[α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→4)]-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→1')-[α-L-Rha*p*(1→3')]-D-Glucitol-**P**-

III: α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→2)-α-L-Rha*p*-(1→1')-[α-L-Rha*p*(1→3')]-D-Glucitol-**P**-

IV: α-L-Rha*p*-(1→3)-α-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-α-L-Rha*p*-(1→3)-β-L-Rha*p*-(1→4)-D-Glc*p*NAc-**P**-

P: Phosphodiester

Figure 1.1: Proposed structure of the Group B Carbohydrate and line structures of its constituent oligosaccharides [44].

1.4.2 Capsular polysaccharides

As mentioned, the ten serotypes of GBS are serologically grouped according to their capsular polysaccharides (Table 1.1). Nearly all of these serotype-specific capsular polysaccharides consist entirely of only four different kinds of monosaccharide unit, namely, D-Galactose (Gal), D-Glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc) and N-Acetylneuraminic acid (Neu5Ac or sialic acid).

Table 1.1: Currently identified Group B *Streptococcus* serotypes and their capsular polysaccharide repeating unit structures. Backbone residues are highlighted in bold.

Serotype	Repeating unit
Ia	**→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
Ib	**→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→3)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
II	**→2)-**[α-D-Neu*p*NAc-(2→3)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-**[β-D-Gal*p*-(1→6)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→3)-β-D-Glc*p*-(1→**
III	**→6)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)]**-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
IV	**→4)-α-D-Glc*p*-(1→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
V	**→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]**-α-D-Glc*p*-(1→4)-**[β-D-Glc*p*-(1→3)]**-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
VI	**→6)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→3)]**-β-D-Glc*p*-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
VII	**→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]**-α-D-Glc*p*-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**
VIII	**→4)-**[α-D-Neu*p*NAc-(2→3)]**-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→4)-β-D-Glc*p*-(1→**
IX	**→4)-**[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]**-β-D-Glc*p*NAc-(1→4)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→**

The one exception to this is serotype VIII which contains a rhamnose sugar. This means that the variety between the CPSs of each serotype is solely due to differences in molar ratio, linkage position and linkage configuration between the monosaccharide units. These capsular

polysaccharides play a major role as a virulence factor for GBS, a fact that is linked to their unanimous terminal sialylation. This is because it has been shown that sialic acid inhibits activation of the complement system of the human immune system as it is prevalent on many human cells, thereby allowing the cell to avoid phagocytosis. This is reinforced by the fact that bacteria that have had the sialic acid residues removed (whether by enzyme or mutation) elicit an immune response as normal [46,47].

1.5 Treatment and prevention of GBS disease

1.5.1 Intrapartum antibiotic prophylaxis

With the rate of invasive GBS infections remaining high in the 1970s and 1980s, clinical trials of intrapartum antibiotic prophylaxis (IAP) were started in the United States and Britain in the 1980s using penicillin and ampicillin since they had been shown to be effective against GBS [48,49]. This was shown to be effective at preventing EOD, but the mortality rate among neonates that did still get GBS disease remained at approximately 15% [50], drawing public attention to GBS as a fatal disease that was readily preventable. It was only the 1990s, however, that protocols for screening, diagnosis and prevention of EOD were published [51]. Candidates for IAP were chosen based on the risk-factors established in the guidelines or after a positive result for GBS colonisation. After the adoption of these guidelines in developed countries, the mortality rate of EOD began to decrease, dropping to approximately 5% [52]. While the reduction in the incidence of EOD and neonate mortality is admirable, the rate of late-onset infection and disease in pregnant women has not noticeably decreased [28]. This is important because LOD often presents as meningitis which, as mentioned, has a high mortality rate and is responsible for significant long-term sequelae [18]. Other challenges with IAP include the need for universal screening, the possibility of vaginal colonisation after screening, the need for IAP to occur timeously (at least 4 hours before delivery) and the possibility of developing antimicrobial resistance among GBS strains. These challenges affect the feasibility of IAP in lower- to middle-income countries since the laboratory infrastructure and antenatal care required for successful implementation of IAP is limited. Some of these issues could be solved with the alternate strategy of maternal vaccination.

1.5.2 Polysaccharide-based vaccines

The seed of the idea of a capsular polysaccharide-based GBS vaccine stems from the early days of GBS research when it was shown that mice could be protected against GBS infection by inoculation with a CPS-specific serum [53]. Over the following decades further research proved that the different CPSs produced by each serotype of GBS are targets of protective immunity [53], that low CPS-specific antibody levels in the mother increase the risk of GBS disease in the infant [54] and that serotype-specific capsular polysaccharide antibodies are successful at preventing infection in an animal model [55]. Additionally, it was shown that a sufficient level of GBS serotype-specific capsular polysaccharide antibody in the mother prevents colonisation in the rectal and vaginal tracts [56] as well as preventing invasive GBS disease in infants [57].

The first clinical GBS vaccine trial occurred in 1978 [58] on 33 healthy adults to investigate the safety and immunogenicity of a type III capsular polysaccharide vaccine. These and subsequent trials showed that the vaccines were safe, even in pregnant women, but their immunogenicity was low. Nonetheless, these studies proved that vaccination would be a feasible approach to preventing invasive GBS disease if the immunogenicity of the candidate vaccines was increased. By the 1980s, a number of trials had been performed, safely vaccinating over 300 healthy adults [59], and the capsular polysaccharides had undeniably been shown to be promising vaccine candidates, but were not immunogenic enough to effectively act as vaccines on their own. Thus, research began to focus on methods by which the immunogenicity of the CPS vaccines could be improved while keeping the polysaccharide antigens intact.

1.5.3 Glycoconjugate vaccines

The success of conjugate vaccines, vaccines in which the antigen has been conjugated to a protein, in the fight against *Haemophilus influenzae* type b (Hib) certainly influenced and spurred on efforts to implement the same technology in an attempt to improve the immunogenicity of GBS vaccines. The first CPS glycoconjugate vaccines of GBS were developed in 1990 wherein the type III CPS was coupled to tetanus toxoid (TT) either with a spacer molecule [60] or by direct linkage of aldehydes formed on some of the sialic acid residues [61]. Significantly, despite the different conjugate vaccines having different coupling methods, polysaccharide size and purification method, they were all more immunogenic than

the attempts at a plain polysaccharide-based vaccine [60,61]. Once the coupling methods had been refined, researchers turned their attention to producing conjugate vaccines against the other GBS serotypes such as Ia, Ib and II and testing them as a tetravalent vaccine in a mouse model [62]. The study showed that the four serotypes did not interfere with each other, meaning there was the potential that a multivalent vaccine could be developed for humans. With the increase in GBS disease arising from serotype V, research has also been conducted towards a serotype V glycoconjugate vaccine linked to either TT or CRM_{197} to be included in a multivalent vaccine [63]. The frequency of serotype IV strains causing invasive GBS disease has also been on the rise in recent years [64], so it's inclusion in any multivalent also seems necessary. Due to the aforementioned significant geographical variation of GBS serotypes, the broad coverage that a multivalent vaccine provides would also make it more useful in the global context. As it stands, some trivalent glycoconjugate vaccine candidates against GBS have passed phase I and II clinical trials but have currently not progressed to phase III [65], while hexavalent glycoconjugate vaccines including the six GBS CPS serotypes covered in this study have been tested in an animal model [66] and a phase 1/2 trial in healthy non-pregnant adults [67].

1.6 Nuclear Magnetic Resonance (NMR)

1.6.1 Short history of NMR technology

Nuclear magnetic resonance is a spectroscopic method that observes the phenomenon that occurs after a collection of nuclear spins in an applied magnetic field has been disturbed by electromagnetic radiation of a certain frequency. The frequency of this resonance is dependent on the chemical environment of the nucleus, as well as the neighbouring nuclei it is bonded to.

This phenomenon was first observed independently by two physicists, Felix Bloch [68] and Edward Purcell [69], in 1945 and was originally seen as an unwelcome complication, but chemists saw this relationship as a way to gain insight into the structure of molecules to a level of detail not previously possible. In the intervening 75 years, NMR analysis has made many great strides forward to become the leading structural analysis technique [70,71], alongside mass spectrometry. Among these advancements are the inclusion of computers in the 1960s leading to radical improvements in experiment time, sensitivity and resolution, as well as the implementation of the Fourier transform method providing a significant increase

in the signal-to-noise ratio [70]. In the 1970s, superconducting magnets allowed for much higher field strengths than previously available, making ^{13}C NMR more commonplace and setting the foundations for observation of other nuclei with low natural abundance. The 70s also saw the emergence of 2D NMR experiments, which would come to be vital in the analysis of increasingly complex molecules [72]. The 1980s and 90s saw the development of an enormous number of new pulse sequences and techniques, such as non-uniform sampling [73], an acquisition method that significantly reduces the time it takes to run 2D experiments by reducing the number of data points and using special reconstruction methods to convert that data into a spectrum as well as diffusion ordered spectroscopy (DOSY) [74] a pulse sequence that suppresses signals from low molecular weight compounds such as solvent. The 90s also saw the coupling of other analytical methods such as liquid chromatography with NMR. Various technological advancements have resulted in more compact instruments with smaller magnetic stray fields and even spectrometers that do not need cryogens, allowing for bench-top spectrometers. However, these are still relatively low-field, so their utility is limited at present. Alongside these advancements, the maximum field strength of NMR spectrometers has been increasing at a rate of approximately 10-20 MHz per year since the 1950's [71], with the strongest commercial NMR unit currently being 1.2 GHz. The main advantages of this increase in field strength are an improved signal-to-noise ratio and a more spread-out spectrum, reducing the overlap between adjacent peaks.

1.6.2 NMR analysis of carbohydrates

Relative to other organic molecules, carbohydrates (and especially complex carbohydrates) have been underrepresented in biological chemistry research. This is due, in part, to the large variety of monosaccharide subunits (over 100 monosaccharides, compared to 4 bases of DNA and 20 amino acids) and the multitude of ways these monomers can connect. The aforementioned bacterial polysaccharides are a good example of this diversity, being long carbohydrate polymers comprised of an oligosaccharide repeating unit (RU). Since these repeating units are structurally identical, their NMR signals are similarly identical (aside from the terminal RUs, but their signals are of negligible intensity in long chain polysaccharides). This means that defining the structure of the repeating unit also defines the structure of the entire polysaccharide. The properties of a repeating unit that describe its structure are the types and molar ratios of monosaccharide units present in the RU, the configurations they are

in (i.e. α- or β-sugar), the presence of non-carbohydrate substituents and finally the sequence and positions of the linkages between these sugar units.

A wide variety of chemical methods are implemented to perform the structural analysis of a polysaccharide, beginning with various acid-based colorimetric assays to determine the purity of the polysaccharide [75]. This is followed by acid hydrolysis and analysis by a chromatographic method such as high-performance liquid chromatography (HPLC), gas chromatography (GC) or high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), which not only provides information on which monosaccharide units the polysaccharide contains, but also their relative molar quantities. Once the composition of the polysaccharide has been determined, the linkage pattern can be elucidated by methylation analysis, a powerful technique in which the carbohydrate hydroxyl groups are methylated and acid hydrolysed [76]. The resulting partially methylated monosaccharides are reduced and acetylated and the partially methylated alditol acetates can then be separated by GC coupled with a mass spectrometry detector [77]. While this method does elucidate the linkage positions of each sugar monomer, it does not, however, provide the sequence in which they are linked, or indeed whether the glycosidic linkages are α- or β-linkages. This information can be acquired by a variety of other degradation methods, mostly selective hydrolysis, but others include, acetolysis, Smith degradation, oxidation with chromium trioxide [78] or even enzymatic methods [79], but these methods are complex and time-consuming.

Early NMR only served a supporting, confirmatory role in the structural elucidation of oligo- and polysaccharides, but with improvements in NMR technology, it has now become possible to fully elucidate the structure of even complex polysaccharides using solely NMR methods [80,81]. This is because NMR can provide, depending on experiment, all the information the aforementioned chemical experiments provide, such as monosaccharide composition and ratios, linkage patterns, sequences and configurations, as well as the position and extent of non-carbohydrate substituents [82]. The biggest advantages of NMR methods over chemical ones are that they are non-destructive, meaning that all the structural information can be obtained by experiments performed on one sample and only requires a relatively small quantity of sample, which is especially important for compounds that are difficult to produce, as many complex carbohydrates are. The main disadvantage of NMR at the moment is that it

is reliant on infrastructure that is significantly more expensive than that required by chemical methods.

1.6.3 Potential of NMR in GBS vaccine development

As will be further examined in Chapter 2, NMR analytical methods can be effectively used in the development of a GBS vaccine and in polysaccharide vaccine development in general for the purposes of identity and purity testing, among others. However, to perform identity and purity testing, reference spectra and proton/carbon assignments of the polysaccharides in question, such as the GBS CPS, are needed. While partial proton assignments, primarily of the anomeric protons, and full carbon assignments have been published of some of the GBS CPS serotypes [83], full proton and carbon assignments have only been published for GBS type III [84]. However, the NMR data currently in the literature was recorded in the temperature range of 290 – 323 K which, as will be shown in Chapter 3, is not the ideal temperature for recording NMR data of GBS CPSs. Thus, there is a need for reference spectra and proton/carbon assignments of the various GBS CPS serotypes at the temperature of 343 K to aid the role of NMR in the area of GBS vaccine development.

1.7 Scope of work

Aims

The aim of this study was to provide a comprehensive NMR analysis of the structures of the capsular polysaccharides of Group B *streptococcus* serotypes Ia, Ib, II, III, IV and V, including complete proton and carbon assignments at 343 K and present them for use as reference spectra and assignments for identity, integrity and purity testing to facilitate the development of GBS capsular polysaccharide and conjugate vaccines. This data will hopefully also allow the creation of a standardised database of GBS CPS NMR spectra and chemical shift data.

Objectives

The objectives of this study were:

1. Acquire 1D and 2D NMR data including ^{1}H, ^{13}C, ^{1}H-^{1}H scalar and dipolar correlation spectra and ^{1}H-^{13}C heteronuclear single-quantum and multiple bond correlation spectra of the six selected GBS CPSs at 343 K, with the data being of sufficient quality

to be used as reference spectra for the purposes of identity, integrity and purity testing of GBS polysaccharide and conjugate vaccine samples.

2. Process the data and establish an approach and methodology for the assignment of every proton and carbon in the CPS repeating unit to achieve the full NMR characterisation of the GBS III CPS, the most prevalent and thus most widely studied serotype.

3. Apply this approach and methodology to perform the full NMR characterisation of the repeating unit structures of the five remaining GBS capsular polysaccharide serotypes covered in this study.

4. Present the NMR analysis and chemical shift data obtained for use as a reference database for GBS CPS serotypes, accompanied by the 1D and 2D NMR identity spectra for each serotype.

The structure of this thesis is outlined below:

Chapter 1 is the introduction and covers Group B *Streptococcus* disease, as well as providing a short description of NMR and its potential in polysaccharide vaccine development.

Chapter 2 covers the NMR experiments performed in this study, as well as how NMR has been used in carbohydrate-based vaccine development in the past.

Chapter 3 establishes and describes, in detail, the methods and analytical approach of the full assignment of the GBS serotype III CPS using NMR data.

Chapter 4 applies the established methods to the NMR data of the two similar CPSs of GBS Ia and Ib, providing the full assignment of both, as well as a comparison between the two.

Chapter 5 describes the full characterisation of the remaining three serotypes; GBS II, IV and V, using the methods established in the previous chapters.

Chapter 6 describes how the NMR data can be used in GBS polysaccharide and conjugate vaccine development.

Chapter 7 contains the conclusions of the study.

Chapter 2. NMR Characterisation of Polysaccharides

As highlighted, NMR has immense potential for the analysis of bacterial polysaccharides and of carbohydrates overall. This chapter will provide an overview of the most common NMR experiments performed in the analysis of carbohydrates, as well as examples of previous carbohydrate-based vaccine research that have made use of NMR.

2.1. NMR experiments

The NMR experiments described in this chapter use standard Bruker pulse sequences, listed in Table 2.4 later in this chapter.

2.1.1. 1D ^{1}H, ^{13}C and ^{31}P experiments

The 1D ^{1}H and ^{13}C (and phosphorous, if applicable) experiments are the simplest in both execution and presentation. They are the starting point for any NMR characterisation, with the ^{1}H spectrum also acting as the reference against which future samples are compared once the analysis is complete. Due to the significantly lower comparative sensitivity of ^{13}C NMR, it is only used when there is sufficient material.

There is extensive overlap in the ^{1}H spectrum of even relatively simple oligosaccharides, especially around the ring proton region of 3.5-4.4 ppm, hence most of the information to be gleaned from the 1D ^{1}H spectrum relates to significantly shielded or deshielded signals that lie outside of that range. Chief among these are the anomeric (H1) signals, which lie between 4.4-5.4 ppm [85–88] and give not only an indication of how many aldose monosaccharides the polysaccharide repeating unit is comprised of, but also their configuration, with α-sugars appearing between 5.0-5.4 ppm and β-sugars appearing between 4.4-5.0 ppm [85–88]. High-field outliers include signals such as H2 of β-Glc at 3.3-3.4 ppm, H3's of 3-deoxy keto sugars like sialic acid at approximately 1.8 and 2.8 ppm [89], methyl signals for O- and N-acetyl groups at 2.0-2.2 ppm and the H6's of 6-deoxy sugars at 1.1-1.4 ppm. Assignment of the rest of the proton signals will rely on more complex correlation experiments to deconvolute the overlapped signals.

The ^{13}C spectrum (nearly always proton-decoupled for simplification) is even more diagnostic than the proton and can also be used as the fingerprint of a compound, since it provides sharp, well-separated signals across a large range (0-200 ppm), but its low sensitivity (approximately 10 000 times less sensitive than ^{1}H NMR) means it can only be used in cases where sufficient

material is available. The chemical shift regions seen in the ^{1}H spectrum are matched in the ^{13}C, with anomeric, ring and methyl regions. Carbonyl carbons appear at extreme low field, with the ones resulting from carboxylic acids or acetyl groups appearing at 170-180 ppm [85–88]. The carbon anomeric region (90-110 ppm) gives a true count of how many monosaccharide units a polysaccharide contains, since it shows the C1's of aldose and C2's of ketose sugars. The anomeric region can also allow differentiation between furanose (100-110 ppm) and pyranose (90-105 ppm) sugars [88]. The ring region occurs between 65-85 ppm and is crowded like the ^{1}H spectrum, but with significantly less overlap. While the ring signals are still too numerous to allow for full assignment by inspection, they are noticeably influenced by nearby glycosylation, experiencing a significant downfield shift of 5-10 ppm at linkage positions and an upfield shift of 0.5-2 ppm when adjacent to a linkage position [87], allowing for easier identification and confirmation of linkage positions. Further upfield (60-65 ppm, but 68-70 ppm when glycosylated), any hydroxymethyl signals, which are lost among the ring signals in the ^{1}H spectrum, appear. Continuing upfield, the ring carbons linked to O- or N-acetyl groups (such as H2 of GlcNAc) appear at 45-60 ppm, the C3 of 3-deoxy keto sugars such as sialic acid at about 40 ppm [89] and the methyl carbons of acetyl groups and 6-deoxy sugars at 20-25 ppm. An additional ^{13}C experiment that can aid in assignment is DEPT-135 (Distortionless Enhancement by Polarization Transfer) [90], which differentiates carbons by the number of attached protons, leaving methyl and methine signals positive, while inverting methylene signals and completely eliminating quaternary carbon signals. Since ^{13}C spectroscopy is significantly less sensitive than ^{1}H, the carbon profile of samples with extremely low availability can alternatively be identified using a proton-detected experiment such as HSQC [91], described later.

While not used in this study, some polysaccharides may contain phosphorous which can be identified by the use of ^{31}P NMR. The presence of phosphorous can be seen in ^{1}H and ^{13}C spectra by significant glycosylation shifts (a chemical shift value that is different than what would be predicted were the proton or carbon unlinked), as well as distinctive splitting in adjacent protons and carbons. The two most common phosphorous moieties, phosphodiester and phosphomonoester groups appear at around 0 and 5 ppm respectively [87].

2.1.1.1 Water suppression experiments (presaturation, DOSY)

A complication in the proton-detected spectra of a sample in deuterated water is the greatly reduced, but still significant, water (HOD) peak. The chemical shift of this peak is heavily temperature dependent, shifting upfield at higher temperatures, and can occur between 4.3 and 4.8 ppm at temperatures between 343 K and 298 K, respectively. This region coincides with the aforementioned anomeric region of the carbohydrate spectrum (4.4 to 5.4 ppm), meaning it often overlaps with and, due to its large intensity, completely obscures diagnostic carbohydrate signals. To avoid this, a number of water suppression methods are available, the first one used in this study being the use of presaturation [92], in which the frequency of the water signal is selectively saturated, but this eliminates all signals at that frequency, thus both the water signal and other signals that may be of interest are removed. The second method used is diffusion ordered spectroscopy (DOSY) [93], whereby magnetic field gradient pulses allow the suppression of signals arising from molecules that have large diffusion through the NMR tube, which is directly dependent on the hydrodynamic radius of said molecules. This means that signals arising from small molecules, such as water, are suppressed, while those from large molecules, such as polysaccharides, are not. This has the desired effect of removing the water signal near the anomeric region without also eliminating any signals that it may be overlapped with.

2.1.2. Scalar homonuclear correlation experiments (COSY, TOCSY)

Building on the starting point of the 1D ^{1}H spectrum, ^{1}H-^{1}H correlation spectroscopy (COSY) [94] and total correlation spectroscopy (TOCSY) [95] experiments allow the partial deconvolution of the ring signal region by showing which signals belong to which spin system and also the order in which they occur in said spin system. COSY only shows correlations between protons on adjacent carbons (vicinal coupling), meaning that correlations can be seen from diagnostic signals such H1 to H2, then from H2 to H3 and so on, as long as there is minimal overlap with other proton signals. TOCSY allows the transfer of magnetisation to continue passing through the bonds of the ring to show signals across more, if not all, of the spin system, meaning that correlations from H1 to H2, H3, H4 or even further can be seen. TOCSY provides more information than COSY but is less specific since the order of correlation is not immediately apparent in the TOCSY. Thus, COSY and TOCSY are often used in tandem, as an overlay. In this way TOCSY provides the signals for the spin system, while the COSY

provides information on how some of those signals are connected. The amount of information these two spectra provide is related to two factors: the identity of each sugar residue and the mixing time of the NMR experiment. The identity is determined by the axial/equatorial configurations of the ring protons, which in turn determine the coupling constants between said ring protons. This means, for example, that the transfer of magnetisation in the TOCSY of a β-Gal residue ends at H4 due to the equatorial H4 resulting in an extremely small $J_{H4,H5}$ coupling constant, as opposed to β-glucose which has exclusively axial ring hydrogens, meaning correlations can often be seen all the way from the anomeric to the H6 signals. The mixing time is a period in the NMR experiment during which the aforementioned transfer of magnetisation is allowed to occur and is directly proportional (to a point), to how far the magnetisation travels around the ring. Longer mixing times result in correlations through a larger number of bonds, but a mixing time that is too long results in spin relaxation which causes a decrease in signal intensity.

Additionally, 1D TOCSY experiments can be performed that selectively irradiate well-resolved signals identified in the 1D ^{1}H spectrum [96], providing higher-resolution projections of the 2D TOCSY. Not only are these useful for resolving ambiguities in the ring region by showing more detail on the shape of the correlation peaks, they can also be used as a way to determine the order of correlation through the ring. This is done by running a set of experiments, increasing the mixing time in a stepwise fashion each time, to show how each subsequent correlation sequentially builds up over longer mixing times as the magnetisation moves around the ring.

The water suppression methods used in these 2D experiments were similar to those used for 1D experiments, with presaturation being the most common. However, to gain information about signals under and around the water signal, DOSY-TOCSY experiments were also performed. These are diffusion filtered TOCSY experiments, meaning that, like conventional 1D DOSY, signals arising from small molecules are suppressed.

2.1.3. Dipolar homonuclear correlation experiments (NOESY)

As described, since the transfer of magnetisation in the COSY and TOCSY experiments may not always reach all the way around the ring via scalar coupling due to small coupling constants between ring protons, an alternate method must be used to identify all the protons in each spin system. Nuclear Overhauser Effect spectroscopy (NOESY) [97] can provide this

solution because it displays correlations that rely on dipolar coupling through space, instead of bonds, meaning that intra-residue correlations between ring protons that are spatially close to each other, such as diaxial H1-H3 or -H5, can be detected. This allows signals that are not reached through TOCSY to be identified through already assigned signals near them, such as H5 of Gal or H6, 7 and 8 of sialic acid. NOESY also allows inter-residue correlations to be seen, which can be very useful to confirm how residues are linked, as an anomeric proton will generally be spatially close to the ring protons on and around the linkage position on its neighbour. Unfortunately, this can also cause even more confounding overlap and because of this, NOESY is often used in conjunction with TOCSY spectra. An overlay of COSY, TOCSY and NOESY often contains enough information to assign most, if not all, of the spin system cross-peaks.

Like TOCSY, selective 1D NOESY experiments [98] can also be recorded to get higher-resolution projections of the 2D spectrum. A major difference between the two experiments, however, is that the variation of peak intensity at varying mixing times is reflective of the physical distance between two protons, not the distance through bonds. This means that, if desired, they can be used to calculate the distance between two protons for aid in conformational analysis [99].

2.1.4. Heteronuclear experiments (HSQC)

Heteronuclear single-quantum correlation (HSQC) spectroscopy [91], which correlates protons to their bonded carbons, serves two important purposes: to assign unidentified carbon signals using the proton assignments determined from the COSY, TOCSY and NOESY, as well as act as the final presentable proof of full structural elucidation, since it contains a cross-peak for every proton-carbon pair in the polysaccharide. The HSQC assignment of cross-peaks outside of the ring region is trivial when comparing both the 1D ^{1}H and ^{13}C data, but elucidation of the ring region itself is still complicated by the significant proton overlap. One of the measures to aid in deconvolution of the ring region is to record an HSQC-DEPT experiment which, much like for the 1D ^{13}C experiment, inverts methylene signals, meaning the H6 signals that are overlapped in the ring region are easier to single out and assign.

Another method used to aid in assignment is the use of hybrid HSQC experiments, such as HSQC-TOCSY [100] which, like conventional ^{1}H 2D TOCSY, shows scalar coupling correlations through the bonds of the spin system, except now the cross-peaks are not to coupled protons,

but instead the carbons those protons are bonded to. This is especially useful when combined with the normal HSQC and selective 1D TOCSY experiments, since the carbons each anomeric signal couples with will be arrayed vertically along the signal, and the protons arrayed horizontally, giving an effective means of assigning the HSQC cross-peaks. When needed, an HSQC-NOESY [101] experiment can also be recorded which similarly shows dipolar correlations, through space, between protons and the carbons bonded to their coupled protons.

2.1.4.1 HMBC experiments

The last commonly used experiment, heteronuclear multiple bond correlation (HMBC) spectroscopy [91], is used to show intra- and inter-residue correlations providing similar information to NOESY, but using coupling through bonds instead of space, giving it some advantages. One of these advantages is that it shows correlations between protons and carbons that are three (and occasionally two) bonds apart but is not reliant exclusively on proton-carbon bonds to transfer magnetisation. This means the coupling can occur through the oxygen of the ring hence intra-residue correlations such as H1 to C3 and C5 can often be seen for α-sugars and sometimes H1 to C5 for β-sugars, while other intra-residue HMBC correlations are the primary way to assign the quaternary carbons or any other carbons that do not have any attached protons. HMBC coupling can also occur across the carbon-oxygen bonds of the glycosidic linkage and these inter-residue correlations allow for confirmation of linkage positions indicated by NOESY or identification of any that aren't and are less ambiguous than NOESY correlations. These correlations manifest as easily identified signals in the anomeric region between the anomeric proton and the linkage carbon of the neighbouring ring and, inversely, the anomeric carbon and the linkage proton for both α- and β-sugars, thus the sequence of the sugar residues in the repeating unit can be readily determined.

2.1.5 Summary

To summarise (Figure 2.1), 1D ^{1}H and ^{13}C (if sufficient sample is available) experiments are performed to identify anomeric and other diagnostic signals. Then, 2D ^{1}H-^{1}H experiments, such as COSY, TOCSY and NOESY identify and characterise the spin systems those diagnostic signals belong to. Once most, if not all, of the proton signals have been assigned, proton/carbon pairs are assigned from the HSQC spectrum with aid from HSQC-TOCSY, HSQC-

NOESY and HMBC, when needed. Once all the spin systems have been assigned, the linkages and sequence of the sugar monomers in the repeating unit are determined by inter-residue signals present in the NOESY and HMBC spectra. Once the structure has been fully elucidated, a fully labelled HSQC spectrum acts as a 2D proton/carbon identity fingerprint since it contains a peak for every proton/carbon pair in the polysaccharide repeating unit.

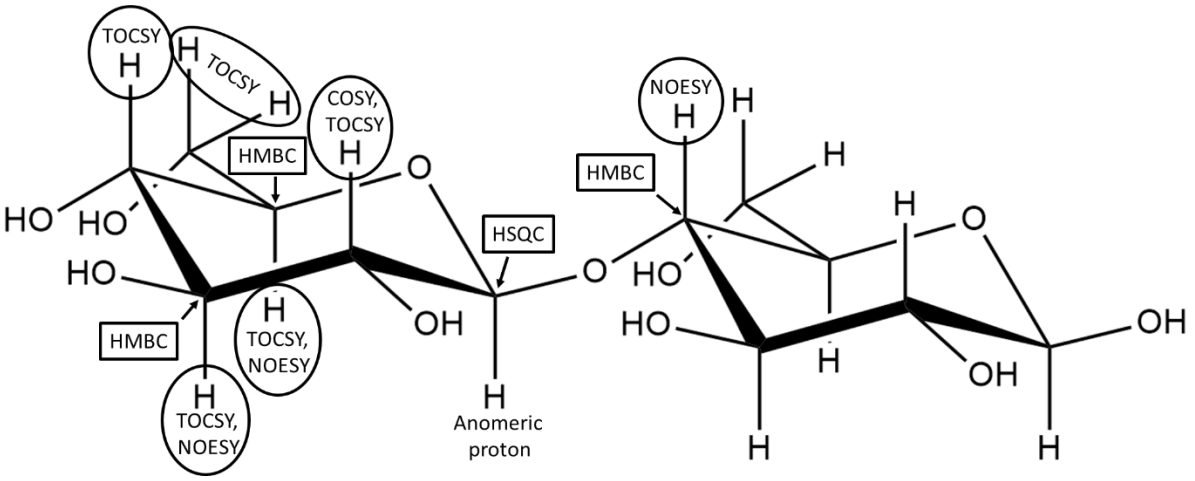


Figure 2.1: Example disaccharide displaying the correlations that would be seen from the anomeric proton using the array of NMR experiments described. Circular highlights represent homonuclear experiments and rectangular ones represent heteronuclear experiments.

2.2. Use of NMR in carbohydrate-based vaccine development

Having now shown how NMR can be used in the structural characterisation of carbohydrates, it is no wonder then that NMR has seen extensive use in the field of carbohydrate-based vaccines, especially for vaccine characterisation and quality control. A variety of NMR methods have been proposed and applied to an equal variety of uses including, but not limited to:

- Identity and determination of polysaccharide antigen structure [80,81,102,103].
- Quantitative measurement of non-carbohydrate substituents potentially related to immunogenicity (e.g. partial O-acetylation) [80,104] or as a measure of successful polysaccharide preparation (e.g. partial N-acetylation).
- A measure of polysaccharide purity by, quantitatively if possible, detecting NMR-active contaminants [105].
- Post-conjugation monitoring of successful conjugation and/or integrity of the conjugated polysaccharide [106].
- Preparation of a reference by which other batches or products can be evaluated.

The possibility of using NMR as an identity, monitoring or release test has also been added to the World Health Organisation recommendations [107] as one of the few methodologies available during the discovery, development and licensure stages of vaccine development.

A number of notable examples of the successful use of NMR in the development, quality control and eventual licensure of carbohydrate-based vaccines are described as follows:

2.2.1. *Haemophilus influenzae* type b (Hib)

Haemophilus influenzae type b (Hib) has been recognised as responsible for meningitis in children between 6 months and 5 years of age since the early 1920's [108], but it was only in the 1970's and 80's when epidemiological studies showed the true extent of the disease burden that study into a Hib polysaccharide vaccine began in earnest. This was spurred on by multiple factors such as, the mortality rate ranging between 5 to 10%, 30% percent of survivors suffering from severe neurologic sequelae [109] and the steady increase of resistance against the antimicrobials, such as ampicillin [110], used to treat Hib meningitis. The first Hib capsular polysaccharide vaccine was licensed in 1985, but it was not effective for children younger than 18 months [111], which are the most vulnerable to Hib disease. The reason for this lack of immunogenicity stems from the fact that the infant immune system has not fully matured and thus does not respond to T-cell independent antigens such as polysaccharides [112]. This problem was solved in the 1990's by the development of conjugate vaccines, namely Hib capsular polysaccharide covalently bound to either diphtheria toxoid, tetanus toxoid or the outer membrane protein complex of a strain of group B meningococcus. This method of increasing the immunogenicity of a ligand, in this case a carbohydrate, by covalently bonding it to a protein was first studied and confirmed by Avery and Goebels in the 1930s [113]. The coupling to a protein carrier transforms the polysaccharide into a T-cell dependent antigen, against which even the infant immune system can mount an immune response.

The ^{1}H NMR spectrum of the Hib polysaccharide, whose repeating unit is [→3)-β-D-Rib*f*-(1→1)-D-Ribitol-(5→***P***→]$_n$, has been established as a "fingerprint", whose complete NMR characterisation and validation as an identity test has been published [102]. It has been shown that NMR can be used to detect degradation of the Hib polysaccharide over time [114], such as depolymerisation or generation of free saccharide, as well as to confirm polysaccharide

integrity through the conjugation process, depending on the type of vaccine and conjugation chemistry [115].

2.2.2. *Neisseria meningitidis* (Nm)

Neisseria meningitidis (Nm), consisting of 12 identified serogroups, was first isolated in 1887 and has been responsible for numerous outbreaks of meningitis in Africa [116] and military populations since the mid-nineteenth century with groups A, B and C accounting for nearly 90% of all cases, with W, Y and X occurring less frequently (Table 2.1). While an antibiotic, sulfa, was developed in the 1930's to treat meningococcal meningitis [117], by the 1960's resistance to this treatment was widespread [118]. This prompted research into a meningococcal vaccine as the best hope to prevent infection. Since there was little, to no, cross-protection between serogroups [119], a meningococcal vaccine would have to be multivalent to provide full coverage. Thus, a tetravalent polysaccharide vaccine was licensed in 1981 that covered groups A, C, Y and W, but it was not optimally immunogenic and was not very effective in children [120]. As with Hib, conjugating the polysaccharide to an antigenic protein to form a glycoconjugate vaccine resulted in an improvement in immunogenicity, and in 1999 a conjugate vaccine for group C was introduced followed by a tetravalent conjugate vaccine, again against groups A, C, Y and W, in 2005.

Table 2.1: Serogroups of *Neisseria meningitidis* and their polysaccharide repeating unit structures.

Serogroup	Repeating unit structure
A	→6)-α-D-Man*p*NAc3/4Ac-(1→***P***→
B	→8)-α-D-Neu*p*NAc-(2→
C	→9)-α-D-Neu*p*NAc7/8Ac-(2→
W	→6)-α-D-Gal*p*-(1→4)-α-D-Neu*p*NAc7/9Ac-(2→
X	→4)-α-D-Glc*p*NAc-(1→***P***→
Y	→6)-α-D-Glc*p*-(1→4)-α-D-Neu*p*NAc7/9Ac-(2→

P is phosphate in a phosphodiester linkage.

Full NMR characterisations for the meningococcal serotypes have been published [80,121], but partial O-acetylation makes fingerprinting difficult due to the number of spin systems resulting in spectral complexity. To solve this, de-O-acetylation is performed in the NMR tube and a second spectrum is recorded. This validated method allows both the confirmation of

the identity of the polysaccharide backbone, as well as the degree of O-acetylation [104] by comparing the relative integrals of acetate and saccharide signals. It has also been shown that NMR can be used to detect degradation of the meningococcal polysaccharide [106], for example hydrolysis of the acid-labile sialic acid residue, by the comparison of spectral profiles from different lots. This means that NMR can be used to confirm polysaccharide integrity, including O-acetylation [80], through the conjugation process. These methods can also be applied to GBS, since it similarly contains sialic acid residues.

2.2.3. *Streptococcus pneumoniae* (Sp)

Streptococcus pneumoniae (Sp), first isolated in 1881, is the primary cause of lobar pneumonia [122], which was the most prevalent and lethal infectious disease in the late nineteenth century and is still globally important today. The pathogen has approximately 100 different identified serotypes, with more still being identified, but the majority of disease is caused by approximately 20 of them. Of the discussed vaccines, pneumococcus has the oldest origin, with the first vaccination attempt against it occurring in South Africa in 1911 [123], using heat-killed bacteria. The first tetravalent trial occurred in the 1940's and in the 1970's hexavalent and 13-valent (Table 2.2) vaccine trials would once again be performed in South Africa [124]. Ultimately, this led to the licensure of a 23-valent polysaccharide vaccine in 1983. As with the previously discussed examples, the relatively poor immunogenicity of the polysaccharide vaccine prompted the development of a glycoconjugate vaccine [125]. So far only 7-, 10- and 13-valent conjugate vaccines have been licensed, but they have already had a noticeable impact on the reduction of invasive pneumococcal disease.

Quality control of the polysaccharide batches used in the preparation of a vaccine with 23 different polysaccharide serotypes poses a unique challenge for NMR since there is an extreme amount of overlap, but focusing on just the anomeric region has allowed polysaccharide identification by the calculation of a correlation coefficient between test and reference spectra [105]. This method was validated by the calculation of a matrix of said correlation coefficients from all possible pair combinations of the 23 serotypes. Contaminants and residuals such as cell wall polysaccharide can also potentially be quantified by the addition of 0.01% DMSO [105]. As opposed to labour- and resource-intensive elemental and colorimetric assays, this NMR method is relatively simple, reproducible and specific [107].

Table 2.2: Polysaccharide repeating unit structures of the *Streptococcus pneumoniae* serotypes present in the 13-valent conjugate vaccine.

Serotype	Repeating unit
1	→3)-D-AAT-α-Gal*p*-(1→4)-α-D-Gal*p*A(2/3Ac)-(1→3)-α-D-Gal*p*A-(1→
3	→3)-β-D-Glc*p*A-(1→4)-β-D-Glc*p*-(1→
4	→3)-β-D-Man*p*NAc-(1→3)-α-L-Fuc*p*NAc-(1→3)-α-D-Gal*p*NAc-(1→4)-α-D-Gal*p*2,3(S)Py-(1→
5	→4)-β-D-Glc*p*-(1→4)-[α-L-Pne*p*NAc-(1→2)-β-D-Glc*p*A-(1→3)]-α-L-Fuc*p*NAc-(1→3)-β-D-Sug*p*-(1→
6A	→2)-α-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→3)-D-Rib-ol-(5→***P***→
6B	→2)-α-D-Gal*p*-(1→3)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→4)-D-Rib-ol-(5→***P***→
7F	→6)-[β-D-Gal*p*-(1→2)]-α-D-Gal*p*-(1→3)-β-L-Rha*p*(2Ac)-(1→4)-β-D-Glc*p*-(1→3)-[α-D-Glc*p*NAc-(1→2)-α-L-Rha*p*-(1→4)]-β-D-Gal*p*NAc-(1→
9V	→4)-α-D-Glc*p*(2/3Ac)-(1→4)-α-D-Glc*p*A(2/3Ac)-(1→3)-α-D-Gal*p*-(1→3)-β-D-Man*p*NAc(4/6Ac)-(1→4)-β-D-Glc*p*-(1→
14	→4)-β-D-Glc*p*-(1→6)-[β-D-Gal*p*-(1→4)]-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→
18C	→4)-β-D-Glc*p*-(1→4)-[α-D-Glc*p*(6Ac)-(1→2)][Gro-(1→***P***→3)]-β-D-Gal*p*-(1→4)-α-D-Glc*p*-(1→3)-β-L-Rha*p*-(1→
19A	→4)-β-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→3)-α-L-Rha*p*-(1→***P***→
19F	→4)-β-D-Man*p*NAc-(1→4)-α-D-Glc*p*-(1→2)-α-L-Rha*p*-(1→***P***→
23F	→4)-β-D-Glc*p*-(1→4)-[α-L-Rha*p*-(1→2)]-[Gro-(2→***P***→3)]-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→

AAT is 2-acetamido-4-amino-2,4,6-trideoxygalactose, Gro is glycerol, Pne is 2-acetamido-2,6-dideoxytalose, Sug is 2-acetamido-2,6-deoxyhexose-4-ulose, Py is pyruvate and ***P*** is phosphate in a phosphodiester linkage.

2.2.4. Group B *Streptococcus* (GBS)

Group B *Streptococcus* is a common cause of neonatal sepsis and meningitis, also sometimes affecting adults and the elderly. As mentioned in Chapter 1, in the 1980s clinical trials demonstrated that intrapartum antibiotic prophylaxis (IAP) was effective at preventing early-onset disease (<7 days after birth) [126], and was successfully implemented in industrialized regions such as Europe and the United States in the 1990s. However, challenges with IAP include the need for universal screening, the possibility of vaginal colonisation after screening, ensuring that IAP occurs timeously (at least 4 hours before delivery), the fact that it does not

effectively protect against late-onset disease as well as the possibility of developing antimicrobial resistance among GBS. An alternative that could address some of these challenges is the development of a glycoconjugate vaccine, as shown for the prevention of bacterial meningitis and pneumonia. A number of studies have shown the immunogenicity of mono- and bi-valent vaccines in pregnant women [127,128] and since the early 2010's, a number of phase II clinical trials have been completed on trivalent (types Ia, Ib and III) conjugate vaccines, with phase III trials under consideration [129].

NMR has played a role in the structural determination of the polysaccharide RU of all ten identified GBS serotypes Ia, Ib, II-IX [130–138] (Table 1.1), ranging from a supplemental role in the earlier determinations to the primary method of elucidation for some of the more recently identified serotypes. NMR fingerprinting has been investigated as a relatively fast and accurate method for tracking and confirming polysaccharide identity of GBS serotypes through the vaccine manufacturing process in both monovalent and multivalent vaccines by the selection and comparison of diagnostic signals in the anomeric region [83].

Table 2.3: Group B *Streptococcus* serotypes covered in this study and their capsular polysaccharide repeating unit structures.

Serotype	Repeating unit
Ia	→4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→
Ib	→4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→
II	→2)-[α-D-Neu*p*NAc-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-[β-D-Gal*p*-(1→6)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→3)-β-D-Glc*p*-(1→
III	→6)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)]-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→
IV	→4)-α-D-Glc*p*-(1→4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→
V	→4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]-α-D-Glc*p*-(1→4)-[β-D-Glc*p*-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

NMR also played a significant role in the identification of one the common contaminants of Group B capsular polysaccharides, Group B Carbohydrate (GBC), which contains numerous

rhamnose residues [139]. These 6-deoxy rhamnose sugars exhibit a diagnostic H6 doublet that can serve as an indicator of GBC contamination.

The NMR characterisation work that has previously been published on the six GBS serotypes covered in this study (Ia, Ib, II, III, IV and V in Table 2.3 above, cumulatively accounting for more than 95% of global GBS disease cases [7]) is limited. Some NMR data of GBS Ia fragments, monomers and dimers [140–142] as well as ^{1}H spectra, anomeric assignments and HSQC spectra recorded at 298 K for both the Ia and Ib serotypes have been published [83,143]. For GBS II, NMR data has been published for a monomer of the CPS [144] as well as the ^{1}H spectra and anomeric assignments at 298 K of the polysaccharide [83], but NMR data for the full polysaccharide has yet to published. The most important serotype, GBS III, understandably has the most published data. High-resolution ^{1}H and ^{13}C NMR data of the RU has been published in a number of studies [83,145–150], and a full proton and carbon assignment of data recorded at 290 K has also been published [84]. Recently, selective 1D TOCSY and HSQC-TOCSY data has been published showing the proton and carbon assignment of most of the Glc, GlcNAc and Gal spin systems at 343 K [87]. For GBS IV CPS, data is available on the anomeric and backbone residue assignments at 310 K [133], as well as ^{1}H spectra and anomeric assignments at 298 K [83]. Lastly, for GBS V, NMR data for a monomer of the CPS has been published [151], as well as data on the anomeric and backbone residue signals at 300 K [134] and ^{1}H spectra and anomeric assignments at 298 K [83].

Thus, while GBS III has received full NMR characterisation at 298 K, no GBS serotypes have received full NMR characterisation at 343 K. The aim of this study is to perform said full characterisations of these six GBS serotypes in an effort to aid in the ongoing development of glycoconjugate vaccines against GBS [83,129], since they are the six serotypes that are the target of a potential hexavalent conjugate vaccine.. These assignments will not only serve as an identity and purity test, but also have the possibility of serving as a means to detect degradation of the polysaccharide and a confirmation of polysaccharide antigen structural integrity after the activation and conjugation processes.

2.3. Materials and methods for NMR structural characterisation

The NMR analytical samples ranging between 5-20 mg of GBS type Ia, Ib, II, III, IV and V polysaccharide as well as Ia-TT, II-TT, III-TT and V-TT conjugate samples were acquired from PATH and vaccine partners supported by a grant from the Bill and Melinda Gates Foundation.

The samples were prepared by dissolving 2-5 mg of lyophilised polysaccharide in 0.7 ml of D_2O (99.9 atom % deuterium acquired from Sigma-Aldrich, #613444). Samples were then subjected to a round of D_2O exchange, mixed, allowed to stand, then transferred to a clean 5 mm NMR tube by lyophilisation/freeze drying and dissolving in a further 0.7 ml D_2O. Some conjugate samples were sonicated prior to transfer to the NMR tube due to poor solubility.

All NMR spectra were recorded at 343 K unless stated otherwise. Spectra were recorded on a Bruker Avance III 600 MHz spectrometer equipped with a BBO Prodigy cryoprobe. The spectra were recorded using standard Bruker pulse programs (Table 2.4) and processed using Bruker software (Topspin 3.6). 1D ^{1}H NMR spectra were recorded with 65,536 data points over a 9-10 ppm sweep width using a 30-degree pulse and a relaxation delay (D1) of 2s, accumulating 64 scans. The 1D ^{1}H pre-saturation experiments were recorded under the same conditions, but with saturation at the water frequency. The 1D DOSY experiments were also run under the same conditions as the 1D ^{1}H experiments, but the number of scans was increased to 128. ^{1}H spectra were weighted with 0.3 Hz line broadening and Fourier-transformed and were referenced relative to the axial H3 signal of the terminal sialic acid residue at 1.79 ppm [89]. ^{13}C spectra were recorded at 150.9 MHz with 65,536 data points over a 230 ppm sweep width, using a 30-degree pulse and a relaxation delay (D1) of 1s, accumulating between 5,000 and 15,000 scans. ^{13}C spectra were referenced relative to the C3 of the terminal sialic acid residue at 40.68 ppm [89].

Two-dimensional ^{1}H-^{1}H homonuclear experiments (COSY pre-sat, TOCSY pre-sat, DOSY-TOCSY and NOESY pre-sat) were also acquired with standard Bruker pulse programs (Table 2.4). Typically, 4,096 and 256 or 512 data points were collected in the F_2 and F_1 dimensions, respectively. On average 80 scans were accumulated with Non-Uniform Sampling (NUS) at 50%. Mixing times were 180 ms and 300 ms for 2D TOCSY and NOESY, respectively. Selective 1D TOCSY and NOESY spectra were collected with 65,536 data points over a 10 ppm sweep width, accumulating 320 scans over a range of mixing times (25-200 ms and 100-500 ms for TOCSY and NOESY respectively). The spectra were weighted with 1.0 Hz line broadening and Fourier-transformed.

Two-dimensional ^{1}H-^{13}C heteronuclear experiments (HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC) were acquired under the same conditions as the 2D homonuclear experiments, but with NUS at 25%. Mixing times were 120 ms and 300 ms for HSQC-TOCSY and HSQC-

NOESY respectively and coupling constants were 145 Hz and 6-8 Hz for HSQC-DEPT and HMBC respectively.

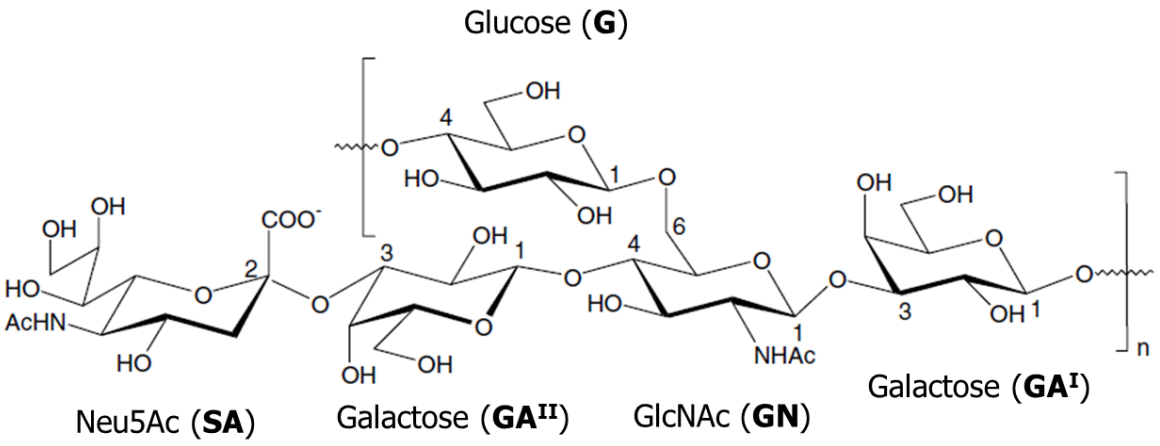
Table 2.4: Bruker NMR pulse programs used during structural characterisation of the GBS serotypes.

Experiment code	Description	Coupling constants/Mixing times
	1D	
zg30	^{1}H	
zgpr	^{1}H with pre-saturation	
ledbpgp2s1d	DOSY	
selmlgp	selective TOCSY	25-200 ms
selnogp	selective NOESY	100-500 ms
zgpg30	^{13}C	
deptsp135	^{13}C-DEPT 135	
	2D	
cosygpprqf	COSY	
mlevphpr.2	TOCSY pre-sat	120-180 ms
ledbpgpml2s2d	DOSY-TOCSY	120-180 ms
noesyphpr	NOESY	300 ms
hsqcedetgpsisp2.3	HSQC-DEPT	145 Hz
hsqcdietgpsisp	HSQC-TOCSY	120 ms
hsqcetgpnosp	HSQC-NOESY	300 ms
hmbcgplpndqf	HMBC	6 or 8 Hz

One online resource, CASPER [152], was used. CASPER is a predictive tool that uses a large database of experimental carbon and proton chemical shifts to calculate the potential chemical shifts of a wide variety of possible oligo- and polysaccharide fragments to displayed margins of error. CASPER was used as a tool to predict the chemical shifts of fragments of the GBS repeating units and facilitated the spectral assignments based on correlations seen in the 2D NMR spectra. Additionally, it provided the unsubstituted monosaccharide data from which glycosylation shifts were calculated.

Chapter 3. NMR characterisation of GBS serotype III

The type III serotype of Group B *Streptococcus* is responsible for the greatest number of GBS disease cases, accounting for approximately 60% of worldwide cases [7]. Since it is arguably the most important GBS serotype for vaccine research (and to prevent undue repetition) an in-depth analysis of the full assignment of the GBS III CPS will be presented, while the analysis of the other five serotypes will be abridged somewhat. The capsular polysaccharide repeating unit (RU) structure of GBS III was published in 1980 [132], determined using chemical methods supported by ^{13}C NMR data, and was the first complete GBS capsular polysaccharide structure to be published. The chemical analysis demonstrated that the GBS III CPS has a pentasaccharide repeating unit containing D-Galactose (Gal), D-Glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc) and N-Acetylneuraminic acid (Neu5Ac or sialic acid) in a molar ratio of 2:1:1:1. The position of the sialic acid-Gal linkage had not been determined correctly, however, so a revision was published in 1987 wherein the structural analysis was supported by high-resolution ^{1}H and ^{13}C NMR data [145] to give the RU structure shown in Figure 3.1.



→6)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)]-β-D-Glc*p*NAc-(1→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Figure 3.1: Repeating unit structure and line structure of the GBS III capsular polysaccharide.

That study included assignments for the anomeric protons of the polysaccharide, a partial assignment of the polysaccharide ^{13}C spectrum as well as a full carbon assignment of the depolymerised pentasaccharide acquired by enzymatic digestion. In the following years a number of attempts were made at the chemo-enzymatic synthesis of ever larger fragments of the capsular polysaccharide [146–150], each of which were supported by high-resolution 1D and 2D NMR, often allowing for the full assignment of the synthesised fragments. The ^{1}H

spectrum and anomeric assignments have been published [83] and an NMR and molecular dynamics study of GBS III that includes full proton and carbon assignments of various fragments as well as the polysaccharide with the spectra recorded at 290 K has also been published [84]. More recently, both selective 1D TOCSY and HSQC-TOCSY data recorded at 343 K has been published showing the proton and carbon assignment of most of the Glc, GlcNAc and Gal spin systems [87]. However, full proton and carbon assignments of the complete CPS RU at 343 K have so far not been published.

3.1. Structural characterisation of the GBS III capsular polysaccharide

Most literature polysaccharide fingerprinting and/or structural elucidation methods generally use temperatures in a range between 298 and 323 K [83,102-105,138,153], but for this study the structural elucidation was performed using spectra recorded at 343 K. While some studies have performed experiments at this temperature [83,136], the practice isn't common, but it provides a number of benefits.

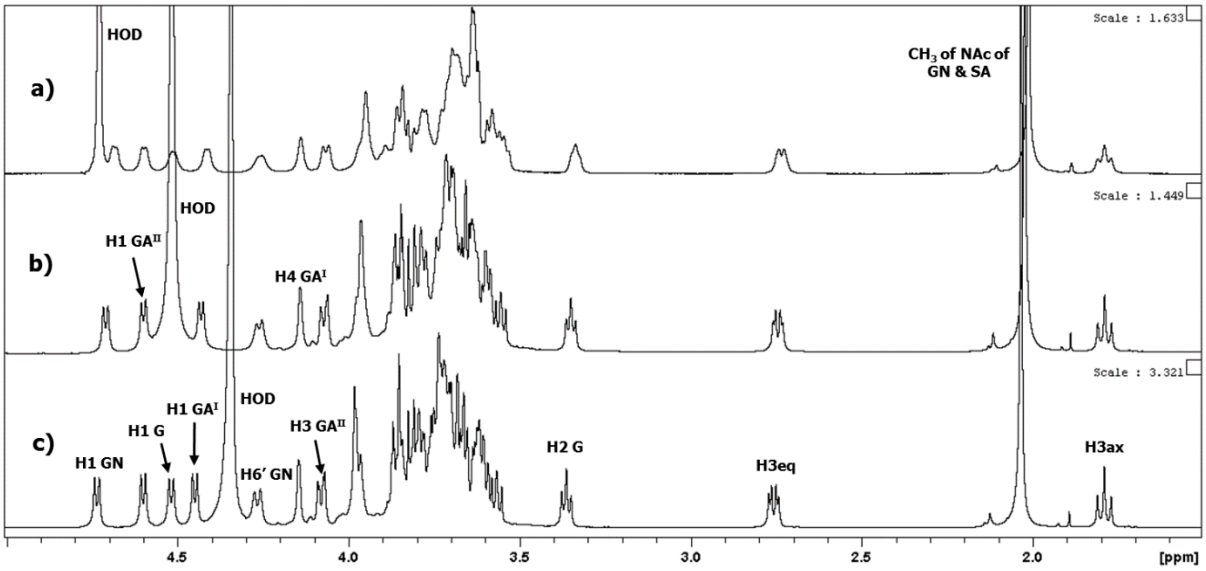


Figure 3.2: Overlay of 1D ^{1}H spectra of the GBS III CPS recorded at 298 (a), 323 (b) and 343 K (c) with diagnostic peaks labelled.

As can be seen from Figure 3.2, recording the spectra at a higher temperature results in sharper peaks. This is because the increased temperature results in increased tumbling of the polysaccharide in solution, which in turn results in a longer relaxation time. Since relaxation time is inversely proportional to the width of a signal, this results in thinner, sharper resonances. The increased temperature also causes a shift of the temperature-sensitive HOD peak away from the anomeric region where it overlaps with and obscures diagnostic anomeric

peaks. This is important not only to get the correct count and chemical shifts of the anomeric peaks but also to facilitate 1D and 2D TOCSY and NOESY correlations from each anomeric peak, which often cannot be seen if the peak is overlapped with the HOD signal. Since, at neutral pH, the polysaccharides do not degrade at this higher temperature, it represents a significant improvement over previously published fingerprinting methods.

3.1.1. 1D ^{1}H and DOSY experiments

Beginning with the 1D ^{1}H spectrum (Figure 3.3a), several well-separated diagnostic peaks could be assigned from the literature [83,154]. This was aided by use of a DOSY spectrum (Figure 3.3b) to confirm that there were no peaks overlapped with and underneath the water signal. Four anomeric signals are present between 4.4-4.8 ppm, suggesting the presence of four sugar residues, but the sialic acid residue is not a standard hexose. It has an exocyclic carbonyl group (labelled C1) attached to the anomeric carbon (thus labelled C2), meaning it does not have an anomeric proton. Additionally, C3 is deoxygenated, meaning it has two attached protons, one axial and one equatorial.

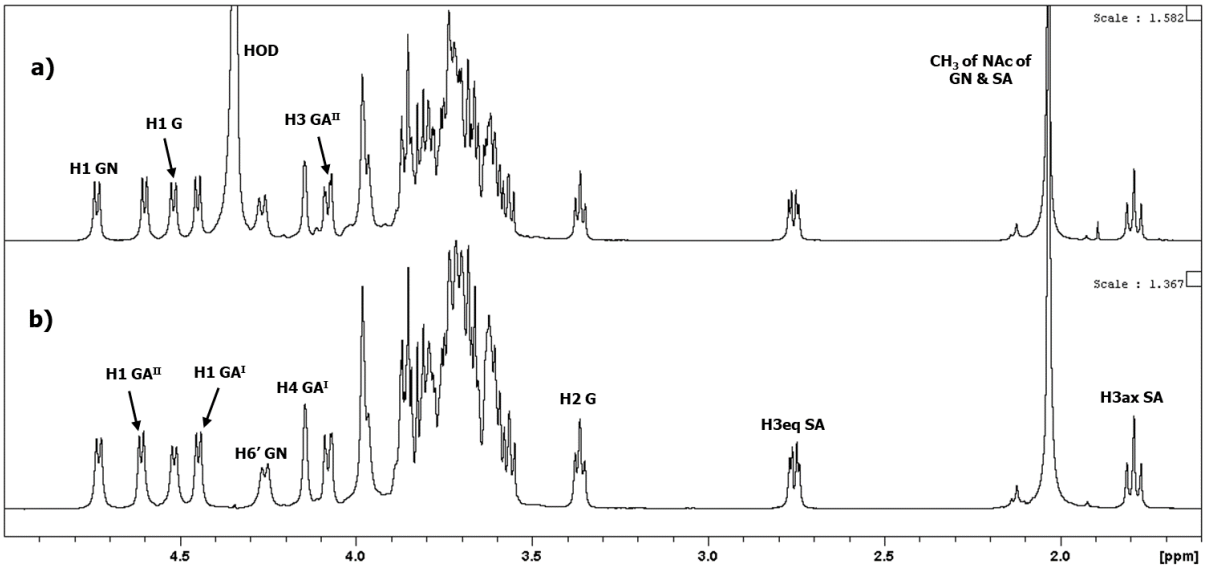


Figure 3.3: Overlay of 1D ^{1}H (a) and DOSY (b) spectra of the GBS III CPS recorded at 343 K with diagnostic peaks labelled.

Therefore, these characteristic H3 axial/equatorial sialic acid (SA) signals at 1.79 and 2.76 ppm in the ^{1}H and the presence of five anomeric signals in the ^{13}C spectrum (Figure 3.15) show a pentasaccharide repeating unit as published. Since the terminal sialic acid is a common subunit between all GBS capsular polysaccharides and remains relatively unaffected by the structural differences between them, the distinctive H3ax peak of SA was set to 1.79 ppm (a

value also given by the CASPER NMR chemical shift prediction tool [152]) and used as the calibration peak for the ^{1}H spectra of all serotypes presented in this study.

The four anomeric peaks (4.74, 4.60, 4.52 and 4.45 ppm) were assigned to H1 of GN, GA^{II}, G and GA^{I} respectively and the fact that all four resonate below 4.9 ppm is consistent with the evaluation that they are β-linked hexoses. Additional peaks outside of the ring region could be assigned by inspection, namely H6' GN, H4 GA^{I}, H3 GA^{II}, H2 G, H3eq/ax SA and the methyl signals from the N-acetyl groups of GN and SA (4.27, 4.14, 4.08, 3.36, 2.76, 1.79 and 2.03 ppm respectively). The assignments made by 1D H^{1} and DOSY experiments are summarised in Figure 3.4. The rest of the ring proton signals are too heavily overlapped, so the analysis turned to correlation experiments to deconvolute and identify each spin system.

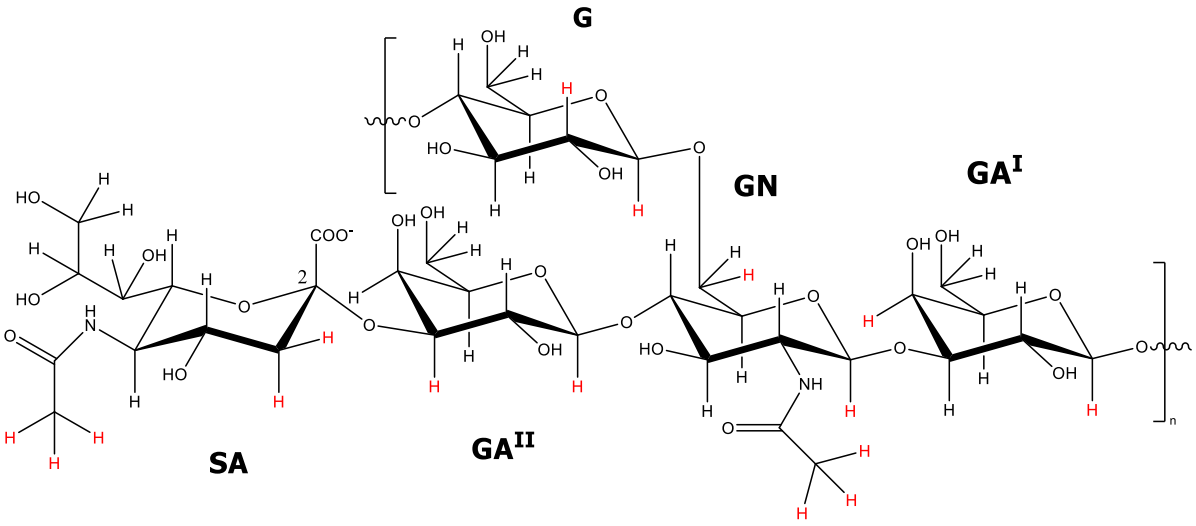


Figure 3.4: GBS III repeating unit structure with the protons assigned by 1D H^{1} and DOSY experiments highlighted in red.

3.1.2. Homonuclear correlation experiments (COSY, TOCSY and NOESY)

Using 2D correlation experiments, such as COSY and TOCSY (Figure 3.5 overlay), with 1D TOCSY providing insight into peak multiplicities (Figure 3.6), most of the spin systems could be preliminarily assigned.

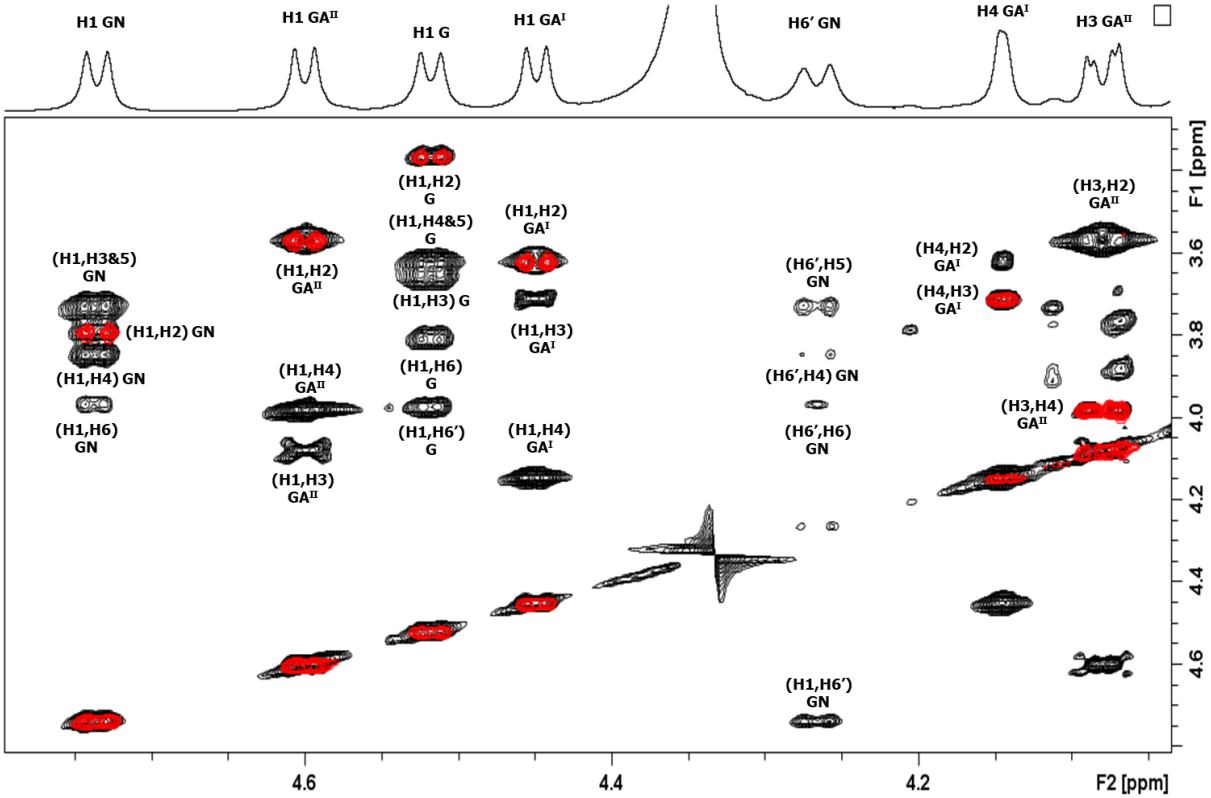


Figure 3.5: Anomeric region of the 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of GBS III with identified peaks for G, GN, GA^{I} and GA^{II} labelled.

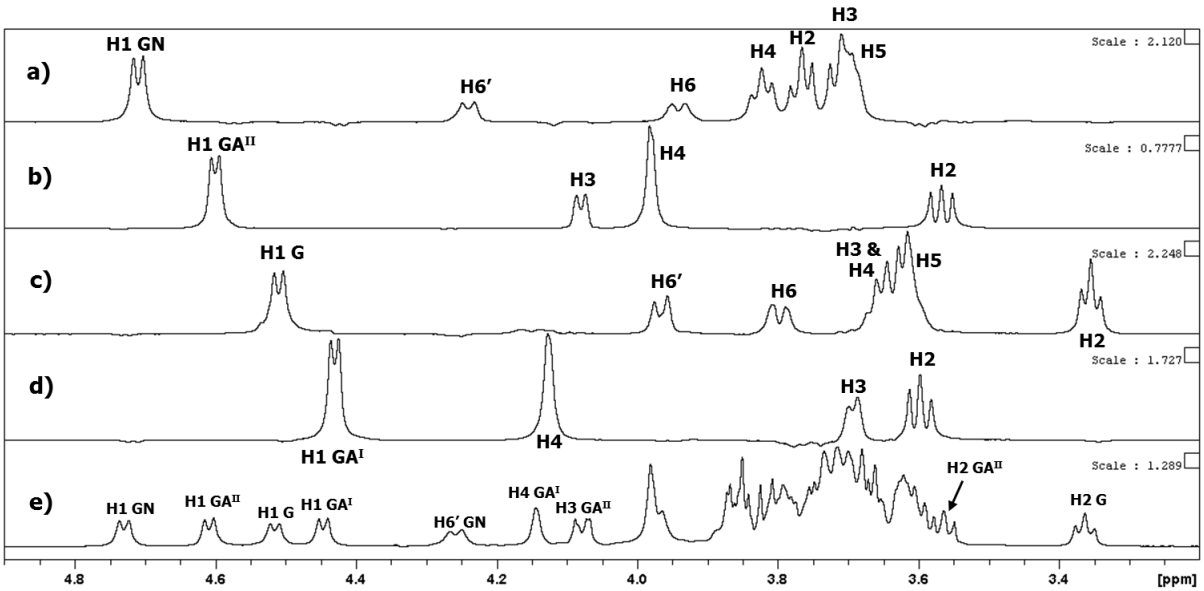


Figure 3.6: Labelled anomeric and ring regions of 1D TOCSY experiments wherein the anomeric protons of GN, GA^{II}, G and GA^{I} (a-d respectively) were selectively irradiated overlaid onto the DOSY spectrum (e) of GBS III.

The most downfield anomeric proton (H1 GN, 4.74 ppm) shows a correlation in the COSY to a signal at 3.79 ppm, assigning it as H2 GN, and this correlation continues in the TOCSY to a large signal at approximately 3.7 ppm, assigned as overlapped H3 GN (3.74 ppm) and H5 GN

(3.72 ppm), a signal at 3.85 ppm, assigned as H4 GN, and a pair of signals at 3.96 and 4.27 ppm assigned as H6/6' GN. Thus, the preliminary assignment of the GlcNAc residue is complete. The anomeric proton of the Glc residue (H1 G, 4.52 ppm) shows a COSY correlation to the previously assigned H2 G signal at 3.36 ppm which, in turn, shows a correlation to a signal at 3.66 ppm (Figure 3.7), assigned as H3 G. This correlation is continued in the TOCSY to an overlapped signal assigned as H4 G (3.64 ppm) and H5 G (3.62 ppm) and a pair of signals at 3.80 and 3.97 ppm assigned as H6/6' G, thus completing the preliminary assignment of the entire glucose spin system. The most deshielded Gal anomeric signal (H1 GA^{II}, 4.60 ppm) shows a correlation in the COSY with a signal at 3.57 ppm, assigned as H2 GA^{II}, which can be continued (Figure 3.7) to the signal at 4.08 ppm previously assigned as H3 GA^{II} which in turn continues to a signal at 3.98 ppm assigned as H4 GA^{II}. This is where the correlations cease, as is characteristic of Gal, and this is due to the small coupling constant, $J_{H4,H5}$, between the equatorial H4 of Gal and its axial neighbour, H5. The final anomeric signal (H1 GA^{I}, 4.45 ppm) similarly correlates to a signal at 3.62 ppm in the COSY, assigned as H2 GA^{I}, and two additional signals in the TOCSY at 3.71 and 4.14 ppm, assigned as H3 and H4 GA^{I} respectively.

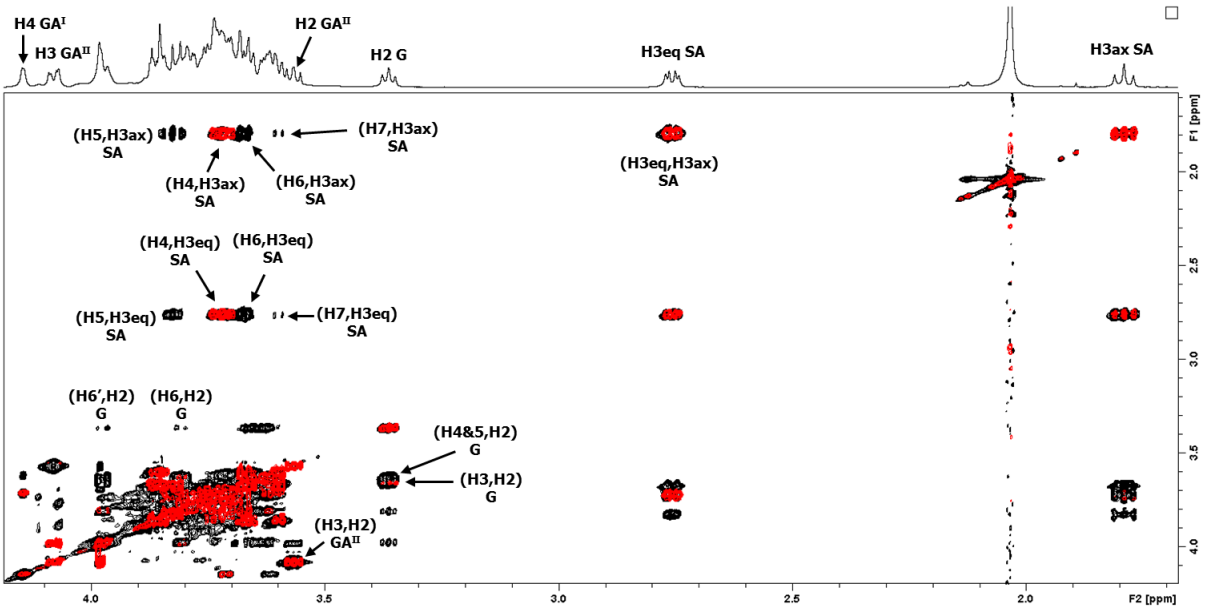


Figure 3.7: Ring and upfield regions of the 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of the GBS III CPS with identified peaks for G and SA labelled.

The last spin system to be assigned was that of sialic acid, which begins at the H3ax/eq signals at 1.79 and 2.76 ppm and shows COSY correlations (Figure 3.7) to a signal at 3.72 ppm, assigned as H4 SA. The TOCSY correlations continue to a signal at 3.82 ppm, assigned as H5 SA, a signal at 3.67 ppm, assigned as H6 SA and lastly a small signal at 3.60 ppm, assigned as

H7 SA. The low intensity of the correlation from H7 to H3 SA and the fact that this is where the TOCSY correlations cease is due to the small $J_{H6,H7}$ coupling constant limiting the transfer of magnetisation. This left H5 and H6 of GA^{I} and GA^{II} as well as H8 and H9 of SA unassigned.

Both 1D and 2D NOESY (Figures 3.8 and 3.9) experiments aided in identifying as yet unassigned peaks through intra-residue correlations as well as confirming the published connectivities between sugar residues in the repeating unit through inter-residue correlations.

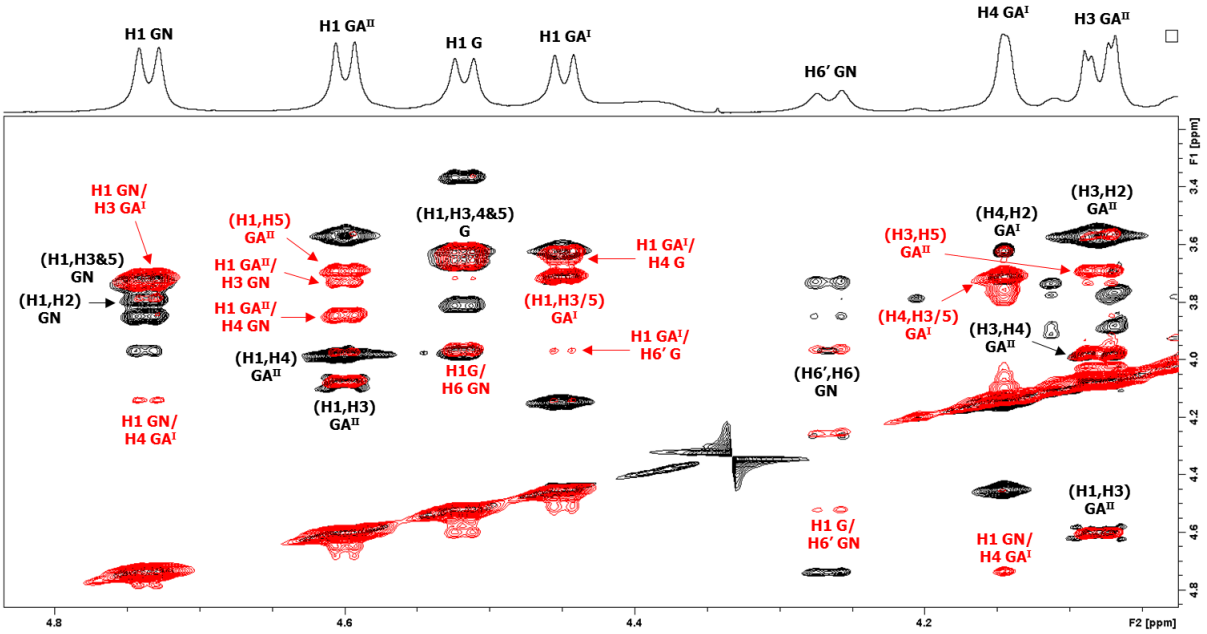


Figure 3.8: Anomeric region of the 2D NOESY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of the GBS III CPS with identified intra- and inter-residue NOESY correlations labelled. Correlations not already seen in through TOCSY are labelled in red.

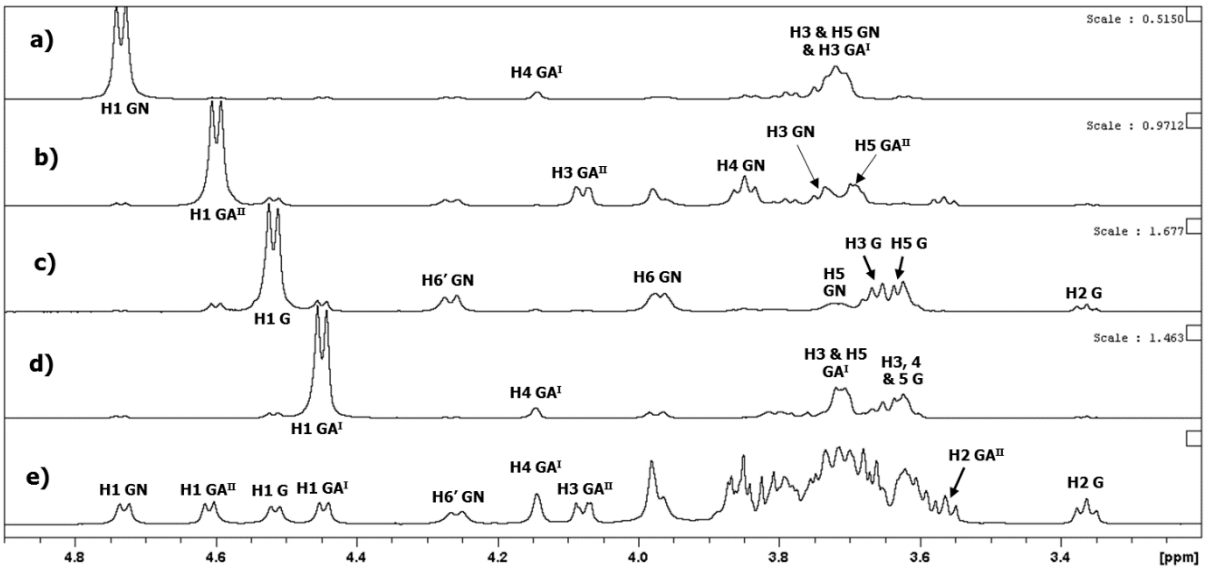


Figure 3.9: Labelled anomeric and ring regions of 1D NOESY experiments wherein the anomeric protons of GN, GA^{II}, G and GA^{I} (a-d respectively) were selectively irradiated overlaid onto the DOSY spectrum (e) of GBS III.

Beginning with the most deshielded anomeric proton, H1 GN (Figure 3.9a), intra-residue NOESY correlations can be seen to the signals for H3 GN and H5 GN (3.72 and 3.74 ppm respectively). These are unfortunately overlapped with the inter-residue correlation to the linkage proton, H3 GA^{I} (3.71 ppm), meaning the GN-(1→3)-GA^{I} linkage cannot be directly confirmed by NOESY, but a weak signal to H4 GA^{I} does provide evidence for the linkage. The GA^{II} anomeric signal (Figure 3.9b) shows a NOESY correlation to a signal at 3.69 ppm that was assigned as H5 GA^{II}, as well as an inter-residue correlation to H4 GN (3.85 ppm) and a weak correlation to H3 GN (3.74 ppm) confirming the GA^{II}-(1→4)-GN linkage. The anomeric signal of Glc (Figure 3.9c) shows intra-residue correlations to the already assigned H3 and H5 G, but also inter-residue correlations to H6/6’ (3.96/4.27 ppm) and H5 GN (3.72 ppm), confirming the G-(1→6)-GN linkage. The NOESY correlation from H1 GA^{I} to a signal at 3.71 ppm likely arises from both H3 GA^{I} and H5 GA^{I} but this cannot be entirely confirmed in the NOESY. Additional inter-residue correlations from H1 GA^{I} (Figure 3.9d) to peaks around 3.64 ppm appear to be overlapped correlations to H3, H4 and H5 G (3.66, 3.64, and 3.62 ppm respectively), confirming the GA^{I}-(1→4)-G linkage.

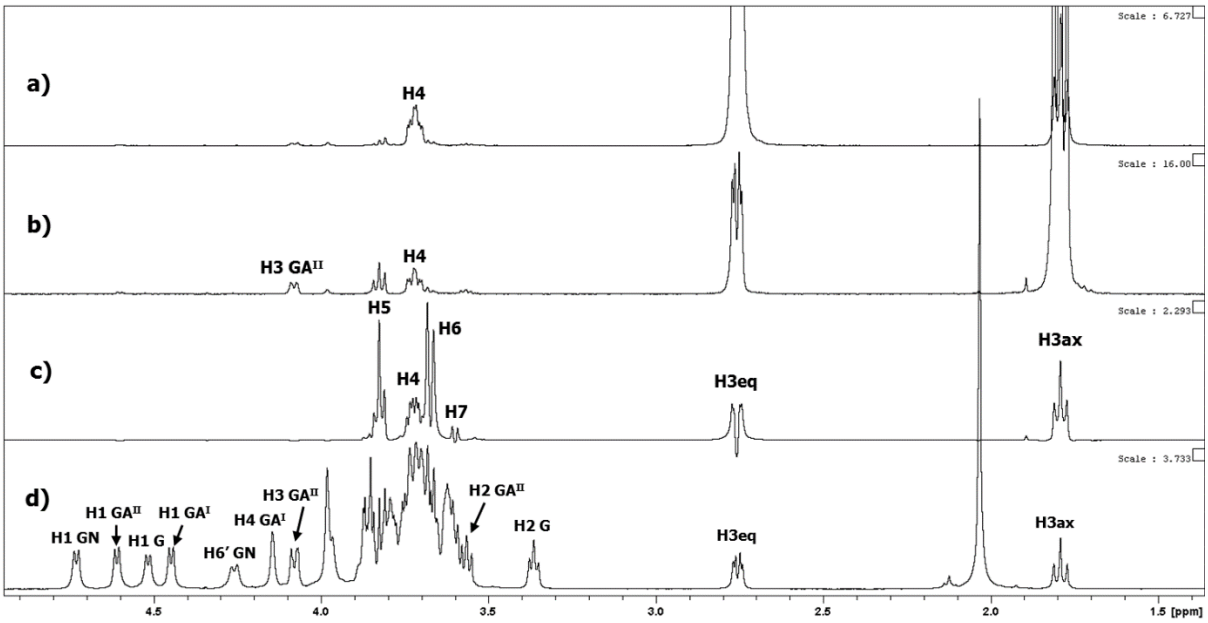


Figure 3.10: Labelled 1D NOESY (a,b) and TOCSY (c) experiments wherein the H3ax (b,c) or H3eq (a) protons of SA were selectively irradiated overlaid onto the DOSY spectrum (d) of GBS III.

Finally, besides showing intra-residue correlations to the already assigned H4 and H5 SA, the H3ax SA proton (Figure 3.10) shows a NOESY correlation to H3 GA^{II}, confirming the SA-(2→3)-GA^{II} linkage. Thus, the NOESY led to the preliminary assignment of H5 of GA^{I} and GA^{II}, as well as evidence or confirmation of all five linkages present in the GBS III CPS repeating unit, leaving only H6 of GA^{I} and GA^{II}, as well as H8 and H9 of SA unassigned. The summary of which protons were assigned by COSY, TOCSY and NOESY is presented in Figure 3.11.

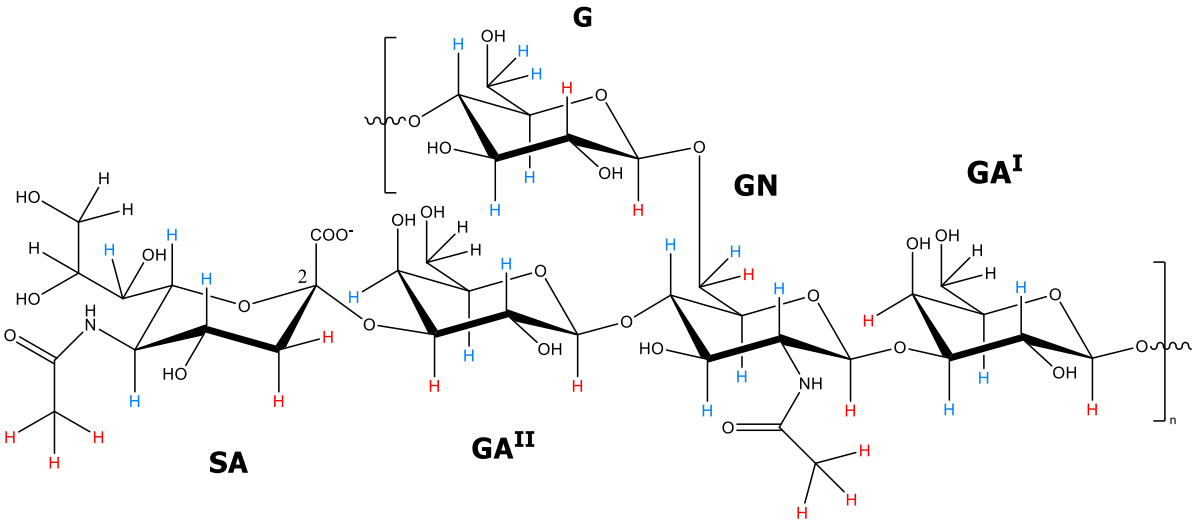


Figure 3.11: GBS III repeating unit structure highlighting the protons assigned by 1D H^1/DOSY (red) and COSY/TOCSY/NOESY (blue) experiments.

3.1.3. Heteronuclear experiments (HSQC, HMBC and hybrid)

Having preliminarily assigned nearly every proton signal, the full ^{13}C NMR assignment and confirmation of all proton assignments began through the combined use of HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC experiments. As with the ^{1}H spectra, the distinctive ^{13}C signal for C3 SA was set to 40.68 ppm (again, a value used by the CASPER NMR chemical shift prediction tool [152]) and used as the calibration peak for the ^{13}C spectra of all the serotypes presented in this study.

The HSQC and HSQC-hybrid experiments contain a wealth of information and allow relatively easy assignment and/or confirmation of nearly all ^{1}H and ^{13}C signals by the use of 1D TOCSY spectra overlaid onto the HSQC-TOCSY and HSQC-DEPT spectra (Figure 3.12 shows the Glc spin system as an example). In this overlay, the carbon chemical shifts are represented vertically above the anomeric proton signal by the HSQC-TOCSY and the proton chemical shifts horizontally by the selective 1D TOCSY, allowing for easier cross-referencing and assignment. Taking one of these overlays (Figure 3.12) as an example, HSQC signals accompanied by 1D and 2D TOCSY correlations show the assignment and confirmation of every proton and carbon in the Glc spin system.

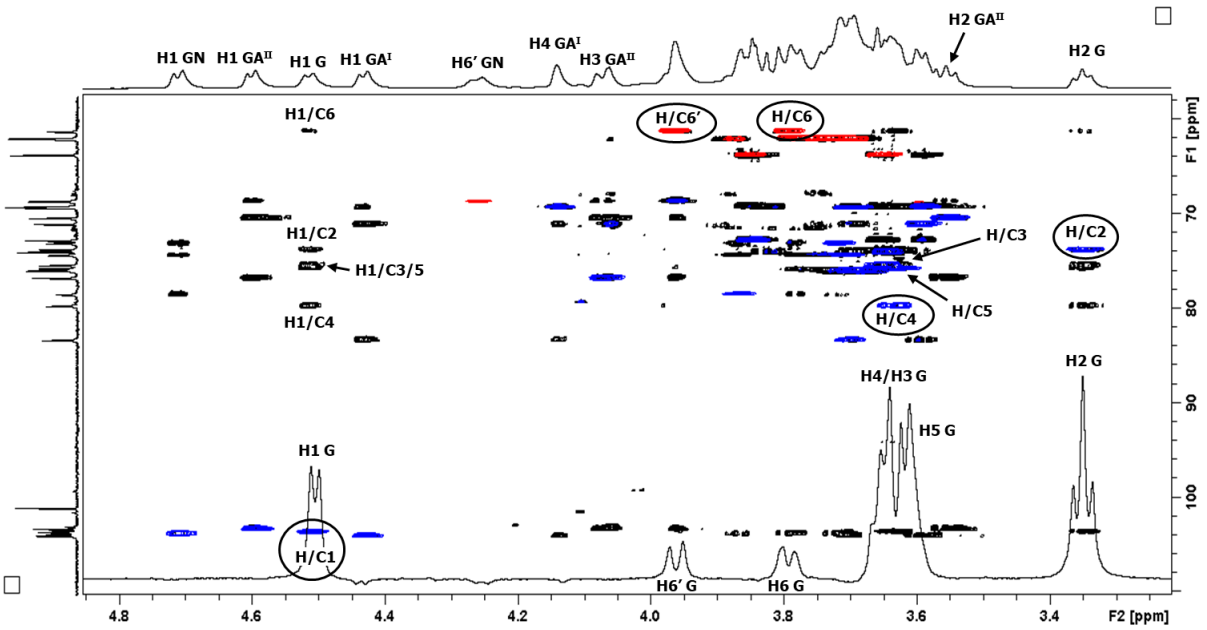


Figure 3.12: Anomeric and ring regions of the selective 1D TOCSY of H1 G (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS III CPS, highlighting correlations that show the proton and carbon assignments of the entire glucose spin system.

Arrayed vertically above the H/C1 HSQC signal, correlations can be seen to all the carbons of the Glc spin system. When cross-referenced against the selective 1D TOCSY spectrum overlaid onto the bottom of the figure, the HSQC signal for each proton/carbon pair in the residue can be identified and assigned. If the sample is particularly complex or there is a significant amount of ambiguity, a similar overlay using HSQC-DEPT, HSQC-NOESY and 1D NOESY spectra can be used to hopefully provide additional assignments and correlations, for example to identify H5 of Gal residues, but this wasn't necessary for the GBS III polysaccharide.

The HSQC-TOCSY (Figure 3.13) also allows assignment of the final unassigned protons, H8 and H9/9' of SA, by their correlations with H7 SA, assigning them as 3.85 and 3.67/3.86 ppm respectively, and H6 of both GA^{I} and GA^{II} by a process of elimination since they, as the only unassigned protons, must be the final large unassigned peak at 3.75 ppm. They were given the same assignment since the data does not give a means to tell them apart in the broad, overlapped signal.

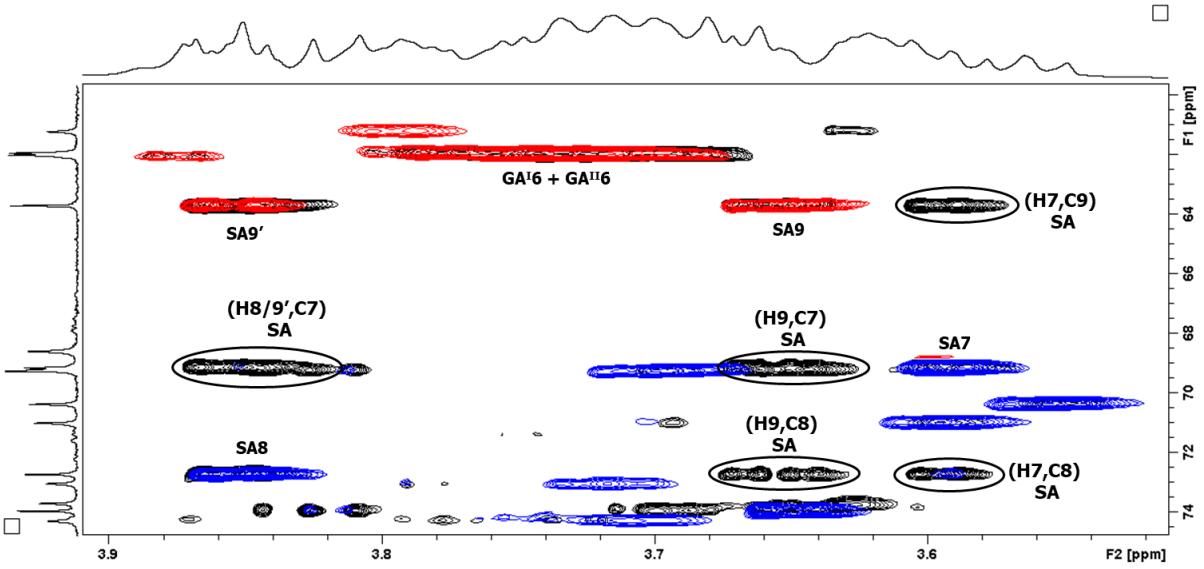


Figure 3.13: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS III CPS, highlighting correlations that led to the assignment of H8 and H9 of SA.

The HMBC (Figure 3.14) spectrum allows the full confirmation of the repeating unit connectivity shown by the NOESY (Figure 3.9) as well as the assignment of the quaternary carbons (carbons with no attached protons) such as C1 and C2 of SA as well as the carbonyl carbons in the N-acetyl groups of SA and GN, by showing intra- and inter-residue correlations through 3-bond coupling.

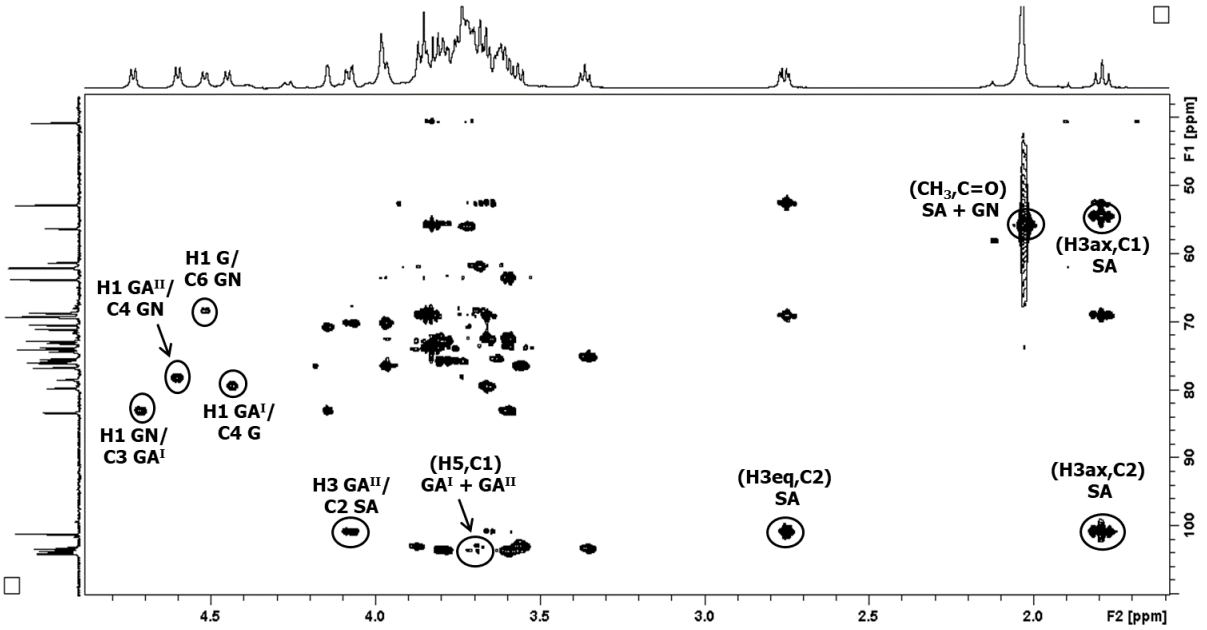


Figure 3.14: HMBC spectrum of the GBS III CPS with important intra- and inter-residue signals labelled.

Four of the five linkages in the polysaccharide, GN-(1→3)-GA^{I}, GA^{II}-(1→4)-GN, G-(1→6)-GN and GA^{I}-(1→4)-G are confirmed in Figure 3.14 by correlations between the anomeric protons and linkage site carbons. Sometimes linkages can also be confirmed by the reverse correlation, namely from the linkage ring protons to the anomeric carbons, like for the SA-(2→3)-GA^{II} linkage (since SA does not have any anomeric protons), though those can often be obscured by larger C1-H2 signals. HMBC can also confirm assignments that are difficult to confirm with other experiments, such as H5 of the Gal residues, which are confirmed by correlations with the anomeric Gal carbons (Figure 3.14). Lastly, HMBC is one of the only experiments that can assign the quaternary carbons in the polysaccharide. The carbon assignment of C2 SA (101.02 ppm) is relatively simple, since it is an anomeric carbon, thus showing intra-residue correlations to H3eq/ax, as well as an inter-residue correlation to H3 GA^{II}, which was also used to confirm the SA-(2→3)-GA^{II} linkage. The carbonyl carbons of the N-acetyl groups in both sialic acid and GlcNAc can be identified through a large correlation to the CH_3 signal at approximately 55 ppm. Since carbonyl carbons are extremely deshielded and the spectral width of the HMBC experiment was only 120 ppm, these appeared as folded peaks in the HMBC and could thus be assigned at 175.74 ppm. Similarly, there is a folded correlation between H3ax SA and the carboxyl carbon, C1 SA, that was assigned as 174.66 ppm.

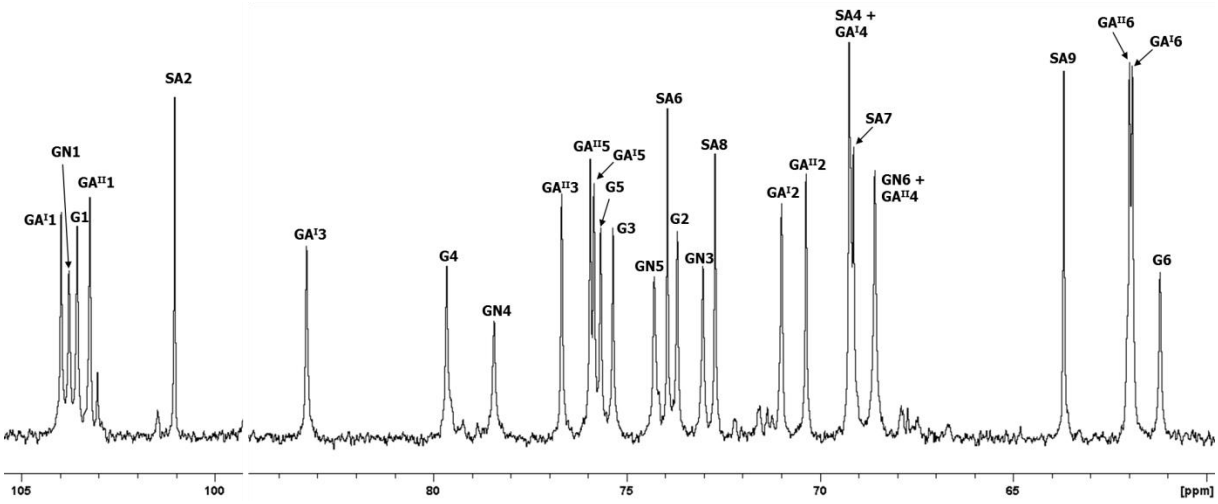


Figure 3.15: Expansion of the anomeric and ring regions of the labelled 1D ^{13}C NMR spectrum of the GBS III CPS.

Once all carbons have been assigned by the combination of HSQC, HSQC-TOCSY and HMBC experiments, the ^{13}C spectrum (Figure 3.15) can be fully labelled. The summary of which protons were assigned by heteronuclear correlation experiments such as HSQC and HMBC is presented in Figure 3.16.

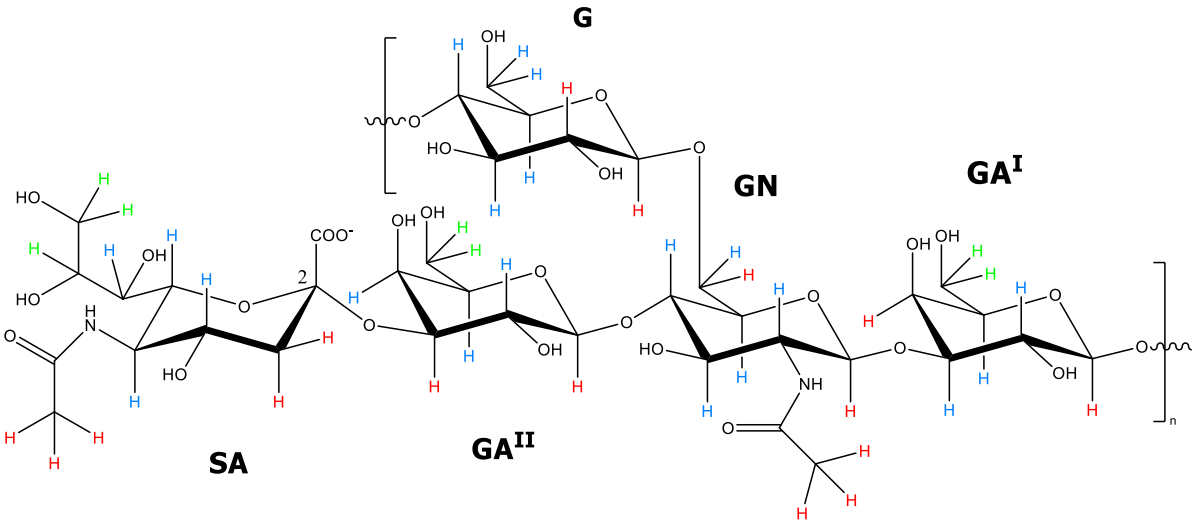


Figure 3.16: GBS III repeating unit structure highlighting the protons assigned by 1D H^{1}/DOSY (red), homonuclear correlation (blue) and heteronuclear correlation (green) experiments.

3.1.4 Full assignment and fingerprint summary

In summary, the 1D ^{1}H spectrum was used to identify well-separated peaks, chiefly the anomeric signals, that could be used as diagnostic peaks to begin the analysis. The majority of the spin systems could then be elucidated with a combination of COSY, TOCSY and NOESY experiments through correlations from these diagnostic peaks, whether by scalar coupling

through bonds or dipolar coupling through space. The NOESY could also confirm the proposed linkage sequence and positions through inter-residue correlations. With most protons preliminarily assigned, 2D heteronuclear experiments such as HSQC and HSQC-TOCSY confirmed those assignments and aided in assigning the ^{13}C spectrum and any unassigned protons. Finally, the HMBC experiment allowed assignment of the quaternary carbons that could not be detected in the HSQC experiments, as well as definitive confirmation of the linkage positions and sequence of the sugar residues in the CPS repeating unit, resulting in a complete assignment of the GBS III capsular polysaccharide. These assignments agree with previously published assignments recorded at 290 K [84], albeit with minor differences of <0.1 ppm attributed to the difference in experiment temperature.

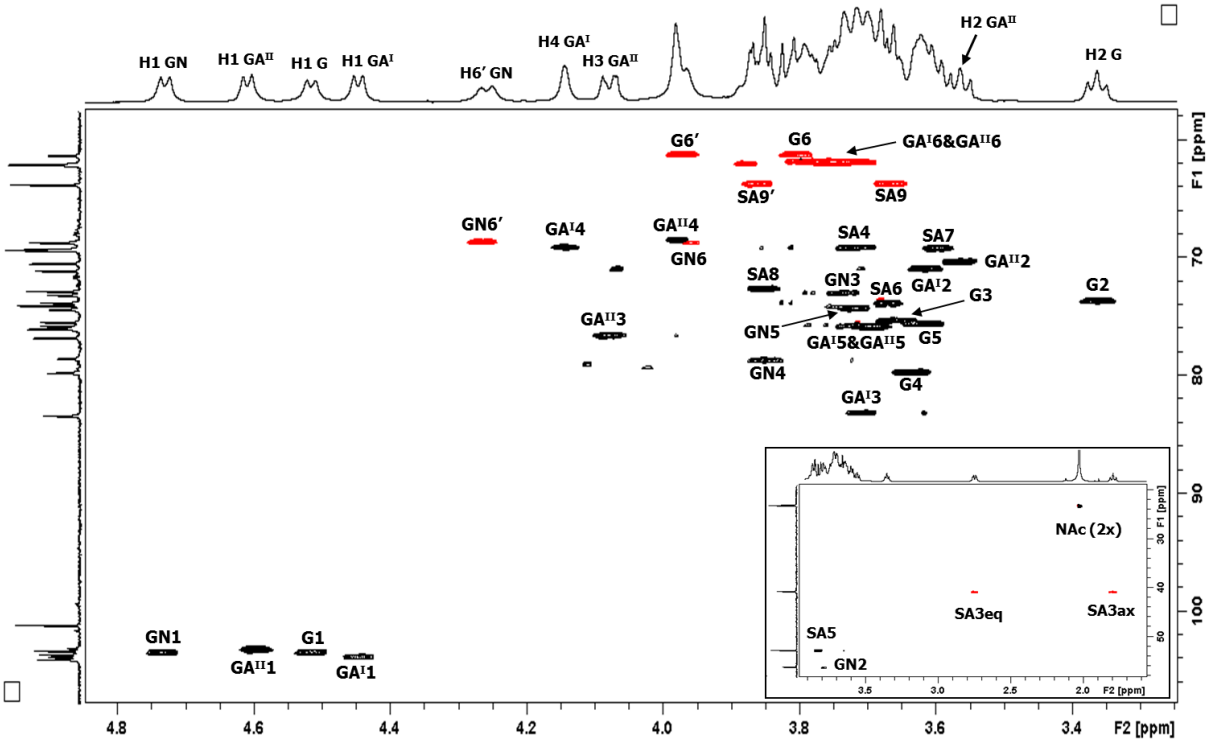


Figure 3.17: Fully labelled HSQC-DEPT spectrum of the GBS III CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

With all assignments made, the HSQC-DEPT spectrum serves as a 2D identity map (Figure 3.17) showing the chemical shift of every proton/carbon correlation in the GBS III CPS repeating unit. The chemical shift data is collected in Table 3.1, with calculated glycosylation shifts further confirming the linkage positions. The glycosylation shifts were taken as the difference between the assigned chemical shift of each RU proton and the chemical shift of that proton in an unsubstituted monosaccharide. The data for the unsubstituted monosaccharides was acquired from the CASPER online resource. Likewise, the 1D ^{1}H and ^{13}C

spectra can be used as simpler fingerprinting methods wherein the identification of a few diagnostic peaks such as the anomerics, H6 GN, H4 GA^{I}, H3 GA^{II} and H3ax/eq SA are sufficient to prove polysaccharide identity.

Table 3.1: Full NMR assignment of the GBS III CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

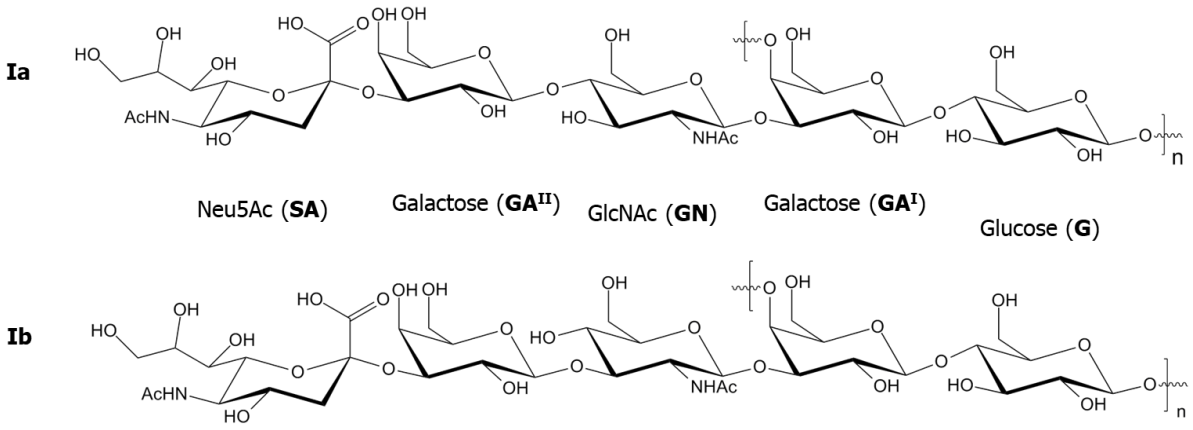
Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→4)-β-D-Glc*p*-(1→ G	4.52 103.51 (6.7)	3.36 73.73 (-1.5)	3.66 75.35 (-1.4)	3.64 <u>79.73</u> (9.0)	3.62 75.66 (-1.1)	3.80/3.97 61.28 (-0.5)	-	-	-
→4,6)-β-D-Glc*p*NAc-(1→ GN	4.74 103.52 (7.7)	3.79 56.18 (-1.7)	3.74 73.00 (-1.8)	3.85 <u>78.77</u> (7.7)	3.72 74.35 (-2.5)	3.96/4.27 <u>68.71</u> (6.9)	-	2.03	-
							175.74	23.18	-
→3)-β-D-Gal*p*-(1→ GA^{I}	4.45 103.89 (6.5)	3.62 70.98 (-2.0)	3.71 <u>83.23</u> (9.5)	4.14 69.16 (-0.5)	3.71 75.79 (-0.1)	3.75 61.88 (0.1)	-	-	-
							-	-	-
→3)-β-D-Gal*p*-(1→ GA^{II}	4.60 103.21 (5.8)	3.57 70.33 (-2.6)	4.08 <u>76.63</u> (2.9)	3.98 68.53 (-1.2)	3.69 75.92 (0.0)	3.75 61.88 (0.1)	-	-	-
							-	-	-
α-D-Neu*p*NAc-(2→ SA[a]	-	-	1.79/2.76	3.72	3.82	3.67	3.60	3.85	3.67/3.86
	174.66	101.02	40.68	69.22	52.74	73.93	69.12	72.69	63.67
			(-1.0)	(-0.2)	(0.3)	(0.5)	(-0.3)	(0.3)	(-0.4)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.03, 22.98 ppm, C=O at 175.99 ppm.

Thus, the structural assignment of the GBS III capsular polysaccharide is complete. The methods established in this chapter will be used in the following two chapters to fully elucidate the CPS structures of four other GBS serotypes, namely GBS Ia, Ib, II, IV and V.

Chapter 4. NMR characterisation of GBS Ia and Ib

The two Group B *Streptococcus* serotypes Ia and Ib are highly prevalent serotypes responsible for approximately 25% of global cases of GBS disease [7] and together with serotype III have been the target of a trivalent vaccine [129]. The fact that Ia and Ib were two different antigens was discovered by Lancefield in 1938 [53], but she could not determine which cellular structure was responsible for the difference. It was only when the repeating unit (RU) structures of the type Ia and Ib capsular polysaccharides were published, many years later in 1983 [130,155], that the reason for the immunological difference was determined. Both structures were determined using chemical methods supported by ^{13}C NMR data demonstrating that the GBS Ia and Ib CPSs both have a pentasaccharide repeating unit containing D-Galactose (Gal), D-Glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc) and N-Acetylneuraminic acid (Neu5Ac or sialic acid) in a molar ratio of 2:1:1:1 (Figure 4.1), differing by only one linkage position, Ia having a GA^{II}-(1→4)-GN linkage, while Ib has a GA^{II}-(1→3)-GN linkage.



Ia: →4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Ib: →4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Figure 4.1: Published repeating unit structures and line structures of the GBS Ia and Ib capsular polysaccharides.

This was the first proof that it was a single linkage difference between the two capsular polysaccharides that was responsible for the immunological difference between the two serotypes. While the GBS Ia and Ib CPSs contain the same sugars in the same ratios as those found in the GBS III CPS, compared to GBS type III, relatively little research has been performed on the NMR of these two serotypes, especially type Ib. Some NMR data of Ia

fragments, monomers and dimers [140–142] as well as ^{1}H spectra, anomeric assignments and HSQC spectra for both serotypes have been published [83,143], but a detailed ^{1}H and ^{13}C NMR study, including full assignments, has not yet been published.

4.1 Structural characterisation of the GBS Ia capsular polysaccharide

The full NMR characterisation of Ia described in this chapter was performed using the methods developed for serotype III in Chapter 3 and will be followed by the characterisation of serotype Ib as well as a comparison of their spectra to highlight key differences between these two structurally similar serotypes.

4.1.1. 1D ^{1}H and DOSY experiments

The 1D ^{1}H spectrum (Figure 4.2a), allowed the assignment of the anomeric and other diagnostic signals from the literature [83]. The four anomeric signals between 4.4 and 5.0 ppm, together with the H3 axial/equatorial sialic acid signals at 1.79 and 2.77 ppm correspond to a pentasaccharide repeating unit as published. The DOSY experiment (Figure 4.2b) reveals that, at 343 K, the water signal is directly overlapped with a signal at 4.35 ppm.

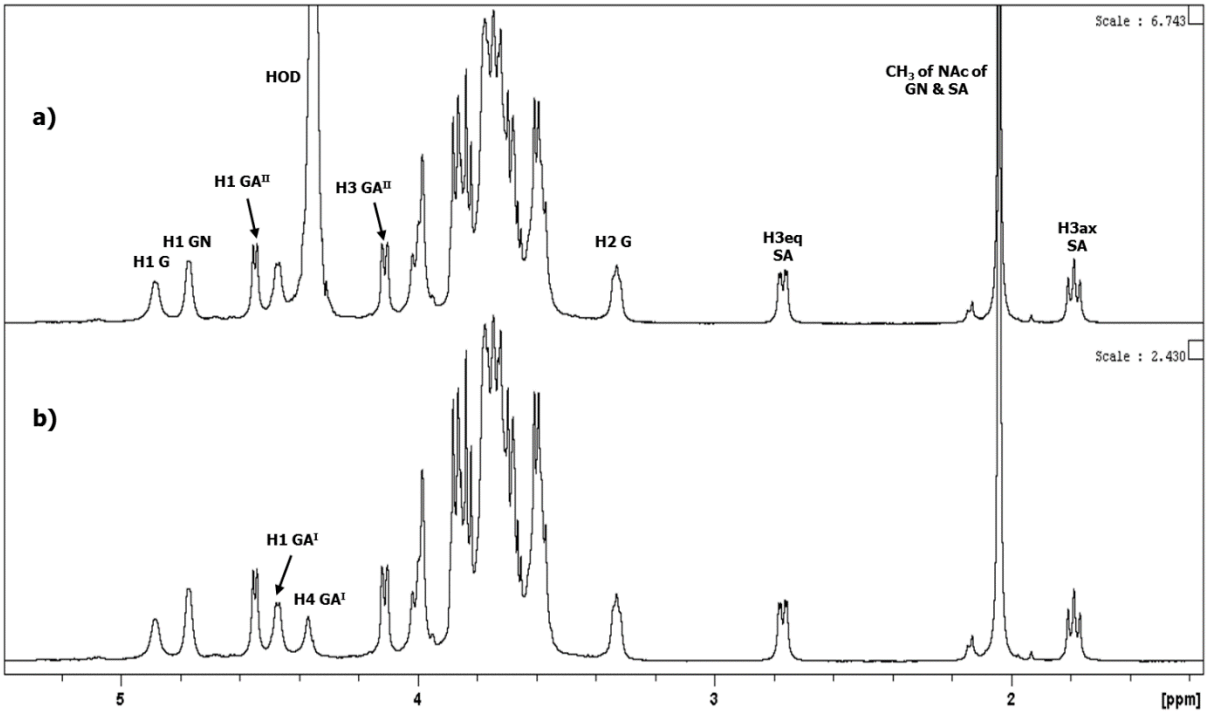


Figure 4.2: Overlay of 1D ^{1}H (a) and DOSY (b) spectra of the GBS Ia CPS recorded at 343 K with diagnostic peaks labelled.

The four anomeric peaks at 4.89, 4.77, 4.55 and 4.48 ppm were assigned to H1 of G, GN, GA^{II} and GA^{I} respectively and the fact that all four resonate below 4.9 ppm is consistent with the

evaluation that they are β-linked hexoses. Additional peaks could be assigned, namely H4 GA^{I}, H3 GA^{II} and H2 G in the ring region (4.35, 4.11 and 3.33 ppm respectively) as well as H3eq/ax SA and the overlapped methyl signals from the N-acetyl groups of GN and SA in the upfield region (2.77/1.79 and 2.04 ppm respectively).

4.1.2. Homonuclear correlation experiments (COSY, TOCSY and NOESY)

The 2D homonuclear correlation experiments are again the source of most preliminary assignments using the identified diagnostic peaks as starting points for correlations, allowing elucidation of most of the spin system signals.

The most downfield anomeric proton (H1 G, 4.89 ppm) shows a COSY (Figure 4.3) correlation to the signal at 3.33 ppm, already assigned as H2 G, and this correlation continues in the 1D TOCSY (Figure 4.4) to three signals at 3.58, 3.63 and 3.69 ppm, assigned as H5, H4 and H3 G respectively. Additionally, two weak TOCSY correlations appear at 3.81 and 3.99 ppm when the intensity of the spectrum is increased, tentatively assigning those as H6/6' G, completing the preliminary assignment of the Glc residue.

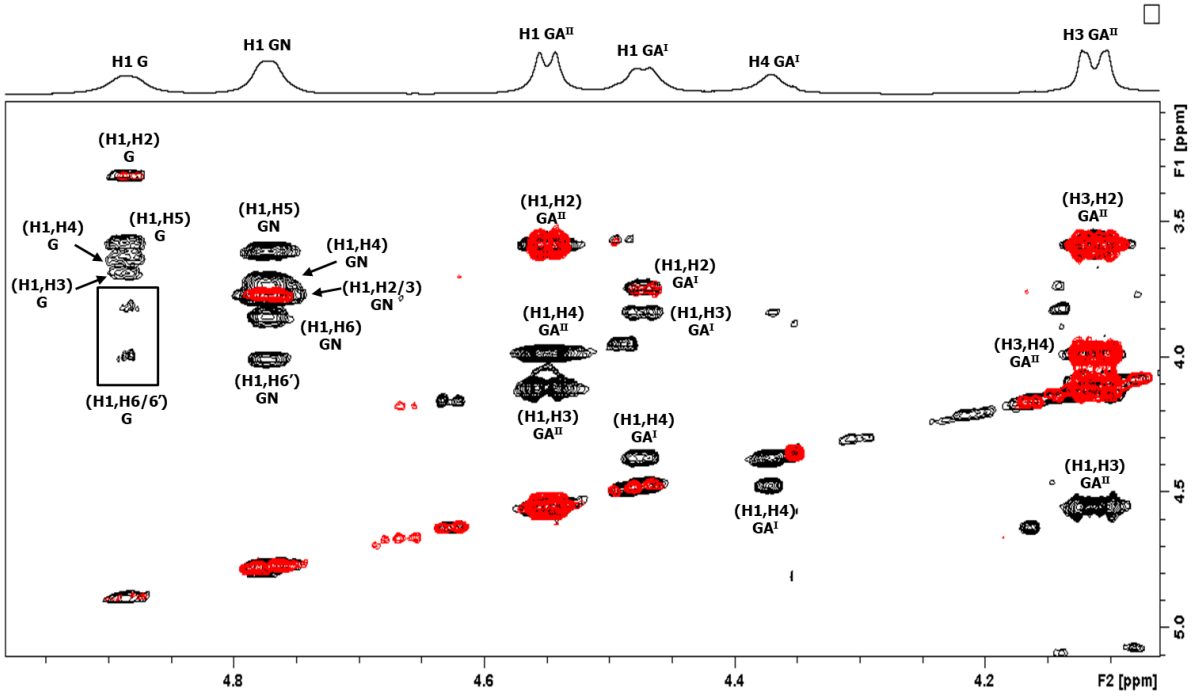


Figure 4.3: Anomeric region of the 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) with identified crosspeaks for G, GN, GA^{I} and GA^{II} labelled. The insert below the glucose anomeric crosspeaks shows low-intensity H6 G crosspeaks.

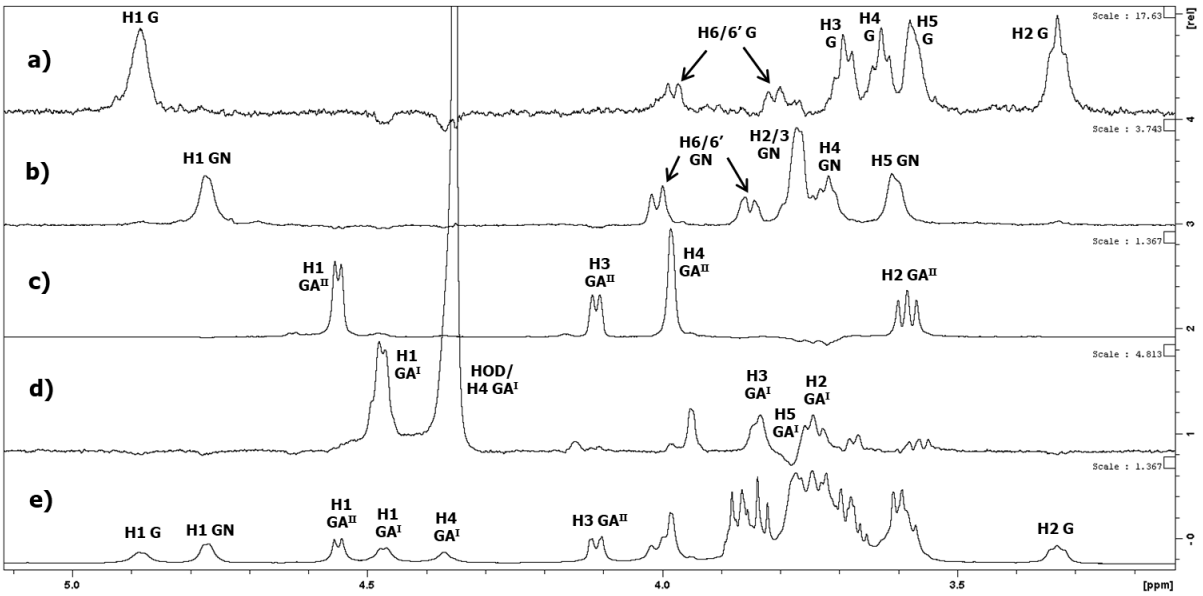


Figure 4.4: Labelled anomeric and ring regions of 1D TOCSY experiments wherein the anomeric protons of G, GN, GA^{II}, and GA^{I} (a-d respectively) were selectively irradiated overlaid onto the DOSY spectrum (e) of the GBS Ia CPS.

Of note is that the anomeric signals of G and GA^{I} are broader and of lower intensity than those of the branch residues, likely due to restricted movement of the backbone resulting in shorter relaxation times. This results in lower intensity crosspeaks in both 1D and 2D correlation experiments. The anomeric proton of the GlcNAc residue (H1 GN, 4.77 ppm) shows a correlation in the COSY to a signal at 3.76 ppm that is larger in the TOCSY spectrum and was thus assigned as the overlapped signals of H2 and H3 GN. Additional TOCSY correlations appear at 3.61, 3.72, 3.85 and 4.01 ppm, which were assigned as H5, H4 and H6/6' GN respectively, completing the preliminary assignment of the GlcNAc spin system. The most downfield Gal anomeric signal (H1 GA^{II}, 4.55 ppm) shows a correlation in the COSY with a signal at 3.58 ppm, assigned as H2 GA^{II}, which can be continued to the signal at 4.11 ppm previously assigned as H3 GA^{II} which in turn continues to a signal at 3.98 ppm assigned as H4 GA^{II}. This, as expected, is where the Gal correlations cease. The final anomeric signal (H1 GA^{I}, 4.48 ppm) similarly correlates to a signal at 3.74 ppm in the COSY, assigned as H2 GA^{I}, and two additional signals in the TOCSY at 3.83 and 4.35 ppm, assigned as H3 and H4 GA^{I} respectively.

The last spin system to be assigned is that of sialic acid, which shows TOCSY correlations (Figure 4.5) from the H3ax/eq signals at 1.79 and 2.77 ppm to signals at 3.60, 3.69, 3.73 and

3.84 ppm which were assigned as H7, H6, H4 and H5 respectively. This leaves H5 and H6 of GA^{I} and GA^{II} as well as H8 and H9 of SA unassigned.

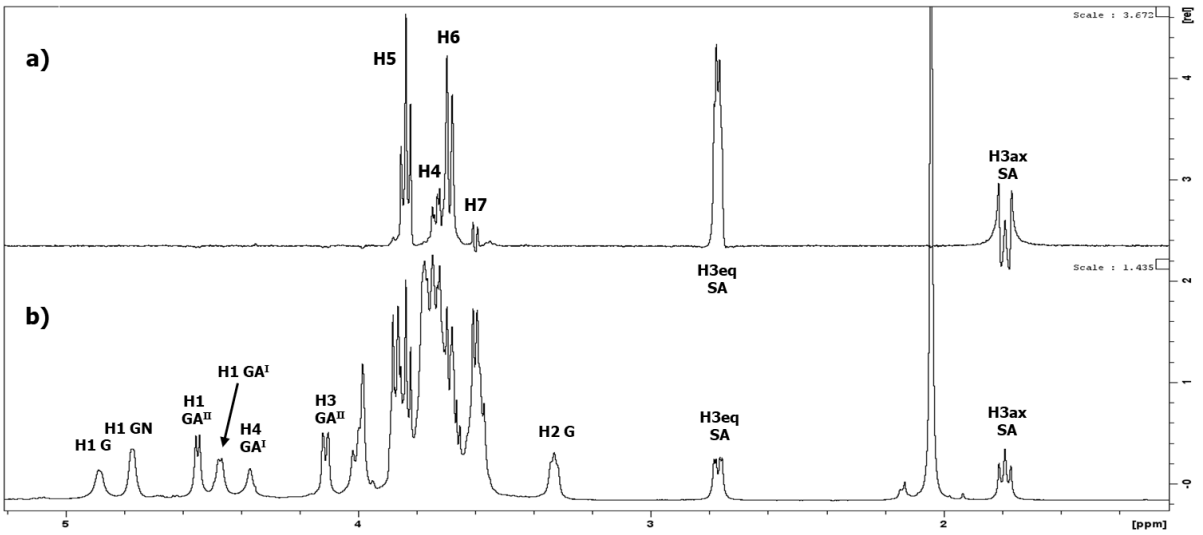


Figure 4.5: Labelled 1D TOCSY experiments wherein the H3eq SA (a) protons were selectively irradiated overlaid onto the DOSY spectrum (b) of the GBS Ia CPS.

In the NOESY (Figure 4.6), selective irradiation of the anomeric signal of Glc (Figure 4.6a) shows a number of inter-residue correlations including ones to H3 and H4 GA^{I}, agreeing with the published G-(1→4)-GA^{I} linkage, while the GN anomeric signal (Figure 4.6b) shows correlations to H2 and H3 GA^{I}, in agreement with a GN-(1→3)-GA^{I} linkage.

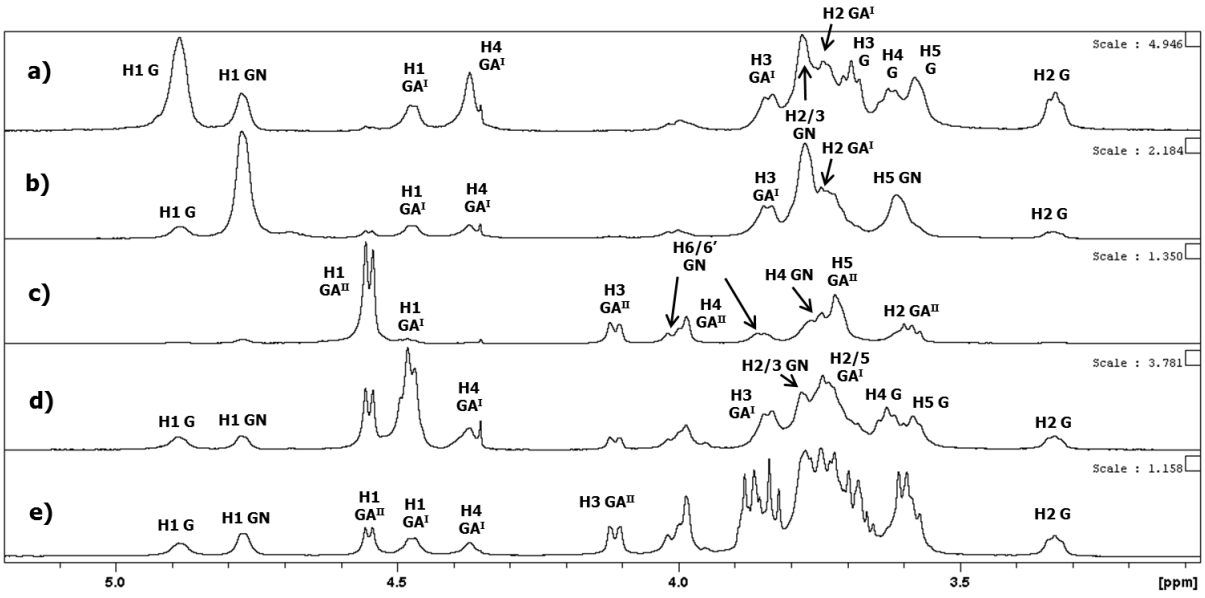


Figure 4.6: Labelled anomeric and ring regions of 1D NOESY experiments wherein the anomeric protons of G, GN, GA^{II}, and GA^{I} (a-d respectively) were selectively irradiated overlaid onto the DOSY spectrum (e) of the GBS Ia CPS.

Additional correlations can be seen from H1 G to H1 and H2/3 GN (and inversely, weak signals from H1 GN to H1 and H2 G) which is not surprising because the GN-(1→3)-GA^{I} linkage is directly next to the G-(1→4)-GA^{I} linkage, meaning H1 G and H1 GN are forced to be in relative proximity.

Intra-residue correlations from H1 GA^{II} (Figure 4.6c) allowed for the assignment of H5 GA^{II} at 3.72 ppm, while inter-residue correlations to H4 and H6/6' GN supported a GA^{II}-(1→4)-GN linkage. The anomeric signal of GA^{I} shows correlations to H4 and H5 GN, which is in agreement with a GA^{I}-(1→4)-G linkage. Additionally, a signal at approximately 3.75 ppm likely arises from both H2 and H5 GA^{I}, but this cannot be confirmed in the NOESY. Lastly, the H3ax/eq SA protons (Figure 4.7) show NOESY correlations to H2, H3 and H4 GA^{II}, confirming the SA-(2→3)-GA^{II} linkage. This left only H6 of GA^{I} and GA^{II}, as well as H8 and H9 of SA unassigned.

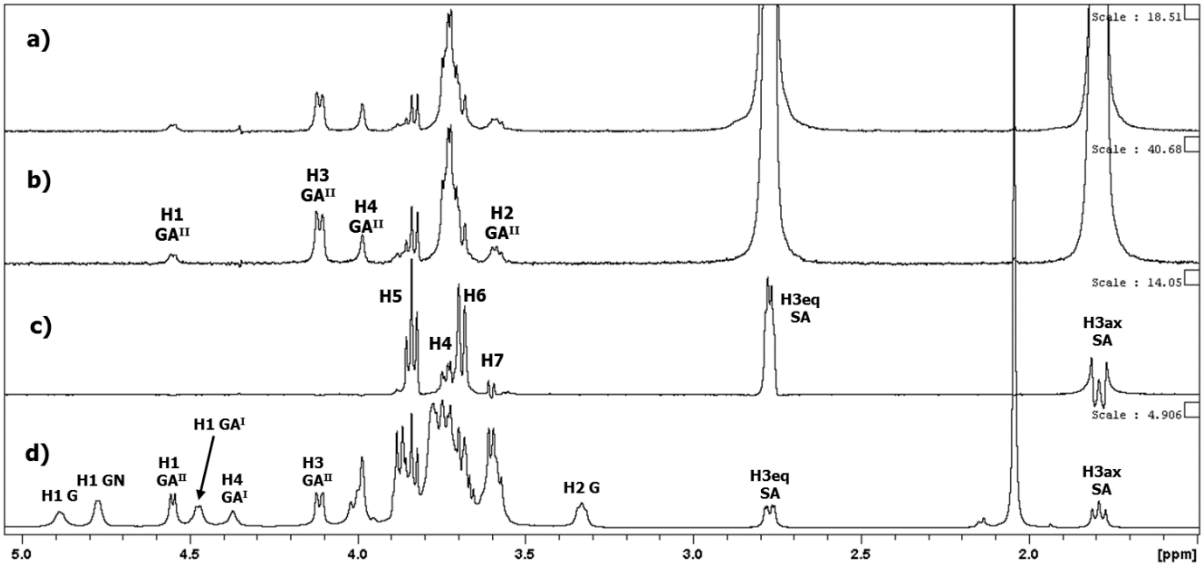


Figure 4.7: Labelled 1D NOESY (a,b) and TOCSY (c) experiments wherein the H3ax (b) or H3eq (a,c) protons of SA were selectively irradiated overlaid onto the DOSY spectrum (d) of the GBS Ia CPS.

4.1.3. Heteronuclear experiments (HSQC, HMBC and hybrid)

As has been established with GBS III in the previous chapter, the full ^{13}C NMR assignment and confirmation of all proton assignments was performed through the combined use of HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC experiments. With the general strategy for assignment of residue spin systems developed for GBS serotype III, the previously described overlay of HSQC-DEPT, HSQC-TOCSY and 1D TOCSY experiments were used to confirm preliminary assignments and assign any remaining unassigned protons. Figure 4.8 shows the

full assignment and confirmation of the GlcNAc spin system as an example. The HSQC-TOCSY signals arranged vertically above the anomeric signal show correlations between the anomeric proton and the ring carbons in the GN residue, allowing for the carbon assignments, especially when cross-referenced with the proton assignments shown by the selective 1D TOCSY overlayed over the bottom of the spectrum.

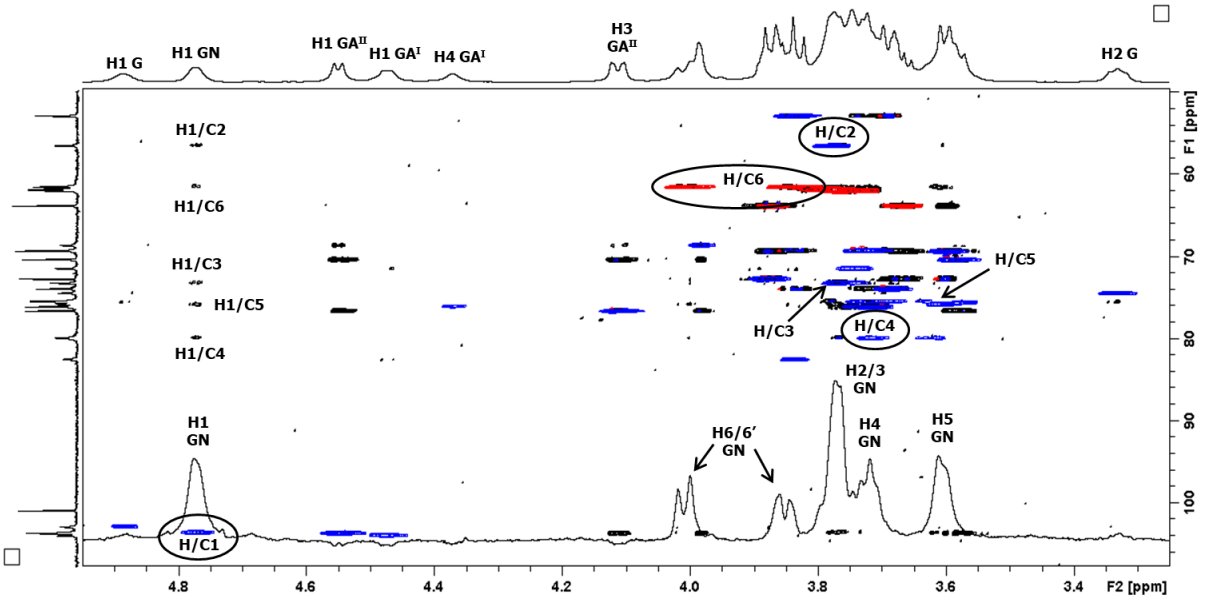


Figure 4.8: Anomeric and ring regions of the selective 1D TOCSY of H1 GN (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS Ia CPS, highlighting correlations that show the proton and carbon assignments of the entire GlcNAc spin system.

The HSQC-TOCSY (Figure 4.9) allowed assignment of the final unassigned protons, H8 and H9/9' of SA, by their correlations with H7 SA, assigning them as 3.88 and 3.67/3.87 ppm respectively, and H6 of both GA^{I} and GA^{II} by a process of elimination since they, as the only unassigned protons, must be the final large unassigned peak at 3.75 ppm. They were given the same assignment since the data does not give a means to tell them apart in the broad, overlapped signal.

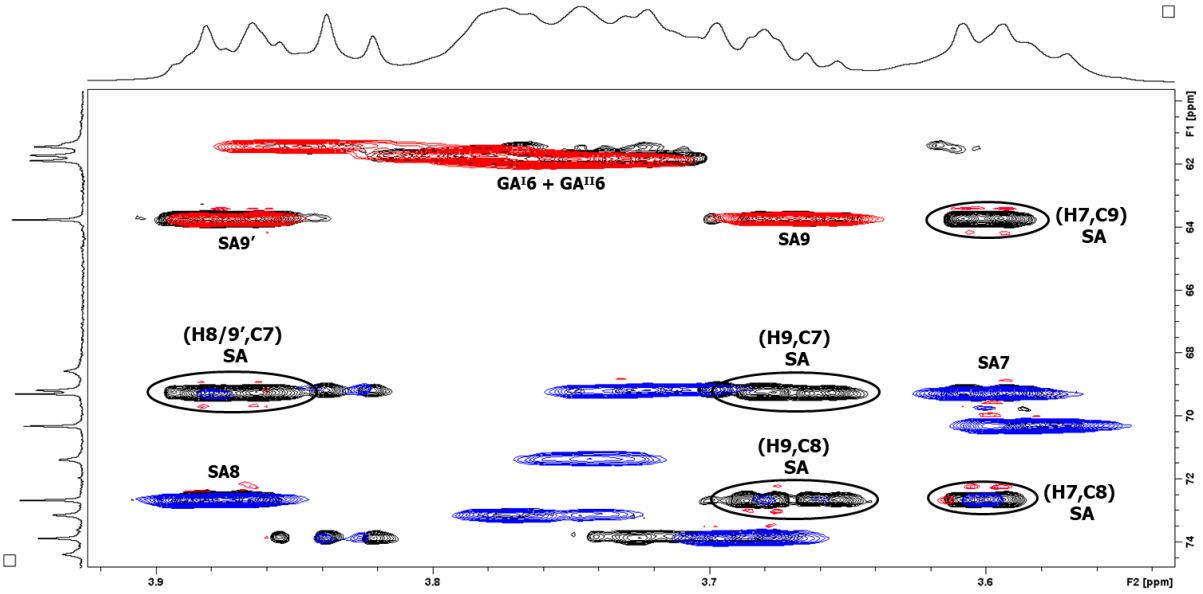


Figure 4.9: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS Ia CPS, highlighting correlations that led to the assignment of H8 and H9 of SA.

The HMBC (Figure 4.10) spectrum allowed some confirmation of the repeating unit connectivity shown by the NOESY (Figure 4.7) as well as the assignment of the quaternary carbons in the polysaccharide, by showing intra- and inter-residue correlations through 3-bond coupling.

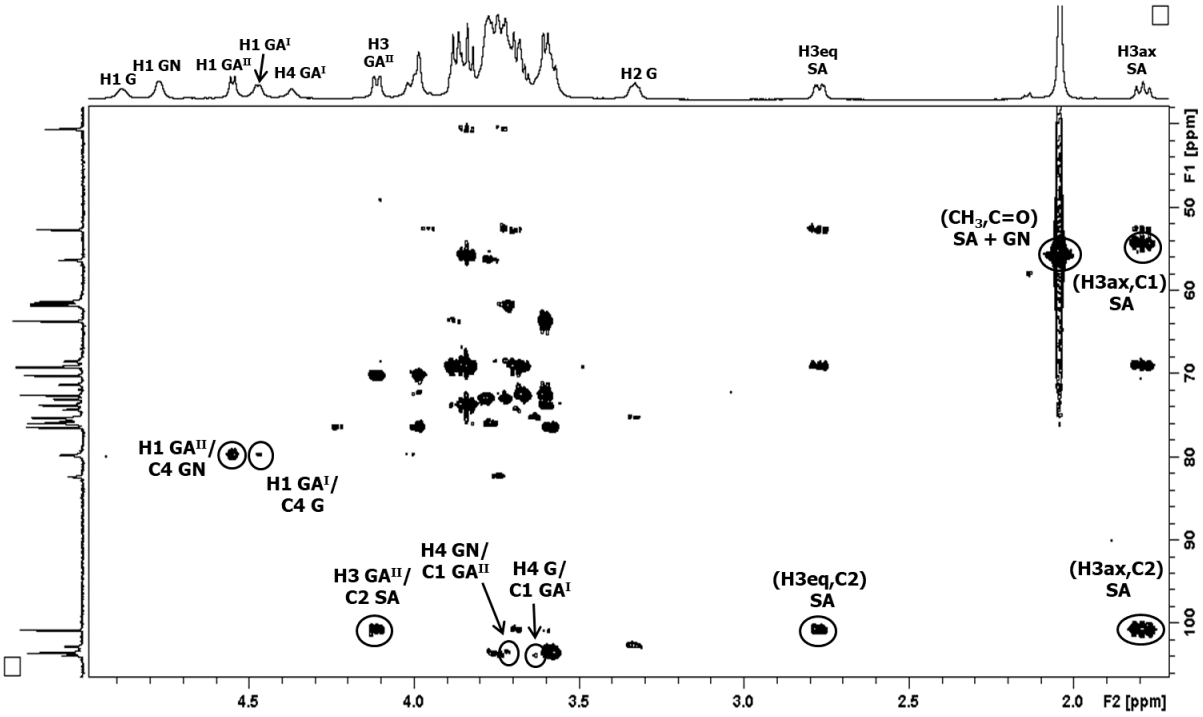


Figure 4.10: HMBC spectrum of the GBS Ia CPS with important intra- and inter-residue signals labelled.

Only two of the five linkages in the polysaccharide, namely GA^{II}-(1→4)-GN and GA^{I}-(1→4)-G, could be confirmed by correlations between the anomeric protons and linkage site carbons in Figure 4.10, with an additional linkage, namely SA-(2→3)-GA^{II}, confirmed by the reverse correlation from the linkage ring proton to the anomeric carbon. Unfortunately, the low intensities of the G and GA^{I} residue signals, likely due to restricted movement in the polysaccharide backbone, result in low-intensity 1D and 2D crosspeaks with signals from those residues, so the GN-(1→3)-GA^{I} and G-(1→4)-GA^{I} linkages could not be confirmed by HMBC.

Lastly, HMBC was used to assign the quaternary carbons in the polysaccharide. The carbon assignment of C2 SA (100.89 ppm) was relatively simple, showing intra-residue correlations to H3eq/ax, as well as an inter-residue correlation to H3 GA^{II}, which was also used to confirm the SA-(2→3)-GA^{II} linkage. The carbonyl carbons of the N-Acetyl groups in both sialic acid and GlcNAc could be identified by a strong correlation to the CH_3 signal at approximately 55 ppm. Since carbonyl carbons are extremely deshielded and the spectral width of the HMBC experiment was only 120 ppm, these are folded peaks and thus assigned at 175.49 and 175.93 ppm for the GN and SA carbonyl carbons respectively. Similarly, there is a folded correlation between H3ax SA and the carboxyl carbon, C1 SA, that was assigned as 174.47 ppm.

In an attempt to confirm the GN-(1→3)-GA^{I} and G-(1→4)-GA^{I} linkages by more than just homonuclear NOESY, an HSQC-NOESY (Figure 4.11) spectrum was performed. While it does show a small correlation confirming the GN-(1→3)-GA^{I} linkage, the G-(1→4)-GA^{I} linkage could not be seen, meaning it was only assigned in the 1D and 2D homonuclear NOESY experiments.

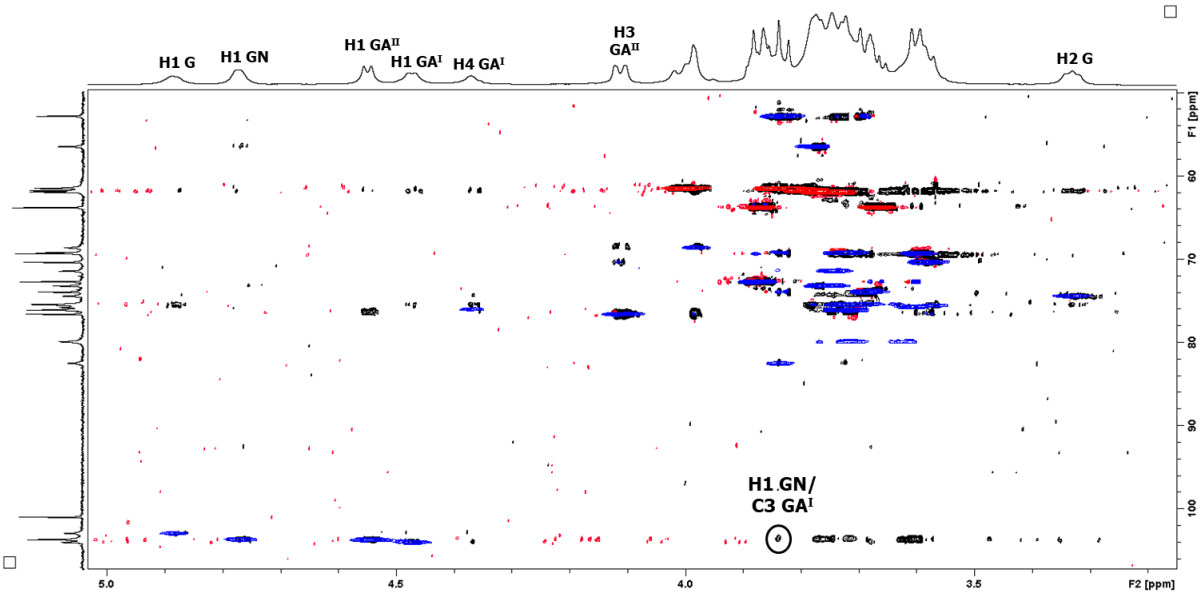


Figure 4.11: Anomeric and ring regions of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-NOESY spectrum (black) of the GBS Ia CPS, highlighting a correlation that confirms the GN-(1→3)-GA^{I} linkage.

Once all carbons have been assigned by the combination of HSQC, HSQC-TOCSY and HMBC experiments, the ^{13}C spectrum (Figure 4.12) can be fully labelled.

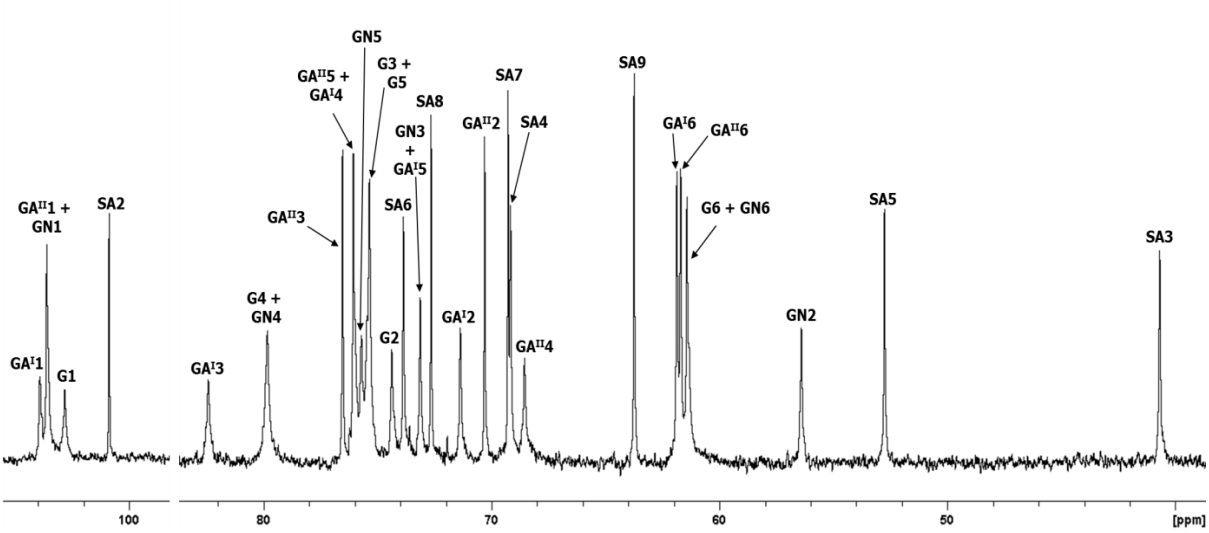


Figure 4.12: Expansion of the anomeric and ring regions of a labelled 1D ^{13}C NMR spectrum of the GBS Ia CPS.

4.1.4 Full assignment and fingerprint summary

With all assignments made, the HSQC-DEPT spectrum constitutes a 2D identity map (Figure 4.13) showing the chemical shift of every proton/carbon correlation in the GBS Ia CPS repeating unit. The chemical shift data is collected in Table 4.1, with calculated glycosylation shifts further confirming the linkage positions. These assignments agree with previously

published anomeric chemical shift values (recorded at 298 K) [83], with slight differences in chemical shift attributed to the difference in experiment temperature. They also agree somewhat with published data of a GBS type Ia CPS dimer [142], but with significant differences since the dimer is just fragment of the intact repeating unit and thus the terminal ends of the dimer will be in different magnetic environments than the linked ends of the RU characterised here. The 1D ^{1}H and ^{13}C spectra can also be used as simpler fingerprinting methods wherein the identification of a few diagnostic peaks such as the anomerics, H4 GA^{I}, H3 GA^{II}, H2 G and H3ax/eq SA are sufficient to prove polysaccharide identity.

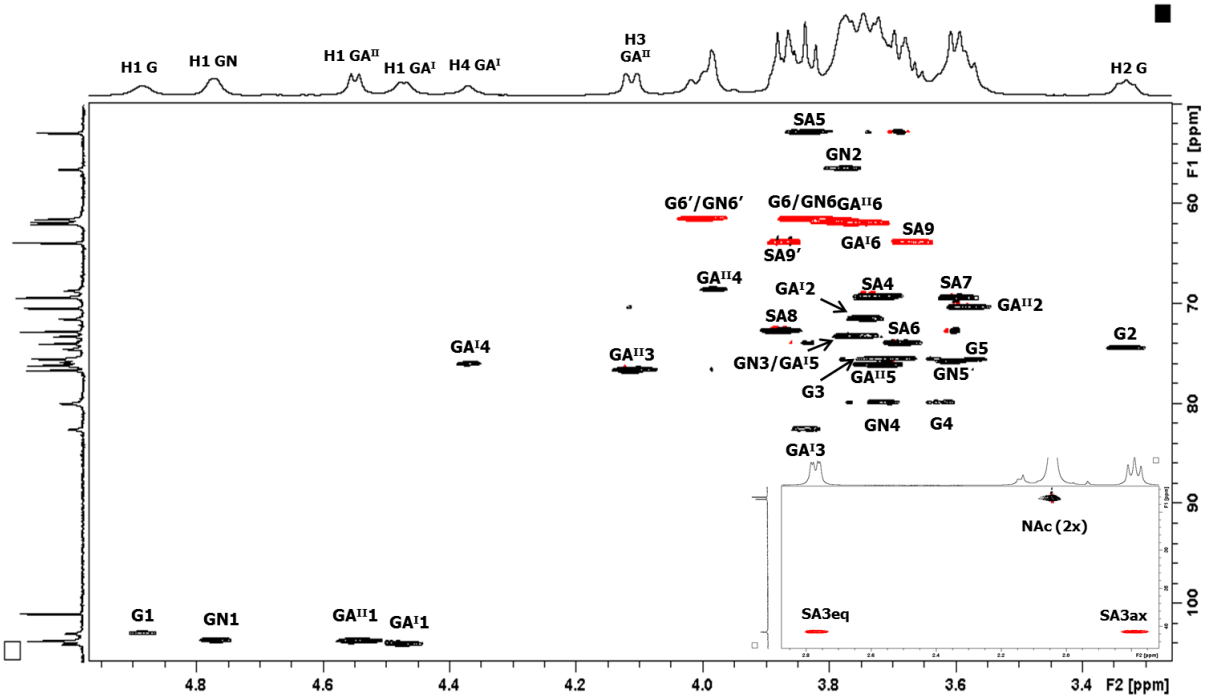


Figure 4.13: Fully labelled HSQC-DEPT spectrum of the GBS Ia CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

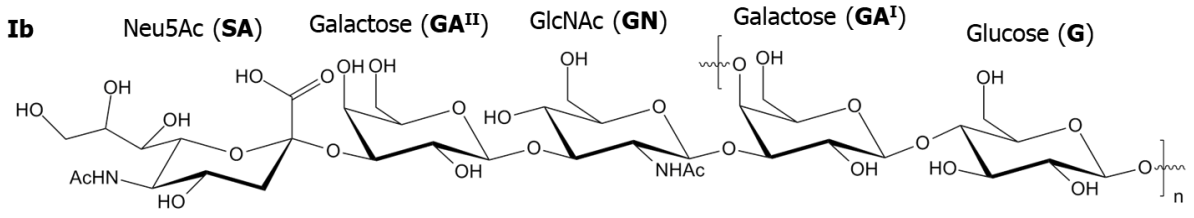
Table 4.1: Full NMR assignment of the GBS Ia CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→4)-β-D-Glc*p*-(1→ G	4.89 102.83 (6.0)	3.33 74.38 (-0.8)	3.69 75.35 (-1.4)	3.63 <u>79.82</u> (9.1)	3.58 75.45 (-1.3)	3.81/ 3.99 61.42 (-0.4)	- -	- -	- -
→4)-β-D-Glc*p*NAc-(1→ GN	4.77 103.57 (7.7)	3.76 56.40 (-1.5)	3.77 73.12 (-1.7)	3.72 <u>79.82</u> (8.8)	3.61 75.71 (-1.1)	3.85/ 4.01 61.42 (-0.4)	- 175.49	2.04 23.24	- -
→3,4)-β-D-Gal*p*-(1→ GA^{I}	4.48 103.93 (6.6)	3.74 71.36 (-1.6)	3.83 <u>82.42</u> (8.6)	4.35 <u>75.98</u> (6.3)	3.77 73.86 (-2.1)	3.75 61.86 (0.0)	- -	- -	- -
→3)-β-D-Gal*p*-(1→ GA^{II}	4.55 103.63 (6.3)	3.58 70.30 (-2.7)	4.11 <u>76.53</u> (2.8)	3.98 68.54 (-1.2)	3.72 76.05 (0.1)	3.78 61.70 (-0.1)	- -	- -	- -
α-D-Neu*p*NAc-(2→ SA[a]	- 174.47	- 100.89	1.79/ 2.77 40.68 (-1.0)	3.73 69.16 (-0.3)	3.84 52.75 (0.3)	3.69 73.86 (0.4)	3.60 69.27 (-0.2)	3.88 72.64 (0.3)	3.67/ 3.87 63.74 (-0.3)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.04, 22.95 ppm, C=O at 175.93 ppm.

4.2 Structural characterisation of the GBS Ib capsular polysaccharide

Following the assignment of the CPS of GBS Ia, the assignment of Ib is relatively simple since their similarities in structure result in similar proton and carbon assignments. The most marked differences in assignment occur around the GN-(1→3)-GA^{I} linkage since that is the only structural difference between the GBS Ia and Ib polysaccharides.



Ib: →4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→3)-β-D-Glc*p*NAc-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Figure 4.14: Published repeating unit structure and line structure of the GBS Ib capsular polysaccharide.

4.2.1. 1D ^{1}H and DOSY experiments

Due to the similarities in structure between the GBS type Ia and Ib polysaccharides, many of their proton assignments are similar, but the most noticeable difference in the 1D ^{1}H

spectrum (Figure 4.10), for fingerprinting purposes, is the upfield shift of H1 GA^{II}, causing it to overlap with H1 GA^{I}. This difference stems from the fact that the GA^{II}→GN linkage is a (1→4) linkage in type Ia and a (1→3) linkage in Ib. The comparison between the spectra of serotypes Ia and Ib is explored further in Chapter 4.3.

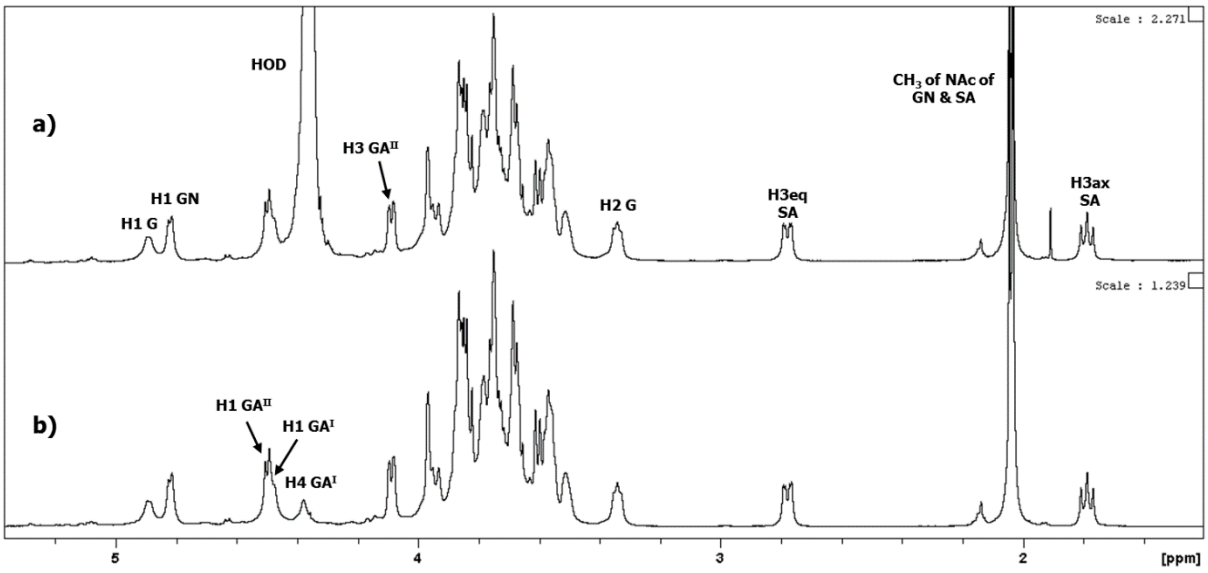


Figure 4.15: Overlay of 1D ^{1}H (a) and DOSY (b) spectra of the GBS Ib CPS recorded at 600 MHz and 343 K with diagnostic peaks labelled.

Four anomeric signals between 4.4 and 5.0 ppm, together with the H3 axial/equatorial sialic acid signals at 1.79 and 2.78 ppm correspond to a pentasaccharide repeating unit as expected. As with type Ia, the DOSY experiment reveals a signal at 4.38 ppm overlapped with the water signal. The two downfield anomeric peaks at 4.89 and 4.82 ppm were assigned as H1 G and H1 GN respectively. The broad signal at approximately 4.5 ppm was assigned as overlapped H1 GA^{II} (4.50 ppm) and H1 GA^{I} (4.48 ppm). All four anomeric signals resonate below 4.9 ppm which is consistent with GBS Ia and the evaluation that they are all β-linked hexoses. Additional diagnostic peaks were also assigned, namely H4 GA^{I}, H3 GA^{II} and H2 G in the ring region (4.35, 4.11 and 3.33 ppm respectively) and H3eq/ax SA in the upfield region (2.78/1.79 ppm). Unlike type Ia, the methyl signals from the N-acetyl groups of GN and SA in type Ib are not completely overlapped, resulting in two identifiable peaks at 2.04 and 2.05 ppm for GN and SA respectively.

4.2.2. Homonuclear correlation experiments (COSY, TOCSY and NOESY)

As with GBS Ia, the 2D homonuclear correlation experiments (Figures 4.3 and 4.4) elucidated most of the spin system signals, from H1-H6 of Glc and GlcNAc and H1-H4 of the two Gal residues.

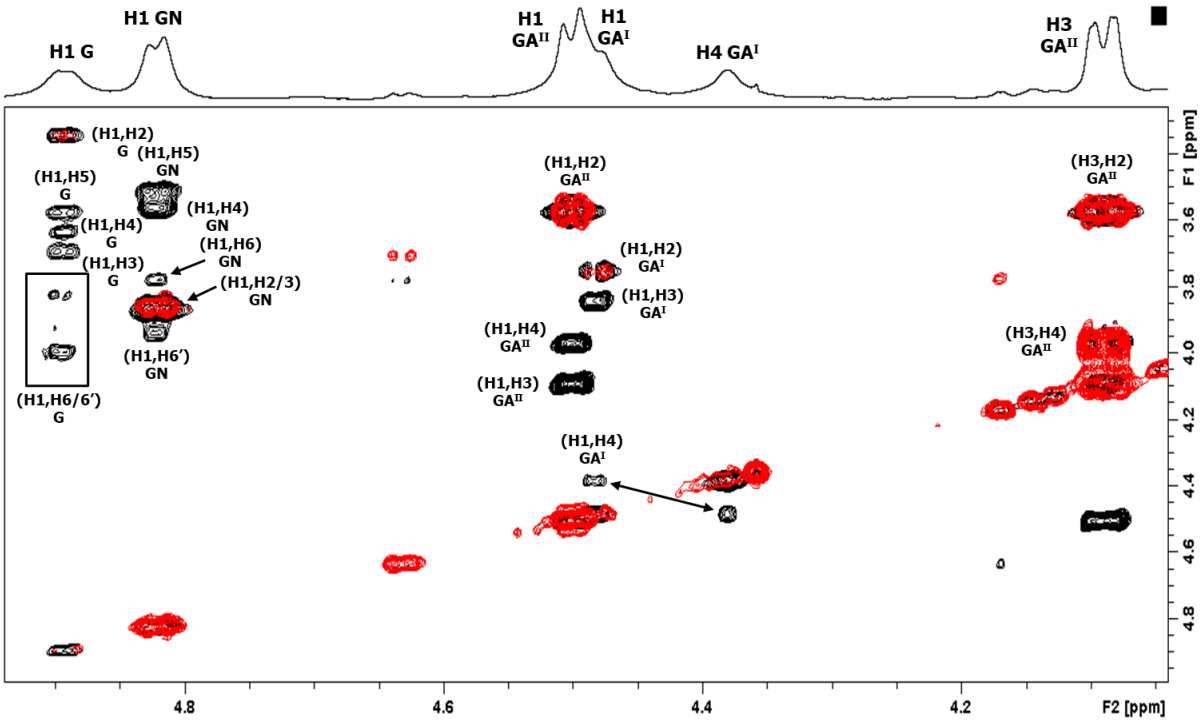


Figure 4.16: Anomeric region of the 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of GBS Ib with identified crosspeaks for G, GN, GA^{I} and GA^{II} labelled. The insert below the glucose anomeric crosspeaks shows low-intensity H6 G crosspeaks.

The COSY/TOCSY overlay (Figure 4.16) shows correlations from the anomeric signal of Glc to signals at 3.34, 3.58, 3.63 and 3.70 ppm, assigned as H2, H5, H4 and H3 G respectively. Since the Glc anomeric signal is broad and of low intensity, as with Ia, the TOCSY correlations to H6/6' G (3.82/3.98 ppm) can only be seen by increasing the spectral intensity (Figure 4.16 insert), thus completing the preliminary assignment of the Glc residue. The anomeric proton of the GlcNAc residue (H1 GN, 4.82 ppm) shows a correlation in the COSY to a large signal at 3.85 ppm, and was thus assigned as the overlapped signals of H2 and H3 GN. More TOCSY correlations appear at 3.51, 3.56, 3.77 and 3.94 ppm, which were assigned as H5, H4 and H6/6' GN respectively, completing the preliminary assignment of the GlcNAc spin system. Even though, H1 GA^{II} and H1 GA^{I} are overlapped, the 2D TOCSY spectrum clearly shows the two separate spin systems and thus the three signals at 3.57, 3.97 and 4.09 ppm could be assigned as H2, H4 and H3 GA^{II} respectively, while the three lower intensity signals at 3.75,

3.84 and 4.38 ppm were assigned as H2, H3 and H4 GA^{I} respectively. This overlap does, however, mean that selective 1D TOCSY (Figure 4.17) or NOESY correlation experiments are less effective.

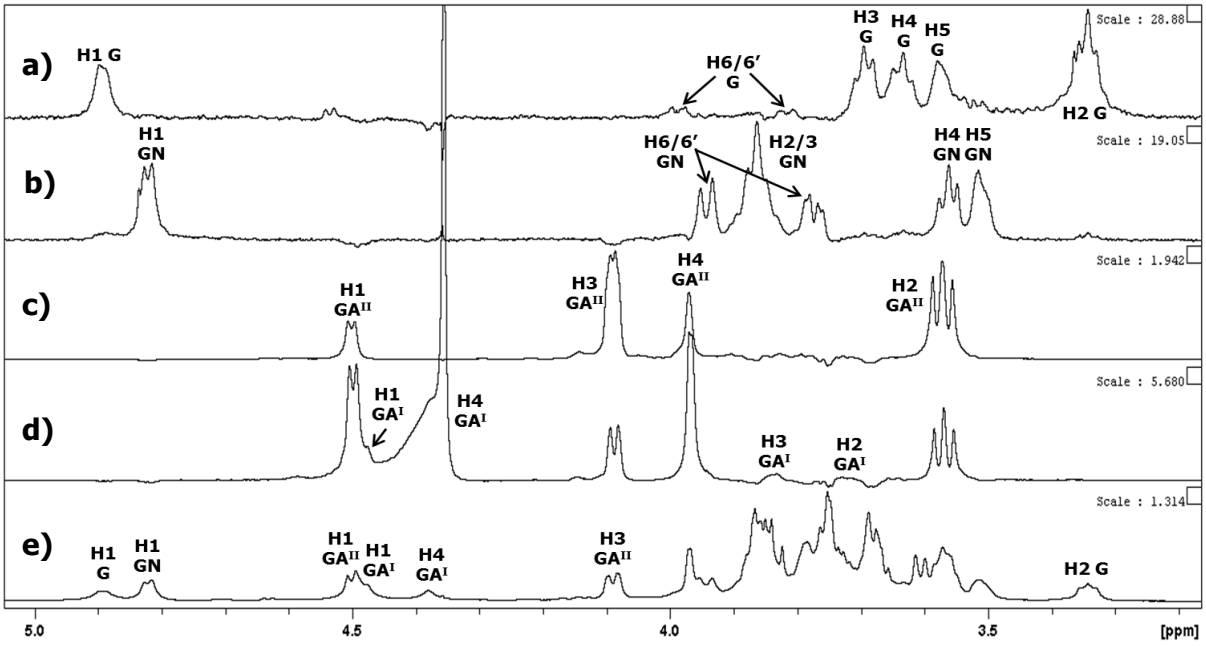


Figure 4.17: Labelled anomeric and ring regions of 1D TOCSY experiments wherein the anomeric protons of G, GN, GA^{II}, and GA^{I} (a-d respectively) were selectively irradiated overlaid onto the DOSY spectrum (e) of the GBS Ib CPS.

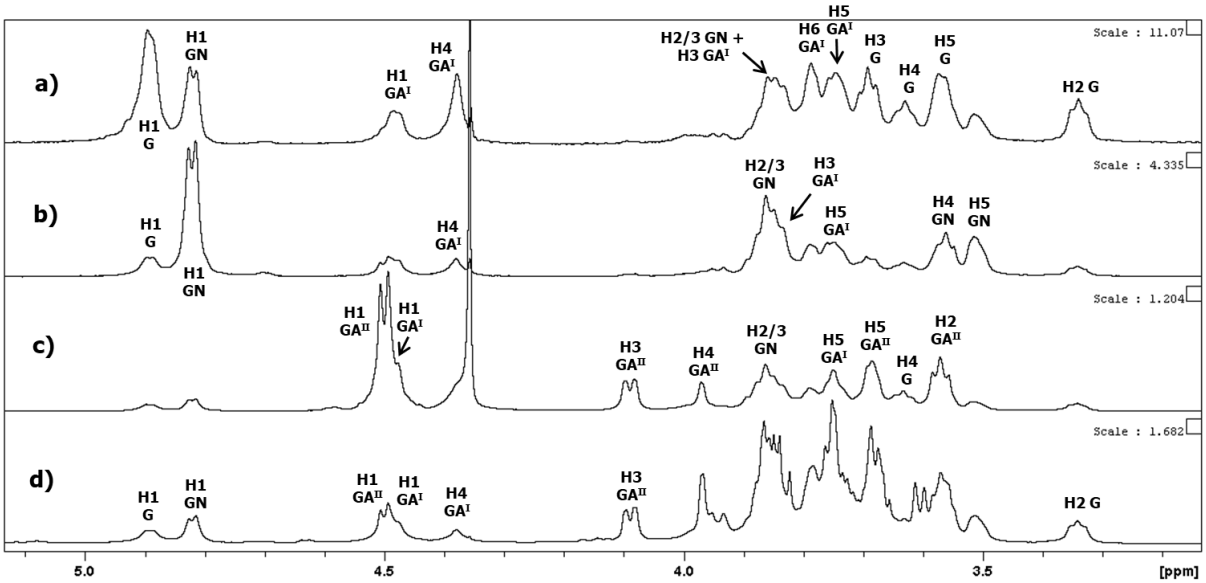


Figure 4.18: Labelled anomeric and ring regions of 1D NOESY experiments wherein the anomeric protons of G, GN, and overlapped GA^{I}/GA^{II} (a-c respectively) were selectively irradiated overlaid onto the DOSY spectrum (d) of the GBS Ib CPS.

In the NOESY (Figure 4.18), selective irradiation of the anomeric signal of Glc (Figure 4.13a) shows a number of inter-residue correlations to H3, H4 and H5 GA^{I} agreeing with the published G-(1→4)-GA^{I} linkage, as well as H1 and H2/3 GN which, as with type Ia, likely arise from the fact that the Glc and GlcNAc residues are linked to adjacent positions of the GA^{I} ring. Correlations from H1 GN to H3 and H4 GA^{I} agree with the GN-(1→3)-GA^{I} linkage (Figure 4.13b).

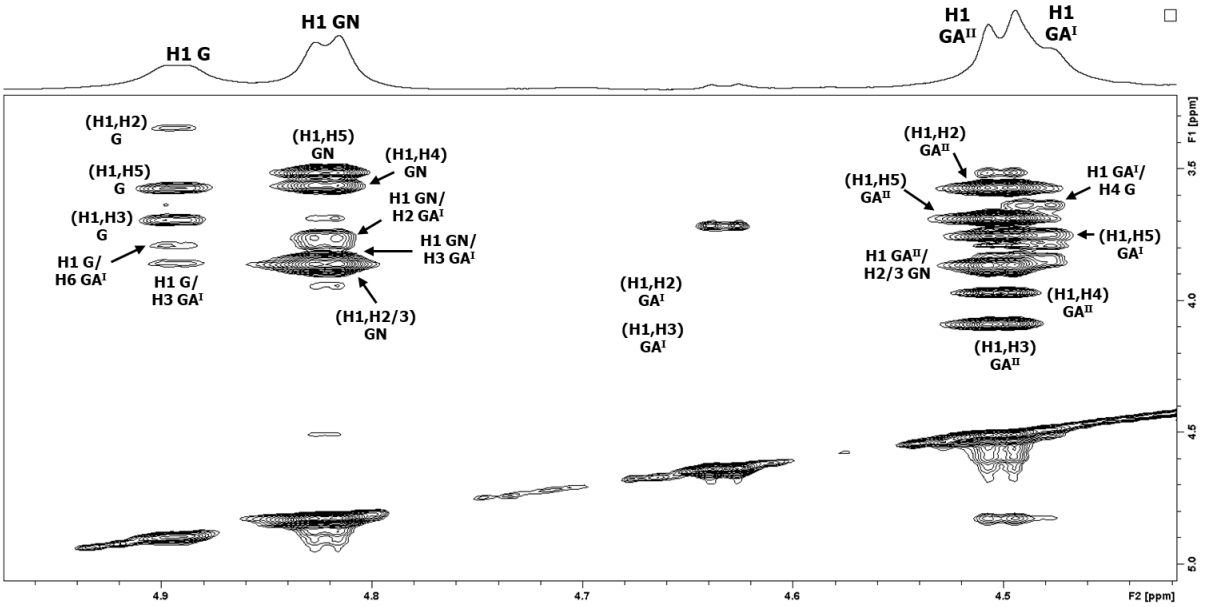


Figure 4.19: Anomeric region of the 2D NOESY spectrum of the GBS Ib CPS with intra- and inter-residue correlations for G, GN, GA^{I} and GA^{II} labelled.

The 1D NOESY of H1 GA^{I} and GA^{II} is ambiguous, since both signals are irradiated at once (Figure 4.13c), but the 2D NOESY (Figure 4.19) helped deconvolute the assignments, much like the 1D and 2D TOCSY spectra. Intra-residue correlations are seen from H1 to H5 GA^{I} at 3.74 ppm and from H1 to H5 GA^{II} at 3.68 ppm, allowing their assignment. Inter-residue correlations are seen from H1 GA^{II} to H2/3 GN, consistent with a GA^{II}-(1→3)-GN linkage, while a weak correlation can be seen from H1 GA^{I} to H4 G, which is evidence of a GA^{I}-(1→4)-G linkage.

Lastly, the sialic acid spin system shows TOCSY correlations (Figure 4.20b) from the H3ax/eq signals at 1.79 and 2.77 ppm to signals at 3.61, 3.68, 3.72 and 3.84 ppm which were assigned as H7, H6, H4 and H5 respectively.

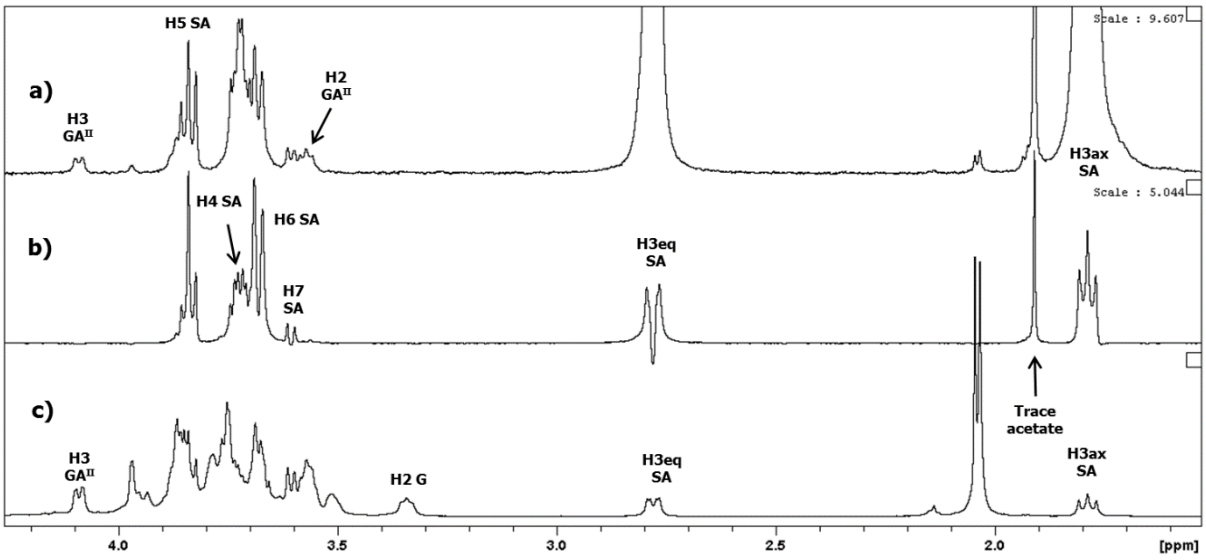


Figure 4.20: Labelled 1D NOESY (a) and TOCSY (b) experiments wherein the H3ax (a) proton of SA was selectively irradiated overlaid onto the DOSY spectrum (c) of the GBS Ib CPS.

This leaves only H6 of GA^{I} and GA^{II}, as well as H8 and H9 of SA unassigned. The H3ax/eq SA protons (Figure 4.20a) also show NOESY correlations to H2 and H3 GA^{II}, confirming the SA-(2→3)-GA^{II} linkage.

4.2.3. Heteronuclear experiments (HSQC, HMBC and hybrid)

As previously, the full ^{13}C NMR assignment and confirmation of all proton assignments was performed through the combined use of HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC experiments. Figure 4.21 shows the full assignment and confirmation of the GlcNAc spin system by a combination of HSQC, HSQC-TOCSY and selective 1D TOCSY as an example.

The HSQC-TOCSY signals arranged vertically above the anomeric signal show correlations between the anomeric proton and the ring carbons in the GN residue, allowing for the carbon assignments, especially when cross-referenced with the proton assignments shown by the selective 1D TOCSY overlayed over the bottom of the spectrum.

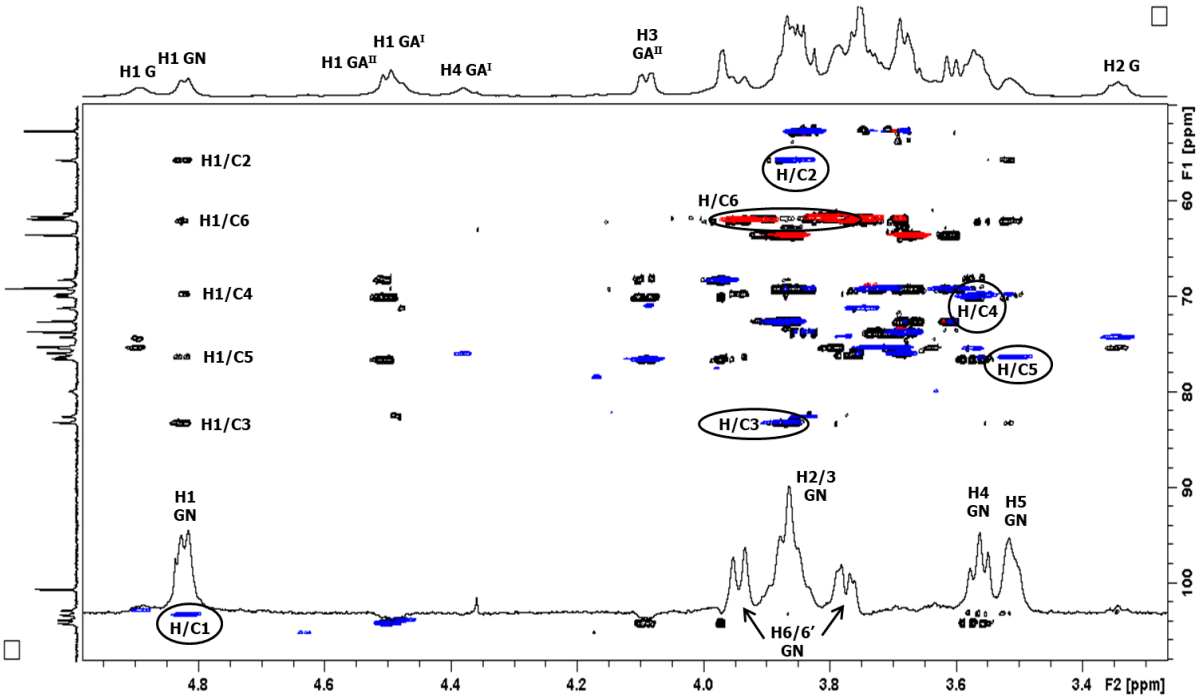


Figure 4.21: Anomeric and ring regions of the selective 1D TOCSY of H1 GN (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS Ib CPS, highlighting correlations that show the proton and carbon assignments of the entire GlcNAc spin system.

Using the HSQC-DEPT and HSQC-TOCSY experiment overlay (Figure 4.22), H8 and H9/9' of SA were assigned as 3.87 and 3.67/3.87 ppm respectively, while the H6 signals of GA^{I} and GA^{II} were assigned as the final overlapped signal at approximately 3.75 ppm since the data does not provide a means to tell them apart.

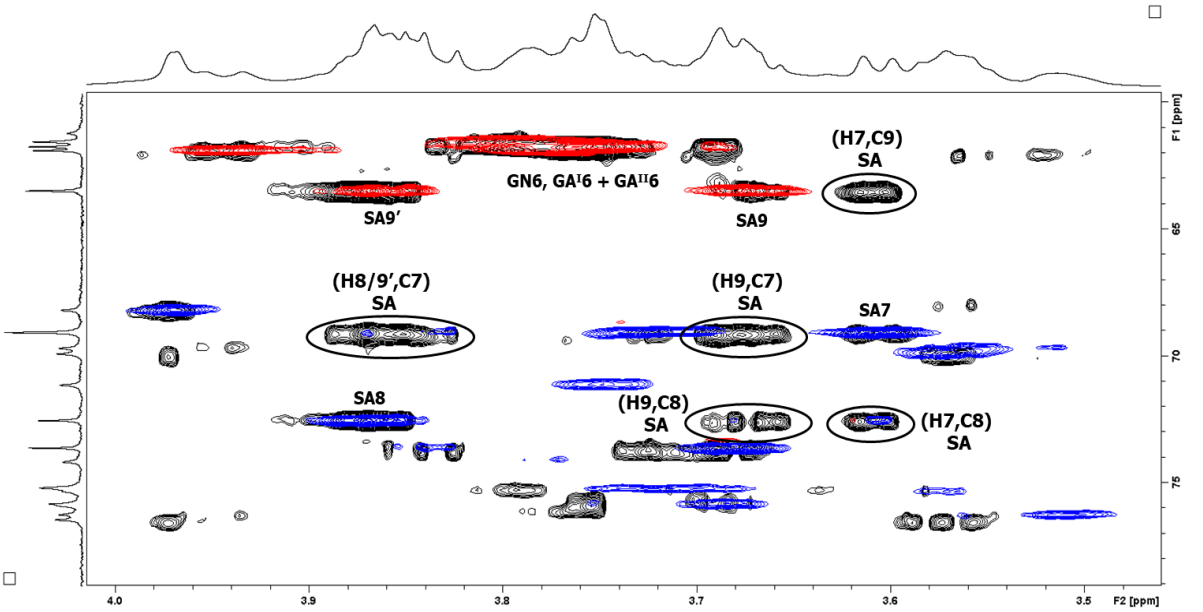


Figure 4.22: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS Ib CPS, highlighting correlations that led to the assignment of H8 and H9 of SA.

The HMBC (Figure 4.23) spectrum allowed full confirmation of the repeating unit connectivity shown by the NOESY as well as the assignment of the quaternary carbons, by showing intra- and inter-residue correlations through 3-bond coupling. Only three of the five linkages in the polysaccharide, namely GA^{II}-(1→3)-GN, GA^{I}-(1→4)-G and GN-(1→)-GA^{I}, could be confirmed by correlations between the anomeric protons and linkage site carbons in Figure 4.23, with an additional linkage, namely SA-(2→3)-GA^{II}, confirmed by the reverse correlation from the linkage ring proton to the anomeric carbon. Unfortunately, like with the GBS Ia CPS, the low intensity of the Glc residue signals, likely due to restricted movement in the polysaccharide backbone, results in low-intensity 1D and 2D crosspeaks with signals from that residue, so the G-(1→4)-GA^{I} linkage could not be confirmed by HMBC.

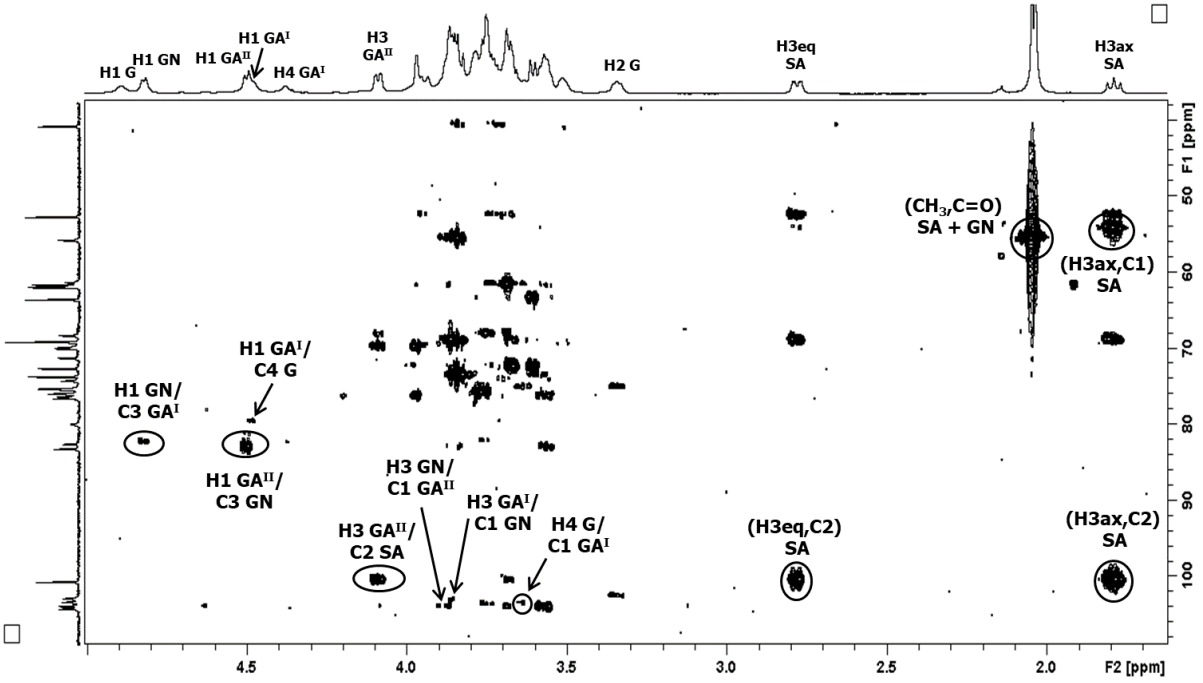


Figure 4.23: The HMBC spectrum of the GBS Ib CPS with important intra- and inter-residue linkage signals labelled.

Lastly, HMBC was used to assign the quaternary carbons such as C1 and C2 of SA as well as the carbonyl carbons in the N-acetyl groups of SA and GN. The carbon assignment of C2 SA (100.56 ppm) was relatively simple, showing intra-residue correlations to H3eq/ax, as well as an inter-residue correlation to H3 GA^{II}, which was also used to confirm the SA-(2→3)-GA^{II} linkage. The carbonyl carbons of the N-acetyl groups could be seen by a strong correlation to the CH_3 signal at approximately 55 ppm. Since carbonyl carbons are extremely deshielded and the spectral width of the HMBC experiment was only 120 ppm, these were taken as folded

peaks and thus assigned at 175.33 and 175.73 ppm for the GN and SA carbonyl carbons respectively. Similarly, there is a folded correlation between H3ax SA and the carboxyl carbon, C1 SA, that was assigned as 174.47 ppm.

Unfortunately, unlike for GBS type Ia, an HSQC-NOESY experiment did not provide any additional correlations allowing the confirmation one of the linkages missing in the HMBC, in this case G-(1→4)-GA^{I}. This means it could only be confirmed in the 1D and 2D homonuclear NOESY experiments.

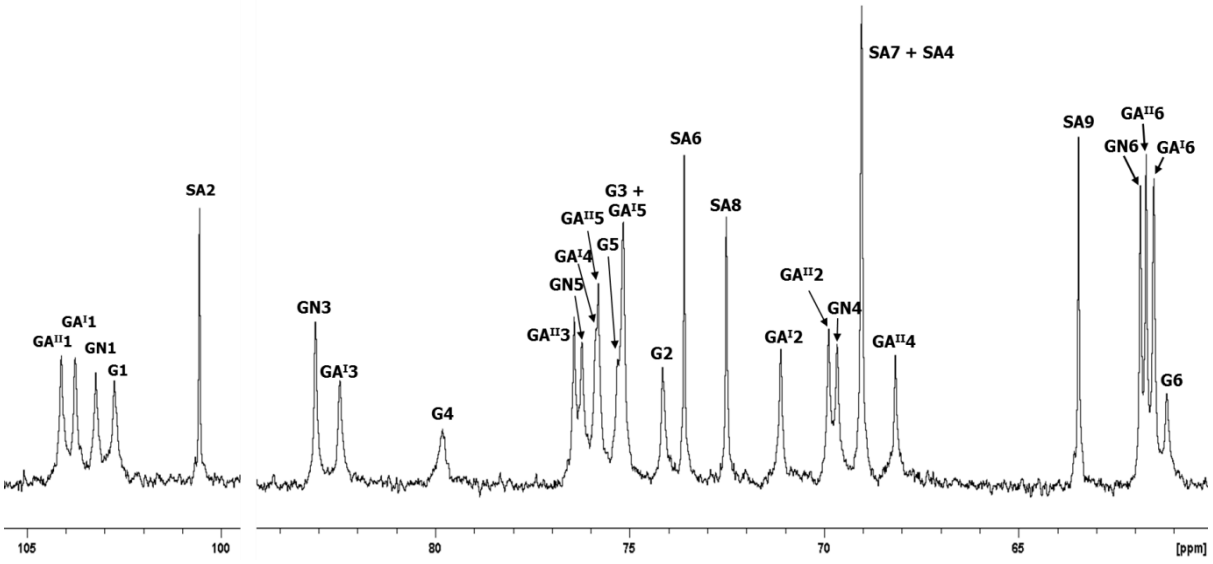


Figure 4.24: Expansion of the anomeric and ring regions of the labelled 1D ^{13}C NMR spectrum of the GBS Ib CPS.

Once all carbons have been assigned by the combination of HSQC, HSQC-TOCSY and HMBC experiments, the ^{13}C spectrum (Figure 4.24) can be fully labelled.

4.2.4 Full assignment and fingerprint summary

With all assignments made, the HSQC-DEPT spectrum serves as a 2D identity map (Figure 4.25) showing the chemical shift of every proton/carbon correlation in the GBS Ib CPS repeating unit. The chemical shift data is collected in Table 4.2, with calculated glycosylation shifts further confirming the established linkage positions. These assignments agree with previously published anomeric chemical shift values (recorded at 298 K) [83], with slight differences in chemical shift attributed to the difference in experiment temperature. The 1D ^{1}H and ^{13}C spectra can also be used as simpler fingerprinting methods wherein the identification of a few diagnostic peaks such as the anomerics, H4 GA^{I}, H3 GA^{II}, H2 G and H3ax/eq SA are sufficient to prove polysaccharide identity.

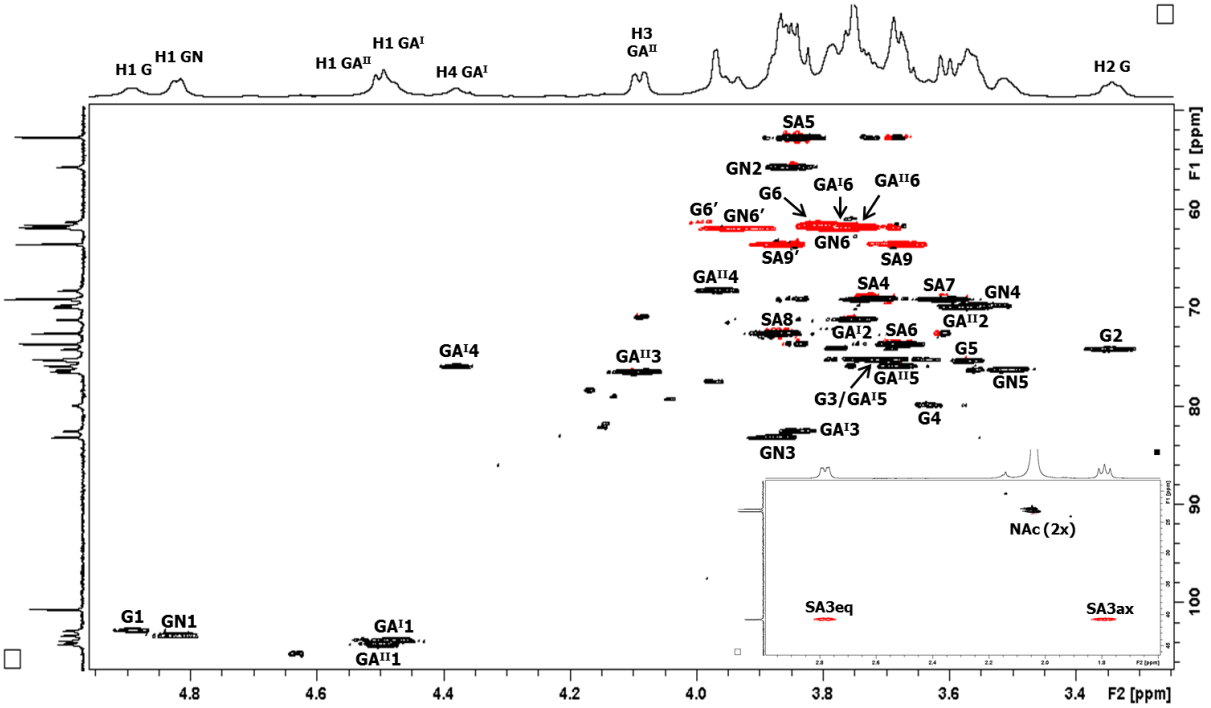


Figure 4.25: Fully labelled HSQC-DEPT spectrum of the GBS Ib CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

Table 4.2: Full NMR assignment of the GBS Ib CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→4)-β-D-Glc*p*-(1→	4.89	3.34	3.70	3.63	3.58	3.82/ 3.98	-	-	-
G	102.75	74.15	75.17	<u>79.81</u>	75.31	61.20	-	-	-
	(5.91)	(-1.05)	(-1.59)	(9.10)	(-1.45)	(-0.64)			
→3)-β-D-Glc*p*NAc-(1→	4.82	3.85	3.86	3.56	3.51	3.77/3.94	-	2.04	-
GN	103.23	55.62	<u>83.08</u>	69.67	76.23	61.87	175.33	23.20	-
	(7.38)	(-2.24)	(8.27)	(-1.39)	(-0.59)	(0.02)			
→3,4)-β-D-Gal*p*-(1→	4.48	3.75	3.84	4.38	3.74	3.79	-	-	-
GA^{I}	103.76	71.12	<u>82.46</u>	<u>75.87</u>	75.17	61.53	-	-	-
	(6.39)	(-1.84)	(8.70)	(6.18)	(-0.76)	(-0.31)			
→3)-β-D-Gal*p*-(1→	4.50	3.57	4.09	3.97	3.68	3.75	-	-	-
GA^{II}	104.12	69.89	<u>76.43</u>	68.17	75.81	61.72	-	-	-
	(6.75)	(-3.07)	(2.65)	(-1.52)	(-0.12)	(-0.12)			
α-D-Neu*p*NAc-(2→	-	-	1.79/ 2.78	3.72	3.84	3.68	3.61	3.87	3.67/ 3.87
SA[a]	174.37	<u>100.56</u>	40.68	69.04	52.59	73.60	69.04	72.52	63.46
			(-1.00)	(-0.38)	(0.15)	(0.17)	(-0.38)	(0.13)	(-0.61)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.05, 22.80 ppm, C=O at 175.73 ppm.

4.3 NMR spectral comparison of GBS Ia and Ib CPSs

Since the GBS Ib CPS differs from GBS Ia by one linkage position, Ia having a GA^{II}-(1→4)-GN linkage, while Ib has a GA^{II}-(1→3)-GN linkage, they can structurally seem very similar and indeed they have similar NMR profiles, but there are key differences that allow NMR to nevertheless distinguing between Ia and Ib.

In the ^{1}H NMR (Figure 4.26), the most noticeable differences of GBS Ib as compared to GBS Ia are the significant upfield shifts of H1 GA^{II} and H5 GN, the different ring region profile as well as the separation of the two methyl signals at approximately 2 ppm. These differences in magnetic environment and thus chemical shift are likely due to the different Gal-GlcNAc connectivity of GBS Ib causing the CPS repeating unit to adopt a different conformation than that of Ia.

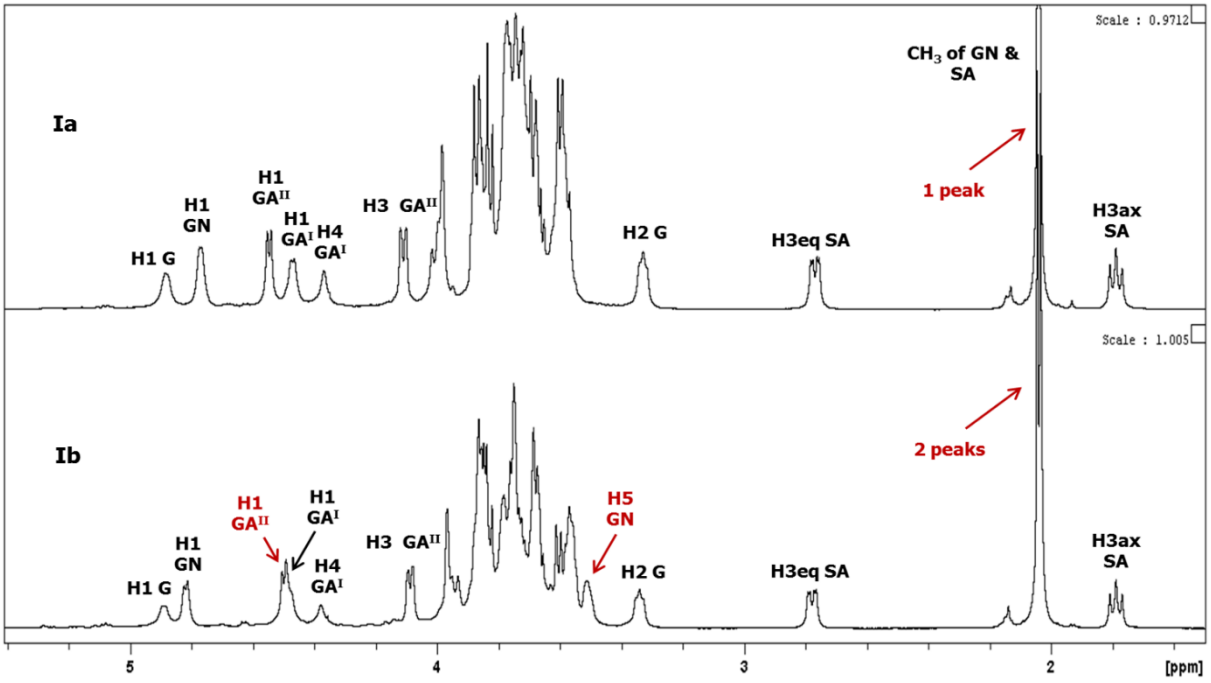


Figure 4.26: Overlay of the ^{1}H NMR spectra of GBS Ia and Ib CPSs with diagnostic peaks labelled and key differences highlighted in red.

Further differences can be noted in the 2D HSQC spectra (Figure 4.27), understandably most prominent in the GN chemical shifts. Since the difference between Ia and Ib is the change from a linkage at H4 GN to H3 GN, those two signals show the greatest change, with GN3 of GBS Ib showing a significant downfield shift for both its ^{1}H and ^{13}C signals and GN4 showing the opposite, a significant upfield shift of both signals. Most of the other GN signals (GN2, 5 and 6') also show noticeable differences.

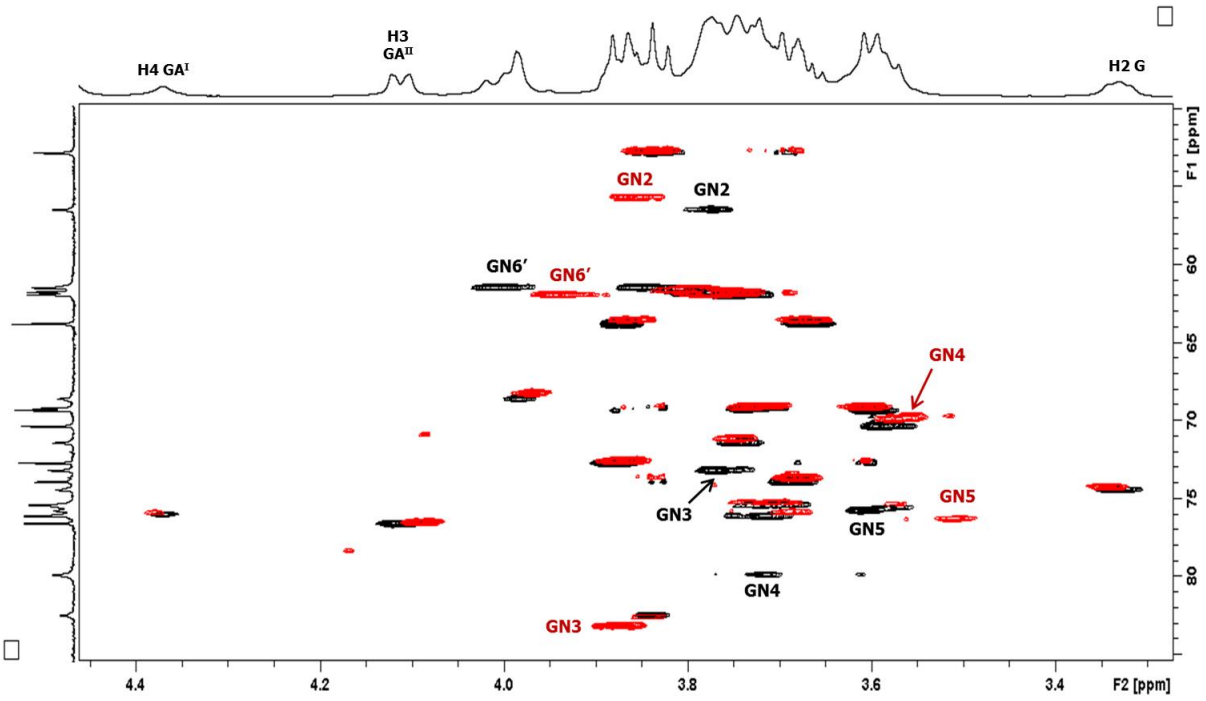


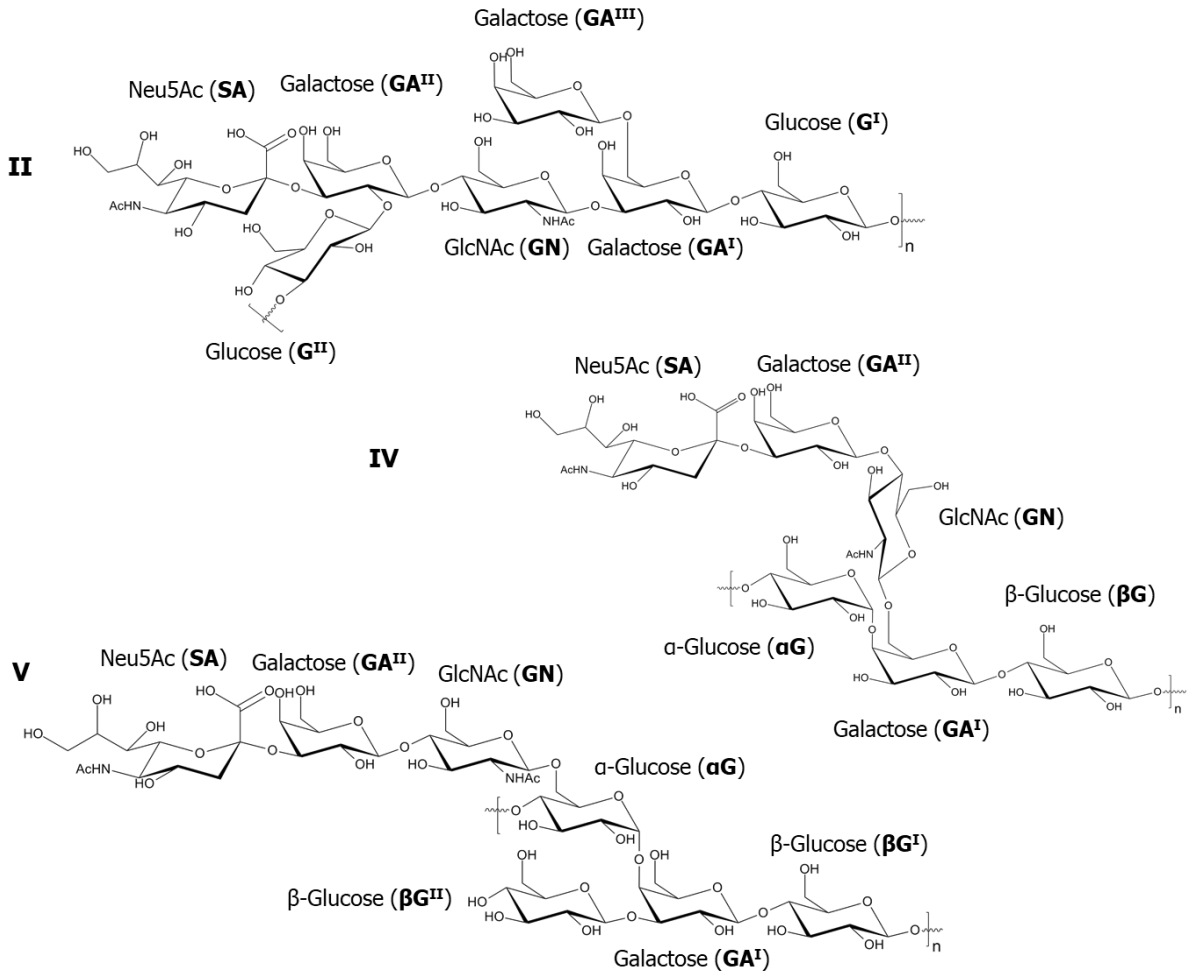
Figure 4.27: Overlaid expansion of the ring region of the 2D HSQC spectra of GBS Ia (black) and Ib (red) CPSs with significantly different peaks labelled.

Thus, while the GBS Ia and Ib capsular polysaccharides are structurally very similar, differing by only one linkage position, NMR data provides a way to easily tell them apart in both the ^{1}H and ^{13}C spectra.

Chapter 5. NMR characterisation of GBS II, IV and V

The three remaining serotypes assigned in this study, GBS types II, IV and V are less prevalent than types Ia, Ib and III, cumulatively only accounting for 12% of global GBS disease cases [7], but they are still important targets for the development of a multivalent vaccine, due to the large variation in GBS serotype distribution both geographically and demographically. For example, serotypes IV and V are the dominant serotypes in the United Arab Emirates and Egypt respectively, while GBS V is also one of the serotypes most responsible for invasive disease in non-pregnant adults, as discussed in Chapter 1. The structures of their capsular polysaccharides were determined between 1983 and 1991, using chemical methods supported either by simple ^{13}C NMR [131] or high-resolution NMR [133,134]. The structural determination demonstrated that the GBS II, IV and V CPSs have either hexa- or heptasaccharide repeating units containing D-Galactose (Gal), D-Glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc) and N-Acetylneuraminic acid (Neu5Ac or sialic acid) in molar ratios of 3:2:1:1, 2:2:1:1 and 2:3:1:1 respectively (Figure 5.1).

As with GBS Ia and Ib, less research has been conducted on serotypes II, IV and V than has been done on serotype III. For GBS II, NMR data has been published for a monomer of the CPS [144] as well as the ^{1}H spectra and anomeric assignments at 298 K of the polysaccharide [83], but NMR data for the full polysaccharide has yet to published. Since the initial structural elucidation of the GBS IV CPS was supported with high-resolution NMR, data is available on the anomeric and backbone residue assignments at 310 K [133], as well as ^{1}H spectra and anomeric assignments at 298 K [83]. Lastly, for GBS V, NMR data for a monomer of the CPS has been published [151], as well as data on the anomeric and backbone residue signals at 300 K [134] and ^{1}H spectra and anomeric assignments at 298 K [83,156]. Thus, full NMR characterisations have so far not been published for the polysaccharide repeating units of any of these aforementioned serotypes.



II: →3)-β-D-Glc*p*-(1→2)-[α-D-Neu*p*NAc-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→3)-[β-D-Gal*p*-(1→6)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

IV: →4)-α-D-Glc*p*-(1→4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

V: →4)-[α-D-Neu*p*NAc-(2→3)-β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→6)]-α-D-Glc*p*-(1→4)-[β-D-Glc*p*-(1→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-(1→

Figure 5.1: Published repeating unit structures and line structures of the GBS II, IV and V capsular polysaccharides.

5.1 Structural characterisation of the GBS II, IV and V CPSs

The methodology established for the NMR characterisation described in Chapters 3 and 4 was also applied to serotypes II, IV and V, however unlike in the previous chapters all three elucidations will be discussed together to avoid undue repetition.

5.1.1. 1D ^{1}H and DOSY experiments

The 1D ^{1}H and DOSY (Figure 5.2) spectra allowed the assignment of the anomeric and other diagnostic signals from the literature [83].

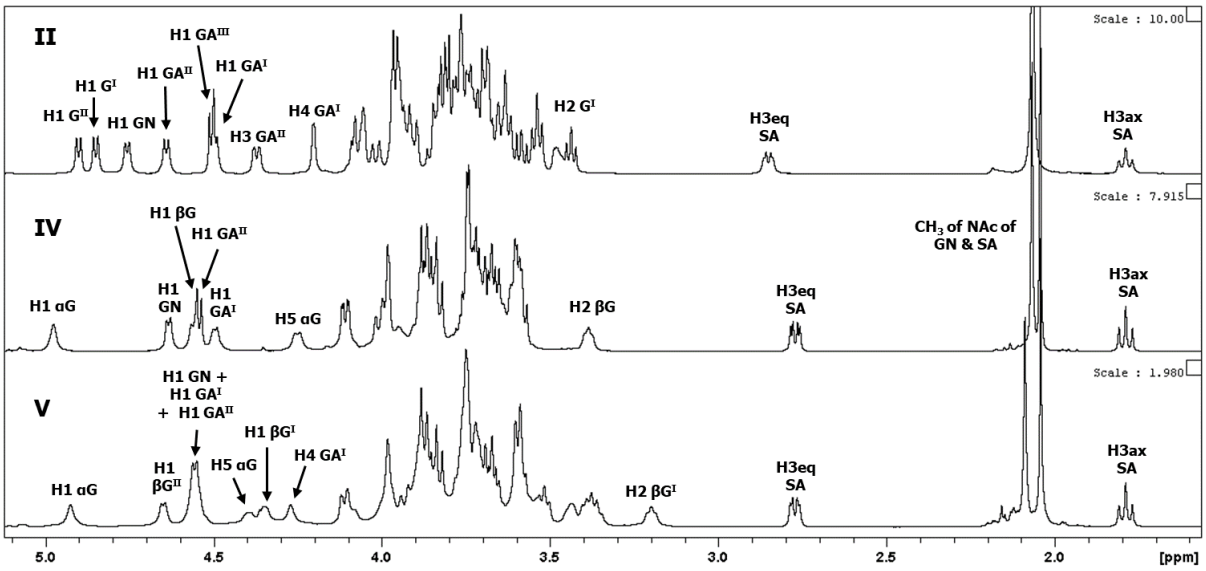


Figure 5.2: Overlay of 1D DOSY spectra of the GBS II, IV and V CPSs recorded at 343 K with diagnostic peaks labelled.

The number of anomeric signals between 4.3 and 5.0 ppm, together with the H3 axial/equatorial sialic acid signals at 1.79 and approximately 2.8 ppm correspond to a hexasaccharide repeating unit for GBS IV and heptasaccharide repeating units for GBS II and V as published.

Starting with GBS II, the four downfield anomeric peaks at 4.90, 4.85, 4.76 and 4.64 were assigned as H1 G^{II}, G^{I}, GN, and GA^{II} respectively, while the large signal at approximately 4.5 ppm was assigned as overlapped H1 GA^{III} (4.51 ppm) and GA^{II} (4.50 ppm). Additional peaks that could assigned in the GBS II ^{1}H spectrum include H3 GA^{II}, H4 GA^{I} and H2 G^{I} in the ring region (4.37, 4.20 and 3.44 ppm respectively) as well as H3eq/ax SA and the overlapped methyl signals from the N-acetyl groups of GN and SA in the upfield region (2.85/1.79 and 2.07 ppm respectively).

For GBS IV, the anomeric peaks at 4.98, 4.64, 4.56, 4.54 and 4.50 ppm were assigned as H1 αG, GN, βG, GA^{II} and GA^{I} respectively, with the βG and GA^{II} signals overlapped. Additional GBS IV peaks that could be assigned were H5 αG, H2 βG in the ring region (4.25 and 3.39 ppm respectively) as well as H3eq/ax SA and the methyl signals from the N-acetyl groups of GN and SA in the upfield region (2.77/1.79 and 2.07/2.04 ppm respectively).

Lastly, GBS V presented some difficulties due to the fact that three of the anomeric peaks are overlapped and it was determined over the course of the structural characterisation that some of the anomeric peaks were incorrectly assigned in the literature [83,134]. Ultimately,

the two downfield anomeric peaks at 4.93 and 4.65 ppm were assigned as H1 αG and βG^{II} respectively, while the large signal at 4.55 ppm was assigned as the overlapped signals of H1 GN, GA^{I} and GA^{II}. Lastly, the anomeric signal of βG^{I} is unusually shielded, appearing at 4.35 ppm, an assignment that will be supported by correlation experiments described later in this chapter. Additional GBS V peaks that could be assigned in the ^{1}H spectrum are H5 αG, H4 GA^{I} and H2 βG^{I} in the ring region (4.39, 4.27 and 3.39 ppm respectively) as well as H3eq/ax SA and the methyl signals from the N-acetyl groups of GN and SA in the upfield region (2.77/1.79 and 2.09/2.04 ppm respectively).

The fact that GBS IV and V each have an anomeric peak that resonates above 4.9 ppm is consistent with the evaluation that they each have one α-linked sugar unit (αG) in addition to sialic acid in the repeating unit, while GBS II exclusively consists of β-linked hexoses and sialic acid.

5.1.2. Homonuclear correlation experiments (COSY, TOCSY and NOESY)

For all three serotypes the 2D homonuclear correlation experiments provided the majority of the preliminary proton assignments by way of correlation with the identified diagnostic peaks, allowing elucidation of most of the spin system signals.

5.1.2.1. GBS II COSY/TOCSY

For GBS II, the COSY/TOCSY overlay (Figure 5.3) shows correlations from the most downfield anomeric proton (H1 G^{II}, 4.90 ppm) to signals at 3.48, 3.54, 3.78, 3.82 and 3.95 ppm which were assigned as H5, overlapped H2/4, H6, H3 and H6' respectively, completing the preliminary assignment of the 3-linked backbone Glc.

The anomeric proton of G^{I} shows COSY/TOCSY correlations to signals at 3.44, 3.68, 3.75, 3.83 and 4.02 ppm, assigned as H2, overlapped H4/5, H3, H6 and H6' respectively, preliminarily assigning the entire spin system of the 4-linked backbone Glc. The entire GlcNAc spin system was preliminarily assigned by correlations from H1 GN to signals at 3.63, 3.70, 3.80, 3.91 and 4.07 ppm. These signals were assigned as H5, H4, overlapped H2/3, H6 and H6' respectively. As expected for Gal (small $J_{H4,H5}$), the TOCSY experiment only shows correlations from H1 to H2, H3 and H4 for the three Gal residues. The anomeric proton of GA^{II} shows correlations to signals at 3.93 and 4.05 ppm which were assigned as H2 and H4, respectively. While it has already been assigned, a correlation to H3 from the anomeric signal cannot be seen vertically,

likely because H3 GA^{II} is overlapped with the water signal and its signal is thus eliminated by the presaturation that removes the water signal, but horizontally, the correlation to the anomeric signal from H3 (4.37 ppm) can still be seen.

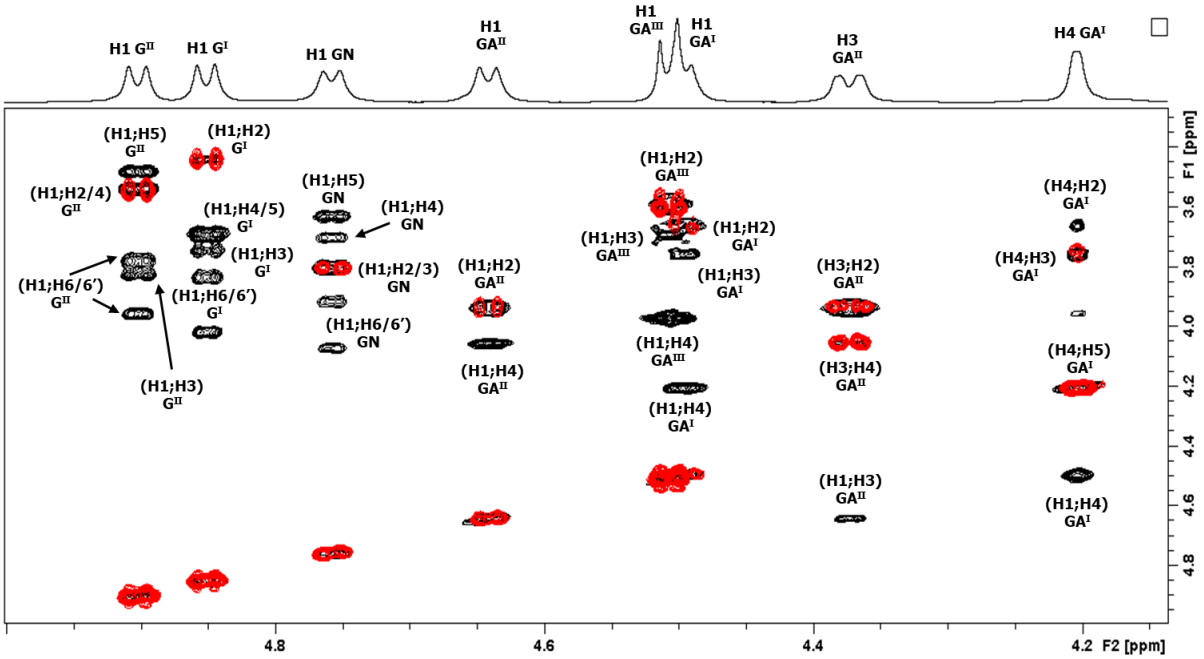


Figure 5.3: Anomeric region of the GBS II 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) with identified crosspeaks for G^{II}, G^{I}, GN, GA^{II}, GA^{III} and GA^{I} labelled.

Even though H1 GA^{III} and H1 GA^{I} are overlapped, the 2D TOCSY spectrum clearly shows the two separate spin systems and thus the three signals at 3.59, 3.69 and 3.97 ppm could be assigned as H2, H4 and H3 GA^{III} respectively, while the three slightly upfield signals at 3.66, 3.76 and 4.20 ppm were assigned as H2, H3 and H4 GA^{I} respectively. This left H5 and H6 unassigned for all three Gal residues of GBS II.

5.1.2.2. GBS IV COSY/TOCSY

Moving on to GBS IV (Figure 5.4), the most deshielded anomeric proton signal belongs to the αG residue and it shows correlations to signals at 3.60, 3.68, 3.85 and 4.25 ppm, assigned as H2, H4, H3 and H5 respectively. Because the αG anomeric signal is relatively broad and of low intensity, correlations to H6 could not be seen from H1. However, TOCSY correlations from H5 αG, another diagnostic peak, to H6/6' allowed them to be preliminarily assigned to a peak near 3.85 ppm, later confirmed as 3.83 and 3.90 ppm respectively, completing the preliminary assignment of the α-Glc residue.

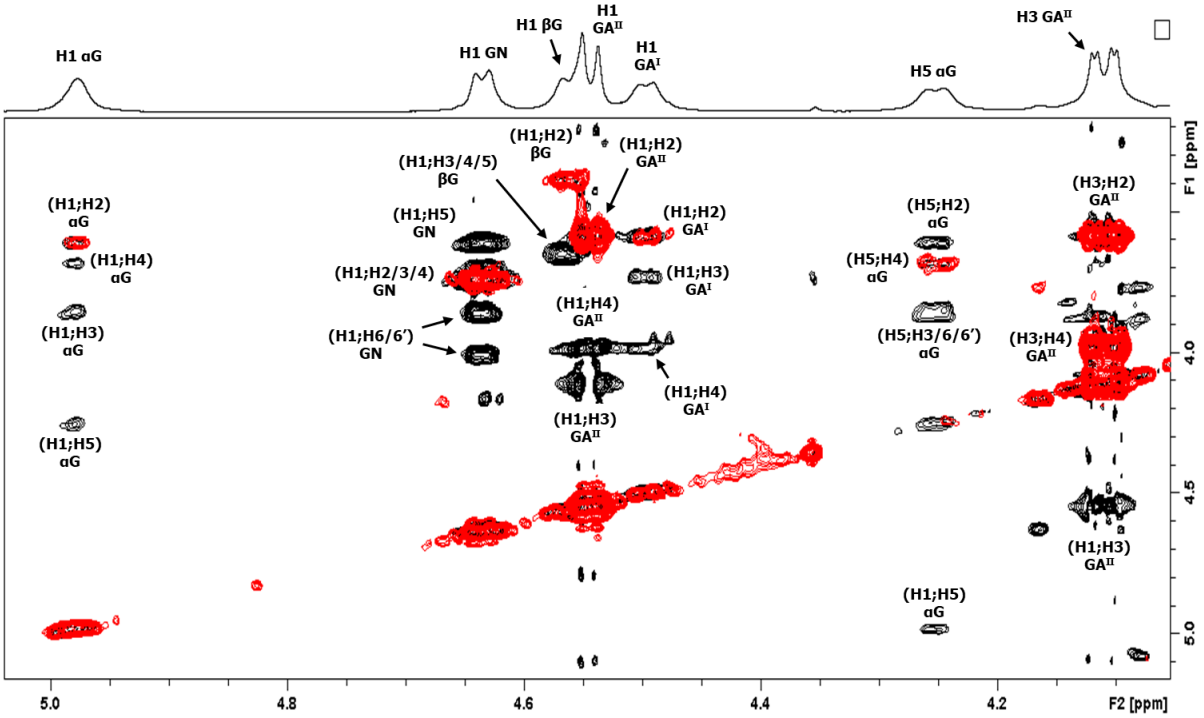


Figure 5.4: Anomeric region of the GBS IV 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) with identified crosspeaks for αG, GN, βG, GA^{II} and GA^{I} labelled.

The anomeric proton of the GlcNAc residue displays correlations to a signal at 3.61 ppm, a very large signal at 3.74 ppm and a pair of signals at 3.86 and 4.00 ppm. These were assigned as H5, overlapped H2/3/4, H6 and H6' respectively, completing the preliminary assignment of the GlcNAc residue. The anomeric protons of βG and GA^{II} are overlapped, but once again the 2D TOCSY allows the two spin systems to be separated. The signal at 3.39 ppm had already been assigned as H2 βG, while the large signal at 3.66 ppm was assigned as the overlapped H3, H4 and H5 βG signals. Similarly, the correlation from H1 GA^{II} to 4.11 ppm had already been assigned as H3 GA^{II}, so the remaining two signals at 3.58 and 3.98 ppm were assigned as H2 and H4 GA^{II} respectively. Finally, the anomeric proton of GA^{I} shows correlations to three signals at 3.59, 3.73 and 3.98 ppm which were assigned as H2, H3 and H4 respectively. This left H6 of βG, as well as H5 and H6 of the two Gal residues of GBS IV unassigned.

5.1.2.3. GBS V COSY/TOCSY

Lastly, for GBS V (Figure 5.5), the anomeric proton of αG displays correlations to signals at 3.55, 3.62, 3.89 and 4.39 ppm which were assigned as H2, H4, H3 and H5 respectively. Unfortunately, unlike GBS IV, no correlations to H6 could be seen from the diagnostic H5 αG signal. The anomeric signal of $βG^{II}$ shows correlations to signals at 3.20, 3.36, 3.44, 3.52, 3.73 and 3.93 ppm, assigned as H2, H4, H5, H3, H6 and H6' respectively. This completes the full

preliminary assignment of the terminal β-Glc residue. The large signal at approximately 4.55 ppm, that had been assigned as the overlapped anomeric signals of GN, GA^{I} and GA^{II}, proved difficult to deconvolute, but the 2D TOCSY does allow some differentiation between the three overlapped spin systems.

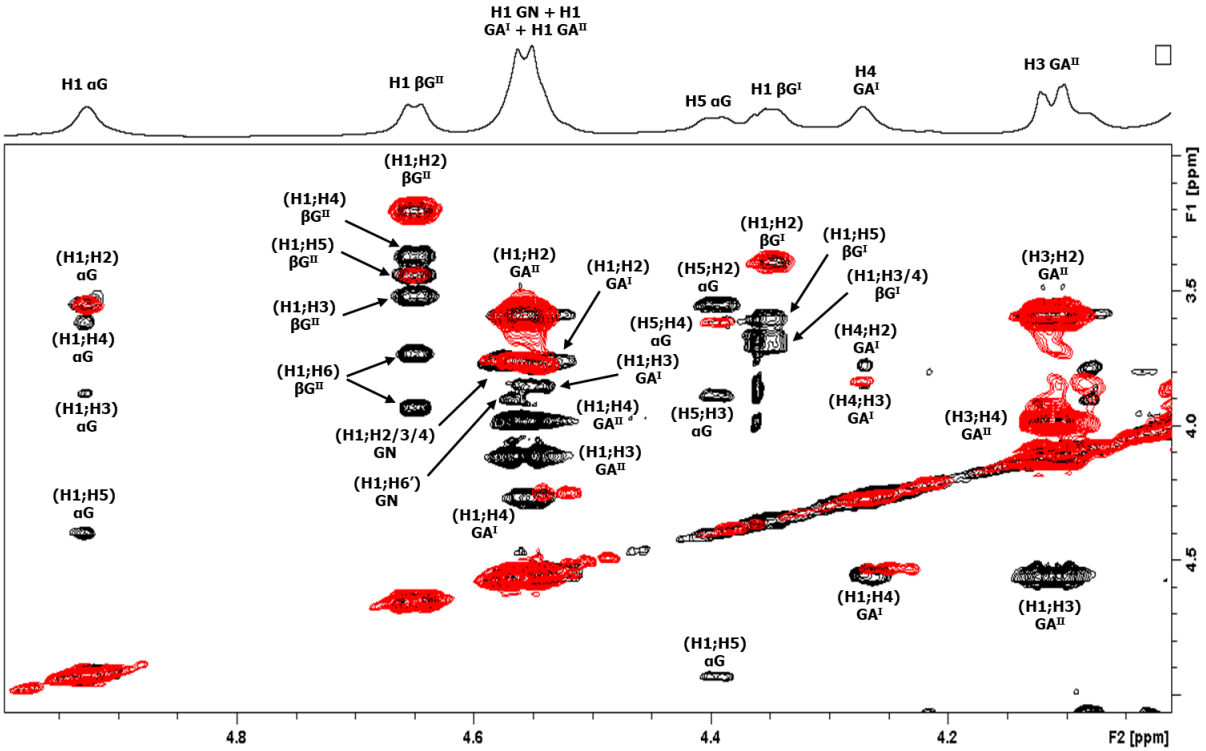


Figure 5.5: Anomeric region of the GBS V 2D COSY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) with identified crosspeaks for αG, βG^{II}, GN, GA^{I}, GA^{II} and βG^{I} labelled.

The large correlation at 3.75 ppm was assigned as the overlapped correlation to H2, H3 and H4 of GN, while the small signal at 3.90 ppm was assigned as one of the H6 protons of GN. This completes the preliminary assignment of the GlcNAc residue. The three signals at 3.77, 3.85 and 4.27 ppm were assigned to the GA^{I} spin system as H2, H3 and H4. The three largest signals, at 3.59, 3.98 and 4.11 ppm, were assigned to the GA^{II} spin system as H2, H4 and H3. Lastly, the significantly upfield anomeric signal of βG^{I} shows three correlations at 3.39, 3.60 and approximately 3.68 ppm which were assigned as H2, H5 and overlapped H3 and H4 respectively. This left H5 of GN, H6 of both αG and βG^{I}, as well as H5 and H6 of the two Gal residues of GBS V unassigned.

5.1.2.4. GBS II NOESY

Progressing to the NOESY spectra and beginning with GBS II (Figure 5.6), some of the as yet unassigned peaks could be identified through intra-residue correlations such as H5 GA^{II} that

could be identified with a correlation to H1 GA^{II} at 3.74 ppm, H5 GA^{I} identified through a correlation to H1 GA^{I} at 3.95 ppm and H6 GA^{I} that was identified with a correlation to H4 GA^{I} at 4.08 ppm. Furthermore, most of the connectivities between spin systems could be confirmed by inter-residue correlations from the anomeric signals. The G^{II}-(1→2)-GA^{II} linkage could be confirmed by a signal from the G^{II} anomeric proton to H2 GA^{II}, while the G^{I}-(1→3)-G^{II} linkage was confirmed by an inter-residue correlation from H1 G^{I} to H3 G^{II}. The GN-(1→3)-GA^{I} linkage was confirmed by a signal from H1 GN to H3 GA^{I} and the GA^{II}-(1→4)-GN linkage was confirmed by a correlation from the GA^{II} anomeric signal to H4 GN.

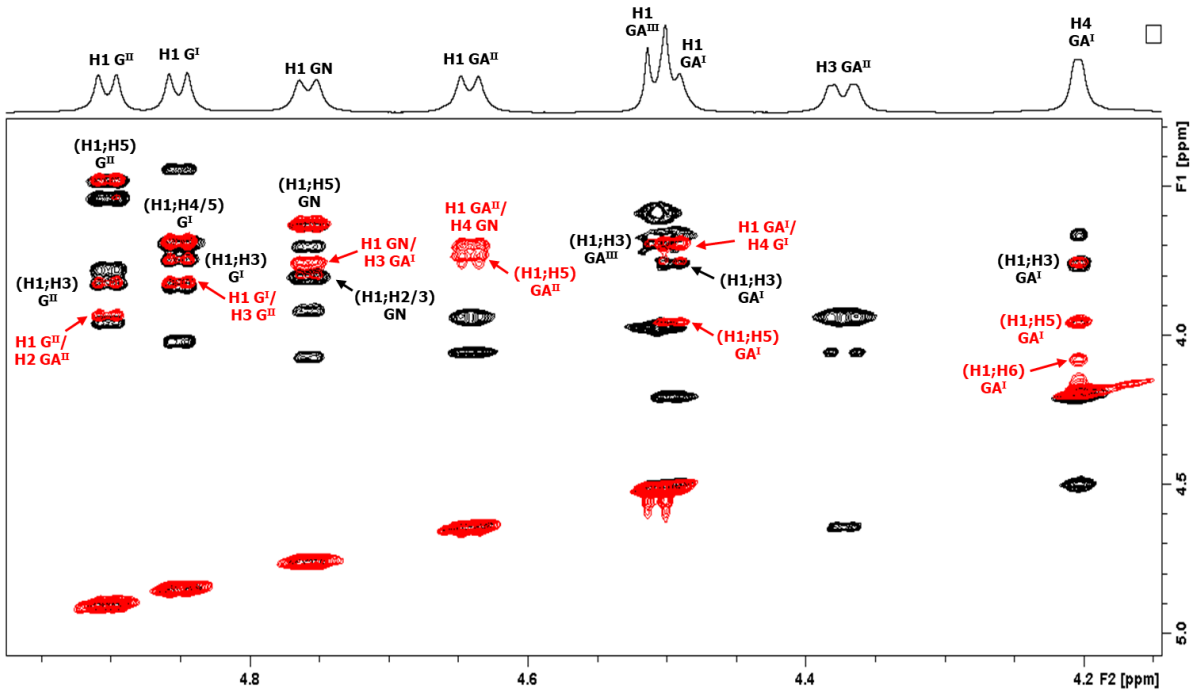


Figure 5.6: Anomeric region of the 2D NOESY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of the GBS II CPS with identified intra- and inter-residue NOESY correlations labelled. Correlations not already seen in through TOCSY are labelled in red.

The last GBS II linkage that could be confirmed through NOESY is the GA^{I}-(1→4)-G^{I} linkage, by a crosspeak from H1 GA^{I} to H4 G^{I}. This left the GA^{III}-(1→6)-GN and SA-(2→3)-GA^{II} linkages of the GBS II polysaccharide unconfirmed by NOESY. H5 and H6 of GA^{III}, H6 of GAII and the sialic acid spin system also remained unassigned.

5.1.2.5. GBS IV NOESY

Unassigned GBS IV (Figure 5.7) peaks that could be identified by intra-residue NOESY correlations include H5 GA^{II} that displays a correlation to H1 GA^{II} at 3.71 ppm that is however overlapped with the inter-residue correlation to H4 GN, making it a somewhat tenuous

confirmation. The H5 GA^{I} assignment was also achieved due to a correlation to H1 GA^{I} at 3.88 ppm. Most linkages in the polysaccharide repeating unit could be confirmed by inter-residue correlations. The αG-(1→4)-GA^{I} linkage could be established by correlations from H1 αG to both H4 and H6/6' GA^{I}, while the GN-(1→6)-GA^{I} linkage was confirmed by a weak correlation from H1 GN to H6 GA^{I}. Additionally, these correlations allowed the assignment of H6/6' GA^{I} (3.95/4.11 ppm). The βG-(1→4)-αG linkage could not be directly confirmed, since that correlation was directly overlapped with the large intra-residue correlations to H3 and H5 βG, but a correlation H3 αG at 3.85 ppm does provide evidence for the linkage. The correlation from H1 GA^{II} to H4 GN that is overlapped with the aforementioned H1 to H5 GA^{II} intra-residue correlation tenuously confirms the GA^{II}-(1→4)-GN linkage. Lastly the GA^{I}-(1→4)-βG linkage was confirmed by a correlation from H1 GA^{I} to H4 βG, leaving the SA-(2→3)-GA^{II} linkage as the only one not yet at least tentatively confirmed by NOESY. The H6 signals of GA^{II} and βG as well as the sialic acid spin system also remained unassigned.

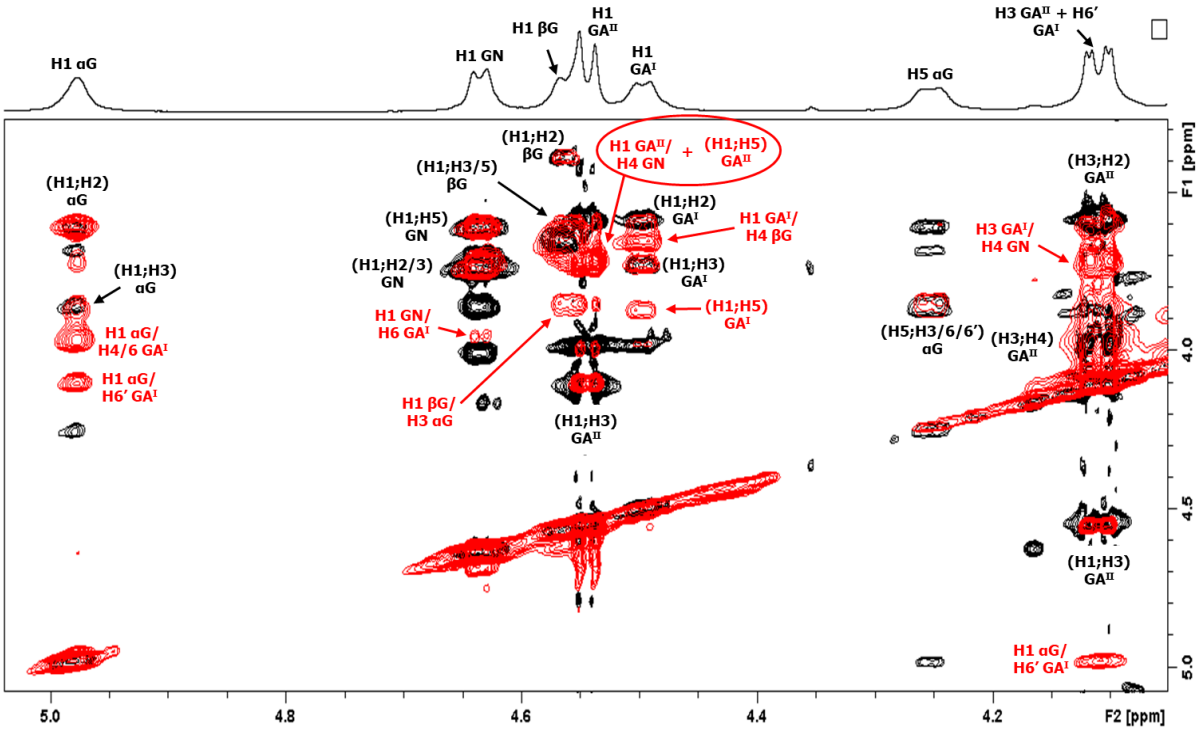


Figure 5.7: Anomeric region of the 2D NOESY spectrum (red) overlaid onto the 2D TOCSY spectrum (black) of the GBS IV CPS with identified intra- and inter-residue NOESY correlations labelled. Correlations not already established through TOCSY are labelled in red.

5.1.2.6. GBS V NOESY

Due to the extensive amount of overlap, the 2D NOESY spectrum of GBS V was quite difficult to deconvolute, but the selective 1D NOESY experiments (Figure 5.8) are comparatively clear.

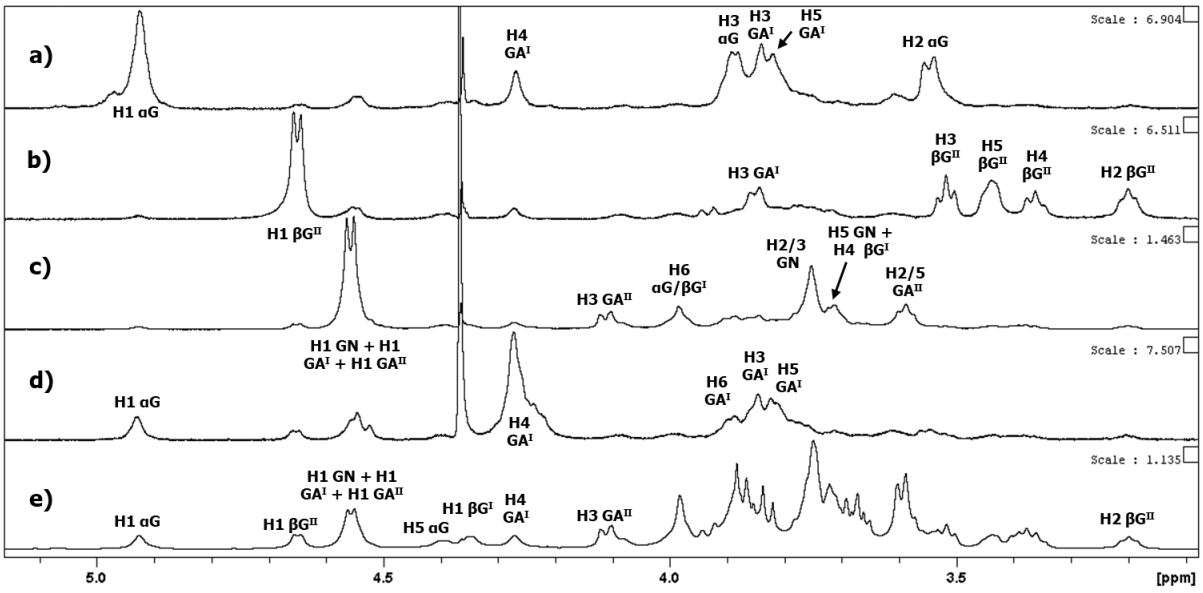


Figure 5.8: Labelled anomeric and ring regions of 1D NOESY experiments wherein H1 αG, H1 βG^{II}, overlapped H1 of GN/GA^{II}/GA^{I} and H4 GA^{I} (a-d respectively) were selectively irradiated and are overlaid onto the DOSY spectrum (e) of GBS V.

Selective irradiation of the anomeric signal of αG (Figure 5.8a) shows a number of inter-residue correlations to H3, H4 and H5 of GA^{I}, confirming the αG-(1→4)-GA^{I} linkage. A correlation from H1 βG^{II} to H3 GA^{I} (Figure 5.8b) confirms the βG^{II}-(1→3)-GA^{I} linkage. Because H1 GN, GA^{I} and GA^{II} are overlapped, they are all excited together during selective irradiation, meaning the 1D TOCSY (Figure 5.8c) shows all correlations for all three, but nevertheless it did still provide some useful information. There are two intra-residue correlations of interest, the first which is the one at 3.58 ppm which was determined to be overlapped H2 and H5 GA^{II}, tentatively assigning H5 GA^{II}. The second one, the peak at 3.72 ppm was determined to be overlapped H5 GN and H4 βG^{I}, due to a combination of intra- and inter-residue correlations from H1 to H5 GN, H1 GA^{I} to H4 βG^{I} and potentially also H1 GA^{II} to H5 GN, allowing the tentative assignment of H5 GN and providing the only NOESY evidence of the GA^{II}-(1→4)-GN linkage. Inter-residue correlations include a signal at 3.99 ppm that was assigned as overlapped signals from H1 GN to H6 αG and H1 GA^{I} to H6 βG^{I}, both tentatively assigning H6 αG and βG^{I} as well as confirming the GN-(1→6)-αG linkage. The combination of the aforementioned H1 GA^{I} to H4 βG^{I} and H1 GA^{I} to H6 βG^{I} correlations also allowed the confirmation of the GA^{I}-(1→4)-βG^{I} linkage. Lastly, in an attempt to gain more information about the GA^{I} spin system, the H4 GA^{I} signal was irradiated (Figure 5.8d), displaying intra-residue correlations to H5 and H6 GA^{I} at 3.81 and 3.89 ppm, allowing their assignment. This

left only the SA-(2→3)-GA^{II} linkage of the GBS V polysaccharide unconfirmed by NOESY. The H6 signals of GA^{II} as well as the sialic acid spin system also remained unassigned.

5.1.2.7. GBS II sialic acid spin system assignment

Finally, the TOCSY and NOESY analysis (Figure 5.9) of the sialic acid spin systems begins with GBS II.

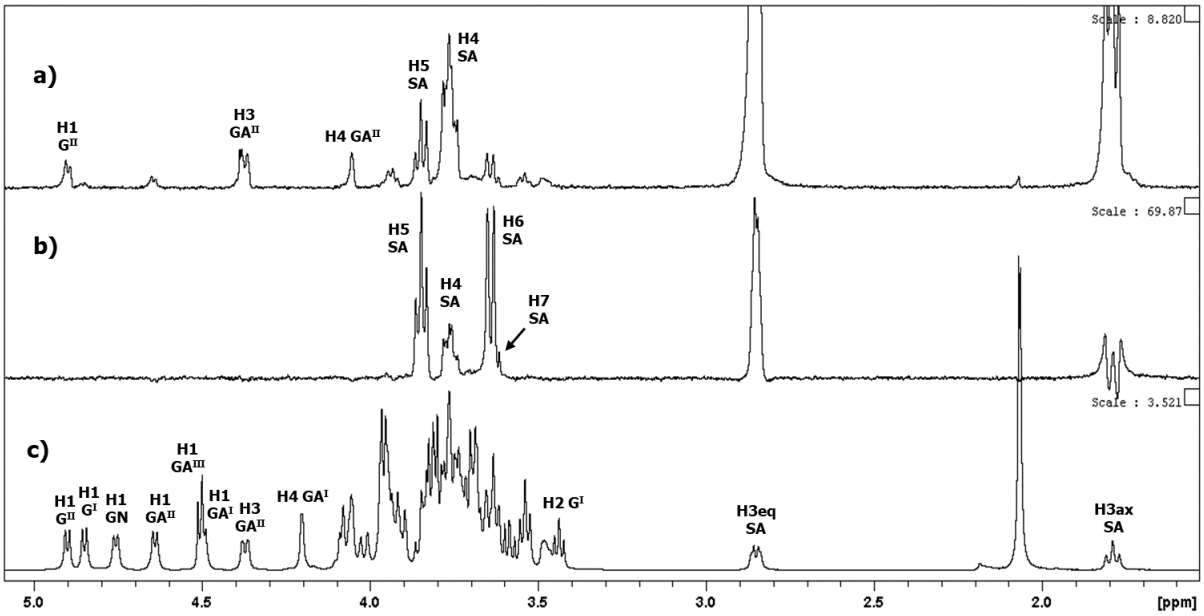


Figure 5.9: Labelled 1D NOESY (a) and TOCSY (b) experiments wherein the H3eq proton of SA were selectively irradiated overlaid onto the DOSY spectrum (c) of GBS II.

The 1D selective TOCSY (Figure 5.9b) shows clear correlations to signals at 3.85, 3.76 and 3.64 ppm, assigned as H5, H4 and H6 SA respectively. Additionally, a very small peak was observed at 3.62 ppm which was preliminarily assigned as H7 SA. The NOESY (Figure 5.9a) shows inter-residue correlations to both H3 and H4 GA^{II}, confirming the SA-(2→3)-GA^{II} linkage. An additional correlation can be seen to H1 G^{II}, which is understandable, since the SA and G^{II} residues are linked to adjacent carbons of GA^{II}, meaning it is likely they would come into close contact with each other.

5.1.2.8. GBS IV sialic acid spin system assignment

The 1D selective TOCSY of H3eq/ax SA in GBS IV (Figure 5.10b) shows much the same information as the one for GBS II, providing the assignments for H5, H4 and H6 of SA (3.84, 3.73 and 3.68 ppm respectively) and the tentative assignment of H7 SA at 3.60 ppm. The 1D NOESY, on the other hand only shows very weak intra-residue correlations, meaning the SA-(2→3)-GA^{II} linkage of GBS IV could not confirmed by NOESY.

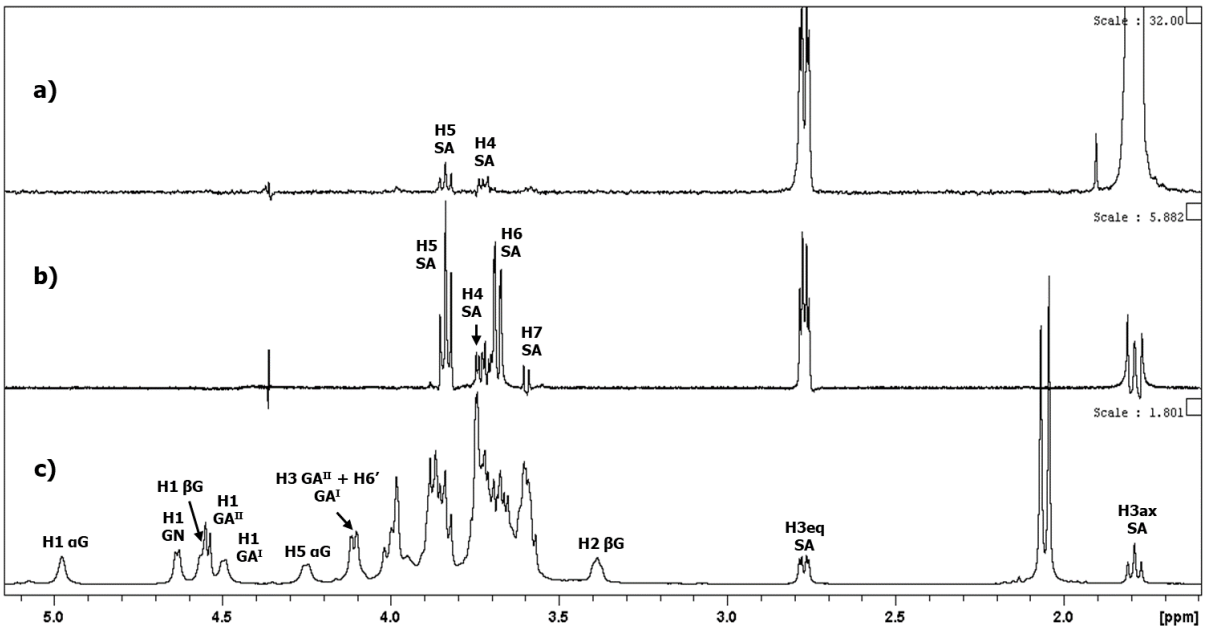


Figure 5.10: Labelled 1D NOESY (a) and TOCSY (b) experiments wherein the H3eq or H3ax proton of SA were selectively irradiated overlaid onto the DOSY spectrum (c) of GBS IV.

5.1.2.9. GBS V sialic acid spin system assignment

Lastly, the TOCSY and NOESY of H3eq/ax SA in GBS V (Figure 5.11) once again shows TOCSY (Figure 5.11b) correlations to H5, H4 and H6 SA (3.84, 3.72 and 3.68 ppm respectively) and a weak correlation to H7 SA at 3.60 ppm. The NOESY (Figure 5.11a) shows inter-residue correlations to both H3 and H2 GA^{II}, confirming the SA-(2→3)-GA^{II} linkage.

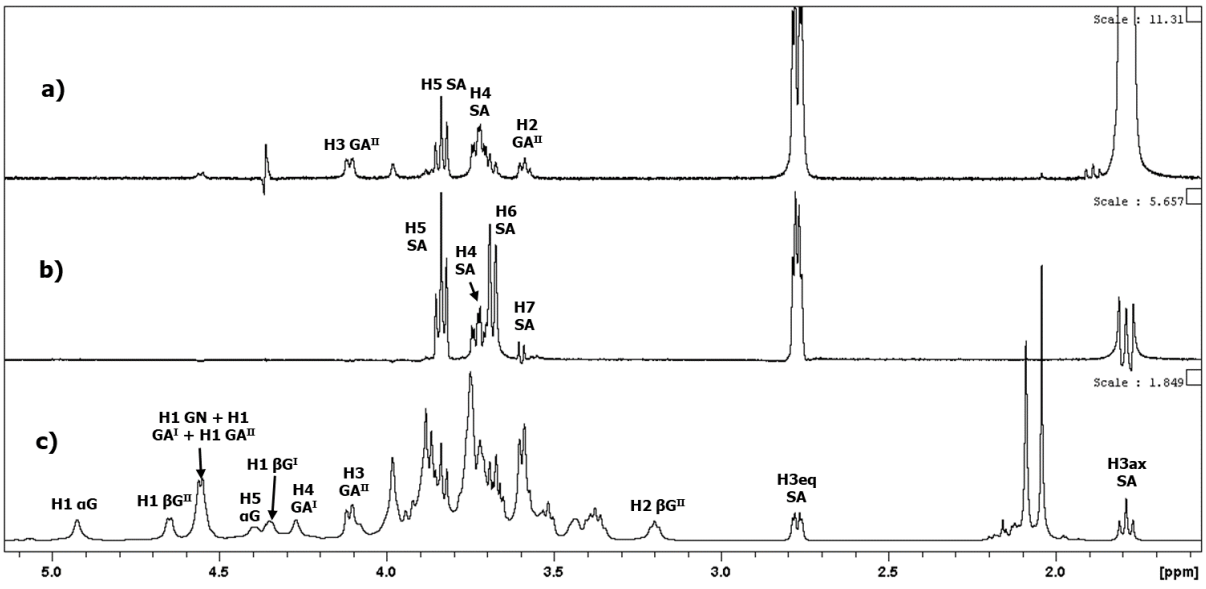


Figure 5.11: Labelled 1D NOESY (a) and TOCSY (b) experiments wherein the H3eq or H3ax proton of SA were selectively irradiated overlaid onto the DOSY spectrum (c) of GBS V.

5.1.2.10. Summary of GBS II, IV and V homonuclear experiments

Thus, after the 2D and 1D homonuclear correlation analysis, H6 of GA^{II} and H5/6 of GA^{III} remain unassigned in GBS II, while H6 of βG and GA^{II} remain unassigned in GBS IV and only H6 GA^{II} remained unassigned for GBS V. Additionally, H8 and H9 of SA were still unassigned for all three serotypes. Each inter-residue linkage of all three serotypes could at least be tentatively confirmed, with the exception of the GA^{III}-(1→6)-GN linkage of GBS II.

5.1.3. Heteronuclear experiments (HSQC, HMBC and hybrid)

As has been established in the previous two chapters, the full ^{13}C NMR assignment and confirmation of all proton assignments in the three serotypes was performed through the combined use of HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC experiments. The previously described overlay of HSQC-DEPT, HSQC-TOCSY and 1D TOCSY experiments were used to confirm preliminary assignments, assign any remaining unassigned protons and identify all of their corresponding carbons.

Figures 5.12, 5.13 and 5.14 show the full assignment and confirmation of one spin system from each of the GBS II, IV and V capsular polysaccharides (G^{II}, GN and $βG^{I}$ residues respectively).

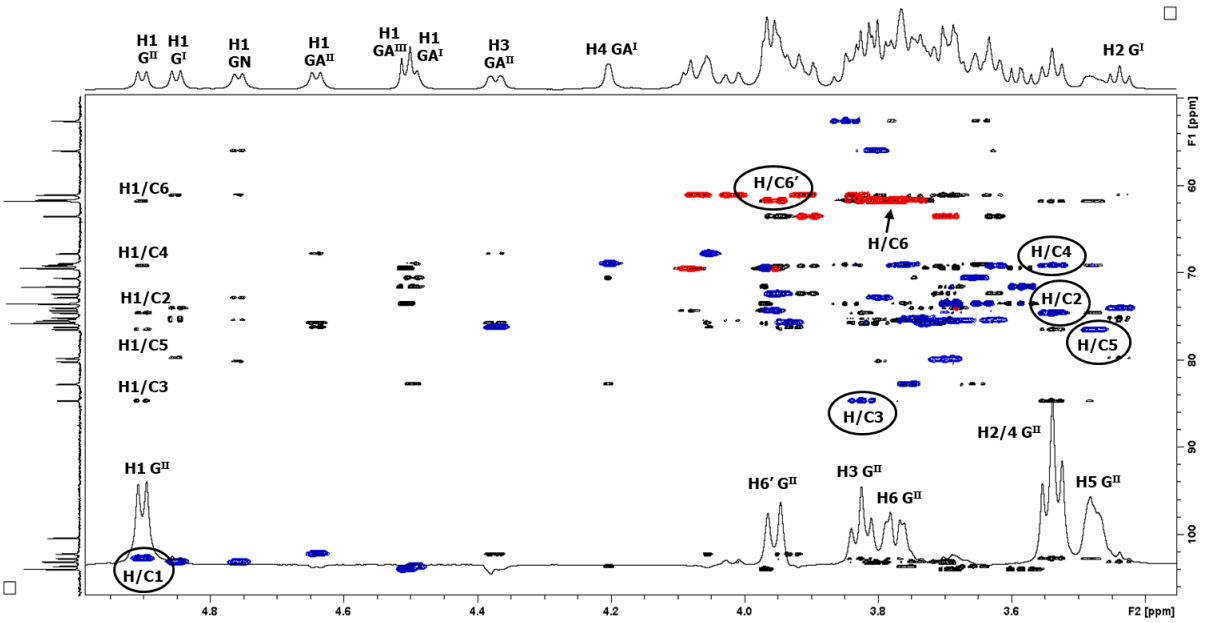


Figure 5.12: Anomeric and ring regions of the selective 1D TOCSY of H1 G^{II} (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS II CPS, highlighting correlations that show the proton and carbon assignments of the entire 3-linked backbone glucose (G^{II}) spin system.

In each, the HSQC-TOCSY signals arranged vertically above the anomeric signal irradiated by the chosen selective 1D TOCSY spectrum show correlations between the anomeric proton and the ring carbons in the residue, allowing for the carbon assignments, especially when cross-referenced with the proton assignments shown by the selective 1D TOCSY overlaid at the bottom of the spectrum.

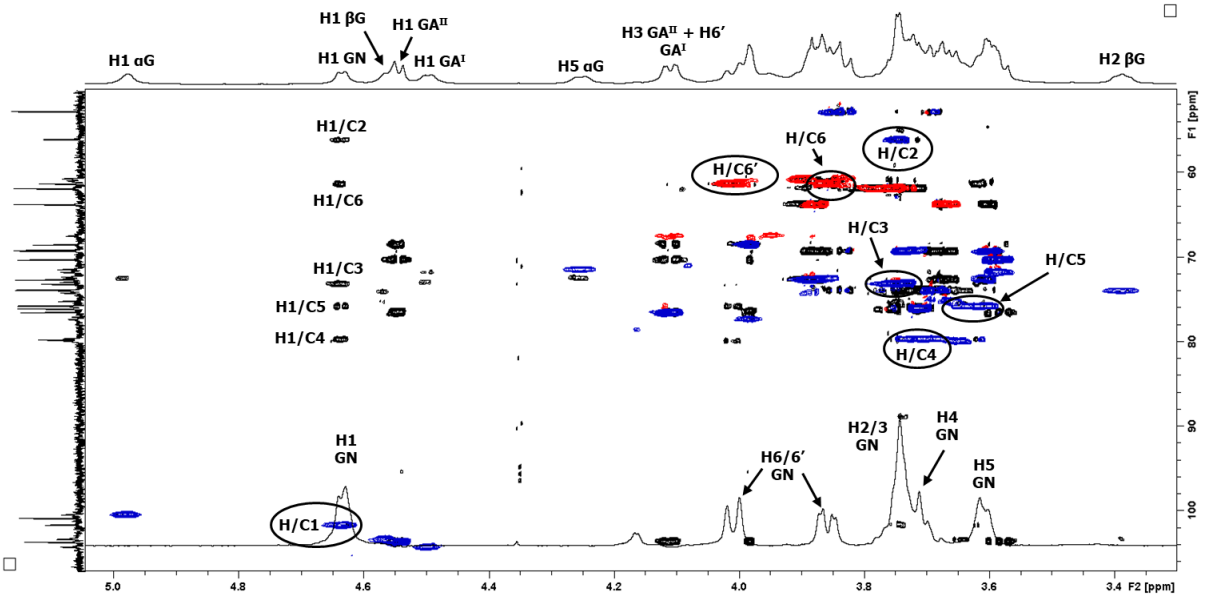


Figure 5.13: Anomeric and ring regions of the selective 1D TOCSY of H1 GN (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS IV CPS, highlighting correlations that show the proton and carbon assignments of the entire GlcNAc spin system.

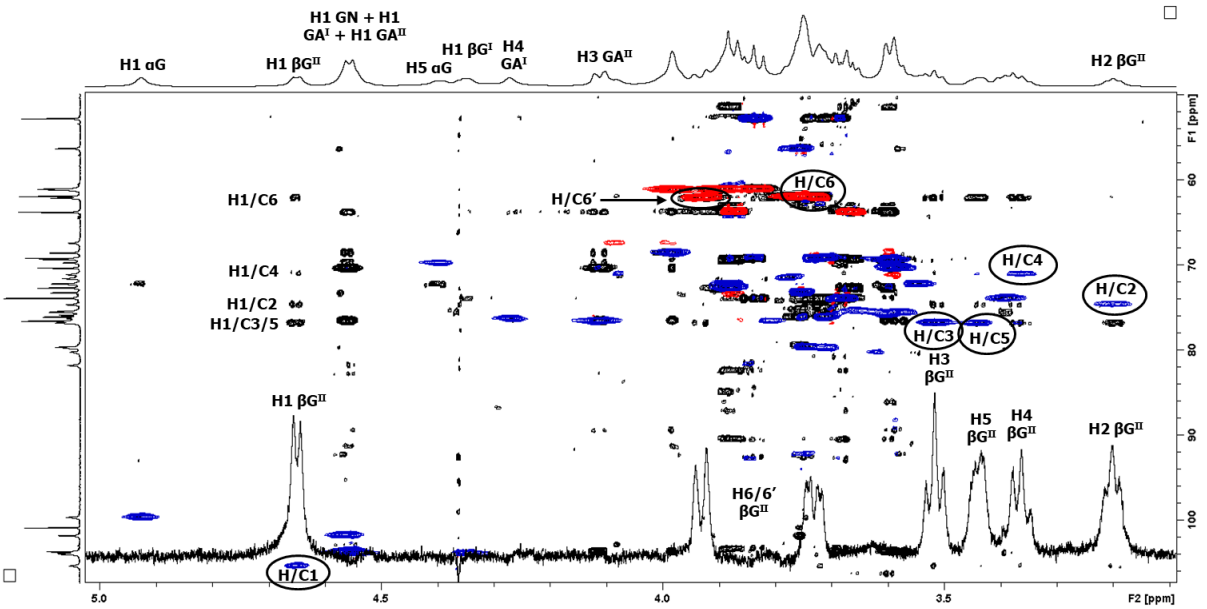


Figure 5.14: Anomeric and ring regions of the selective 1D TOCSY of H1 βG^{I} (black) and the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS V CPS, highlighting correlations that show the proton and carbon assignments of the entire terminal glucose (βG^{II}) spin system.

To go into greater detail on how the final unassigned protons were identified, we will begin with GBS II. Using the HSQC-DEPT and HSQC-TOCSY experiment overlay (Figure 5.15), H8 and H9/9' of SA were assigned, by multiple correlations to H7 SA and each other, as 3.95 and 3.70/3.90 ppm respectively. Additionally, H5 and H6 of GA^{III} were identified by an HSQC-TOCSY correlation, assigning them as 3.72 and 3.82 ppm respectively. Finally, H6 GA^{II} was assigned by a process of elimination as a negative signal at approximately 3.74 ppm since it was the only remaining unassigned proton in GBS II.

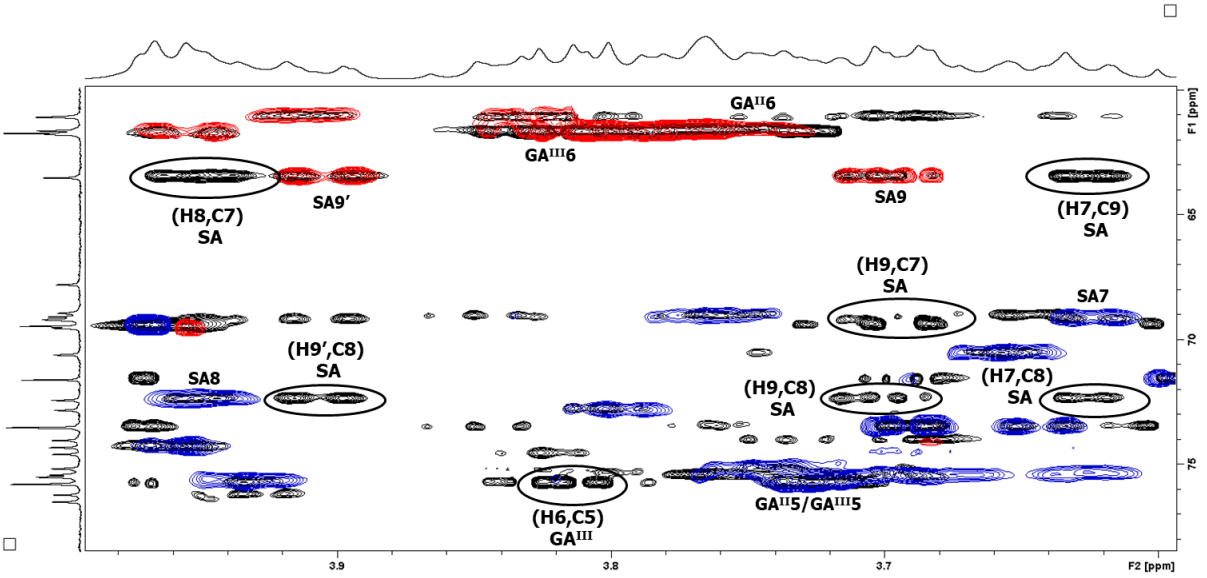


Figure 5.15: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS II CPS, highlighting correlations that led to the assignment of H8 and H9 of SA, H5 and H6 of GA^{III} as well as H6 GA^{II}.

Similarly for GBS IV (Figure 5.16), H8 and H9/9' of sialic acid could be confirmed by a number of correlations to H7 and each other, assigning them as 3.88 and 3.67/3.88 ppm respectively. Additionally, H6 of GA^{II} was identified by its HSQC-TOCSY correlation to H5 GA^{II}, assigning it as 3.76 ppm. Lastly, H6/6' βG were assigned by a process of elimination as a weak pair of signals at approximately 3.85/3.99 ppm since they were the only remaining unassigned protons in GBS IV.

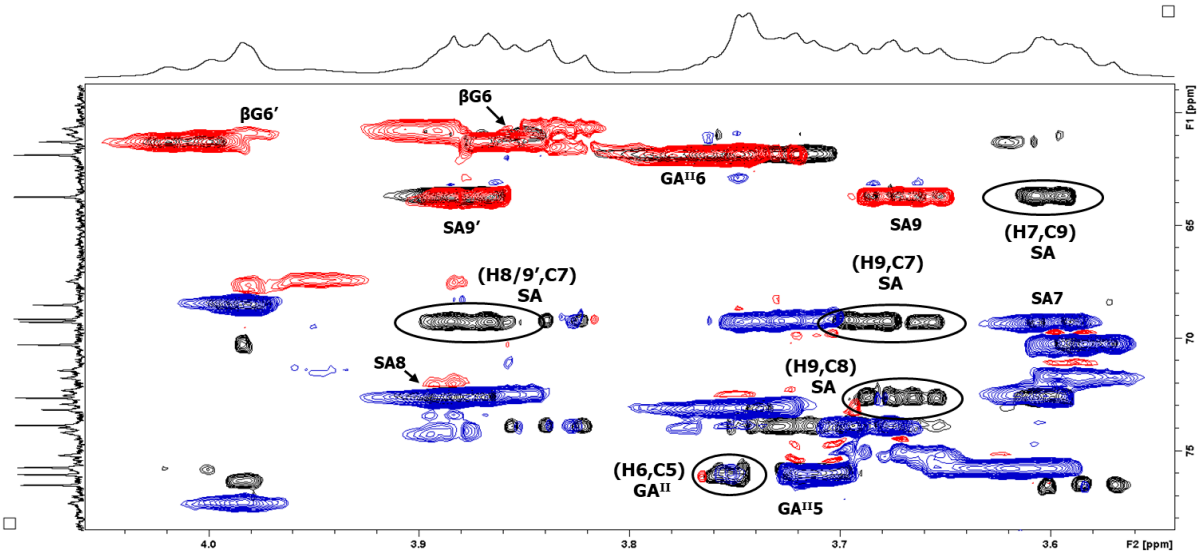


Figure 5.16: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS IV CPS, highlighting correlations that led to the assignment of H8 and H9 of SA as well as H6 of βG and GA^{II}.

For GBS V (Figure 5.17), H8 and H9/9' of sialic acid were, in the same way as for GBS II and IV, assigned as 3.88 and 3.67/3.87 ppm respectively.

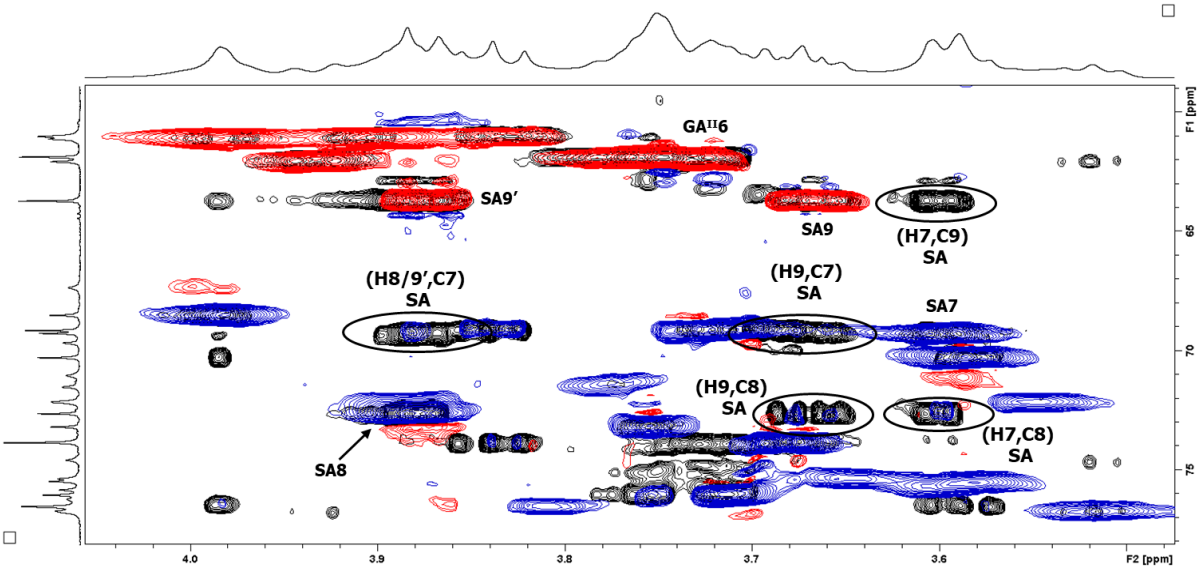


Figure 5.17: Expansion of the ring region of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS V CPS, highlighting correlations that led to the assignment of H8 and H9 of SA as well as H6 of GA^{II}.

This left H6 of GA^{II} as the last unassigned proton of GBS V. It was assigned by a process of elimination as overlapped with H6 of βG^{II} in the large signal at approximately 3.75 ppm.

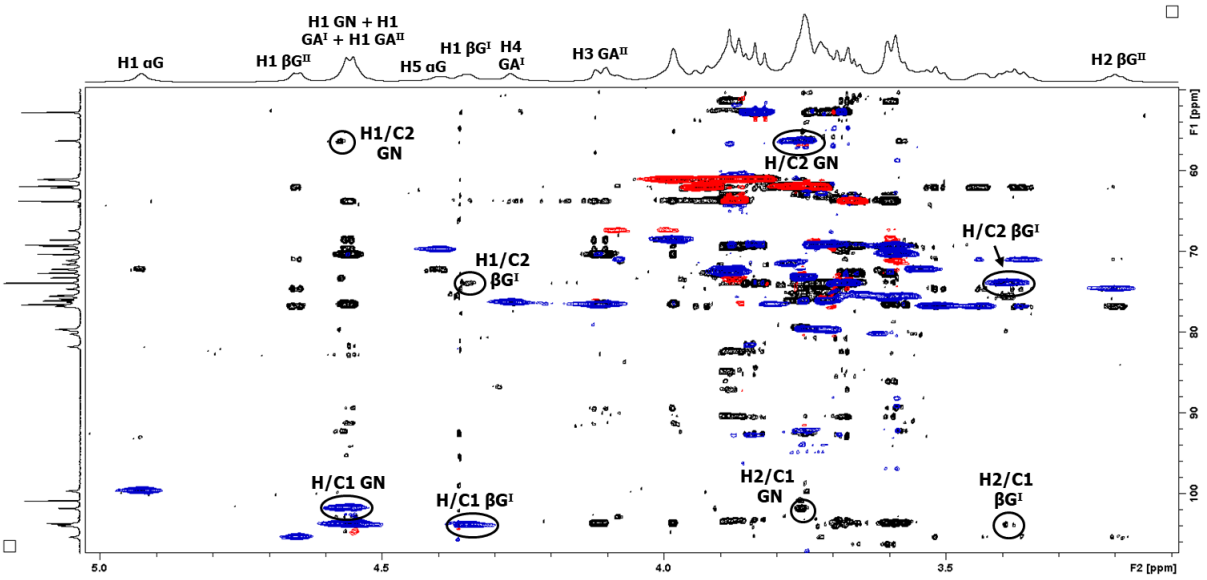


Figure 5.18: Anomeric and ring regions of the HSQC-DEPT spectrum (red/blue) overlaid onto the HSQC-TOCSY spectrum (black) of the GBS V CPS, highlighting correlations that provide evidence that the GN and βG^{I} are indeed at the presented assignments, contrary to the literature.

Lastly, the HSQC-TOCSY provides sufficient evidence that H1 GN and H1 βG^{I} of GBS V are indeed at the presented assignments of 4.56 and 4.35 ppm respectively instead of the inverse of 4.42 ppm and approximately 4.6 ppm as reported in the literature [83,134]. As can be seen from Figure 5.18, the HSQC signal at 4.56/101.72 ppm is the one that shows an HSQC-TOCSY correlation to the easily assignable and diagnostic H/C2 of GN, which has a more upfield carbon signal due to its N-acetyl substitution. Meanwhile, the HSQC signal at 4.35/103.93 ppm only shows a weak correlation to the signal at 3.39/73.82 ppm, which would not make sense for the H/C2 of a GN residue but is instead where one would expect the H/C2 of a β-Glc residue to resonate. Thus, H/C1 of GN must be the HSQC signal at 4.56/101.72 ppm, while the signal at 3.39/73.82 ppm must belong to H/C1 of βG^{I}.

The HMBC spectra allowed partial or full confirmation of the repeating unit connectivity of each residue shown by the NOESY as well as the assignment of the quaternary carbons, by showing intra- and inter-residue correlations through 3-bond coupling.

For GBS II (Figure 5.19), all seven linkages in CPS repeating unit could be confirmed by HMBC. From the anomeric protons, correlations can clearly be seen to their respective linkage carbons, from H1 G^{II} to C2 GA^{II} for the G^{II}-(1→2)-GA^{II} linkage, H1 G^{I} to C3 G^{II} for the G^{I}-(1→3)-G^{II} linkage and similarly for the GN-(1→3)-GA^{I}, GA^{II}-(1→4)-GN, GA^{III}-(1→6)-GN and GA^{I}-

(1→4)-G^{I} linkages. The last linkage, SA-(2→3)-GA^{II}, was confirmed by the reverse correlation from the linkage ring proton, H3 GA^{II}, to the anomeric carbon, C2 SA.

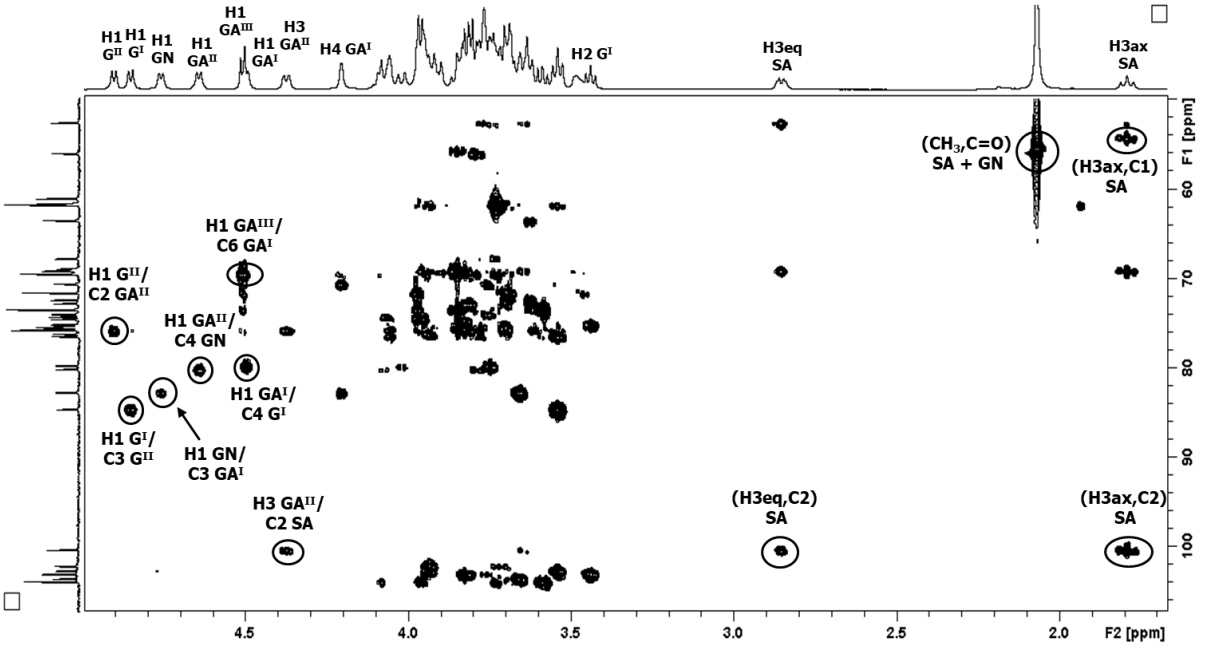


Figure 5.19: HMBC spectrum of the GBS II CPS with important intra- and inter-residue linkage signals labelled.

The quaternary carbons of GBS II could also be assigned by the HMBC. The assignment of C2 SA (100.36 ppm) was made from three correlations including the aforementioned H3 GA^{II} to C2 SA correlation as well as intra-residue correlations to H3eq/ax SA. The carbonyl carbons of the N-acetyl groups in both sialic acid and GlcNAc could be identified by a strong correlation to the CH_3 proton signals, resonating at approximately 55 ppm. Since carbonyl carbons are extremely deshielded and the spectral width of the HMBC experiment was only 120 ppm, these are folded peaks and thus assigned at 175.37 and 175.66 ppm for the GN and SA carbonyl carbons respectively. Similarly, there is a folded correlation between H3ax SA and the carboxyl carbon, C1 SA, that was assigned as 174.17 ppm.

For GBS IV (Figure 5.20), all six linkages in CPS repeating unit could be confirmed by HMBC. From four of the anomeric protons, correlations can be seen to their respective linkage carbons for αG-(1→4)-GA^{I}, GN-(1→6)-GA^{I}, GA^{II}-(1→4)-GN and GA^{I}-(1→4)-βG linkages. The reverse correlations to C1 of βG and C2 of SA confirmed the other two linkages, namely the βG-(1→4)-αG and SA-(2→3)-GA^{II} linkages. The C2 SA quaternary carbon was assigned as 100.84 ppm, while the three folded correlations to the carbonyl carbons of the N-acetyl

groups of SA and GN as well as C1 SA were assigned as 175.91, 175.20 and 174.46 ppm, respectively.

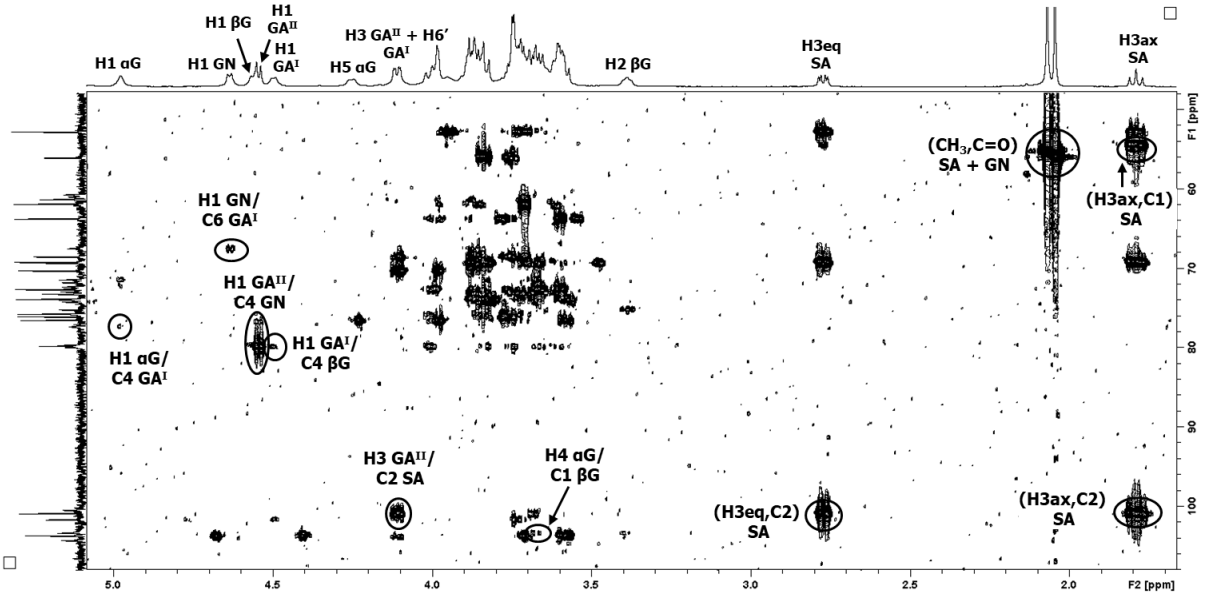


Figure 5.20: HMBC spectrum of the GBS IV CPS with important intra- and inter-residue linkage signals labelled.

Lastly, for GBS V (Figure 5.21), due to the poor quality of the spectrum, only three of the seven linkages could be confirmed by HMBC. The GA^{II}-(1→4)-GN linkage was confirmed by a strong correlation from H1 GA^{II} to C4 GN, while the GA^{I}-(1→4)-βG^{I} linkage was confirmed by a weak correlation from H4 βG^{I} to C1 GA^{I}.

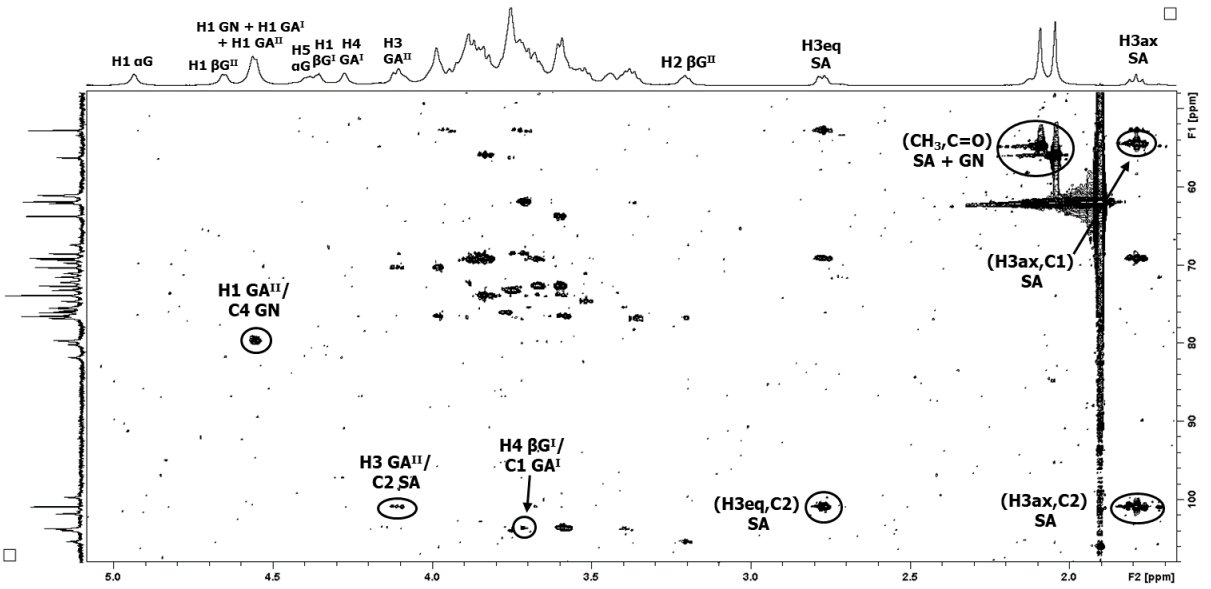


Figure 5.21: HMBC spectrum of the GBS V CPS with some important intra- and inter-residue linkage signals labelled.

The SA-(2→3)-GA^{II} linkage could also be confirmed by a linkage from H3 GA^{II} to C2 SA. The quaternary carbons could however all be assigned, with C2 SA being assigned as 100.81 ppm, while the three folded correlations to the carbonyl carbons of the N-acetyl groups of SA and GN as well as C1 SA were assigned as 175.90, 174.75 and 174.45 ppm respectively.

5.1.4 Full assignment and fingerprint summary

With all assignments made, the HSQC-DEPT spectra serve as 2D identity maps (Figures 5.22, 5.23 and 5.24) showing the chemical shift of every proton/carbon correlation in the GBS II, IV and V CPS repeating units. The chemical shift data is collected in Tables 5.1, 5.2 and 5.3, with calculated glycosylation shifts further confirming the established linkage positions.

The GBS II assignments (Figure 5.22 and Table 5.1) agree with previously published anomeric chemical shift values (recorded at 298 K) [83], with slight differences in chemical shift attributed to the difference in experiment temperature. They also agree to some extent with published data of a GBS type II CPS monomer [144], but with significant differences since the monomer is just a fragment with additional terminal residues compared to the intact repeating unit characterised here.

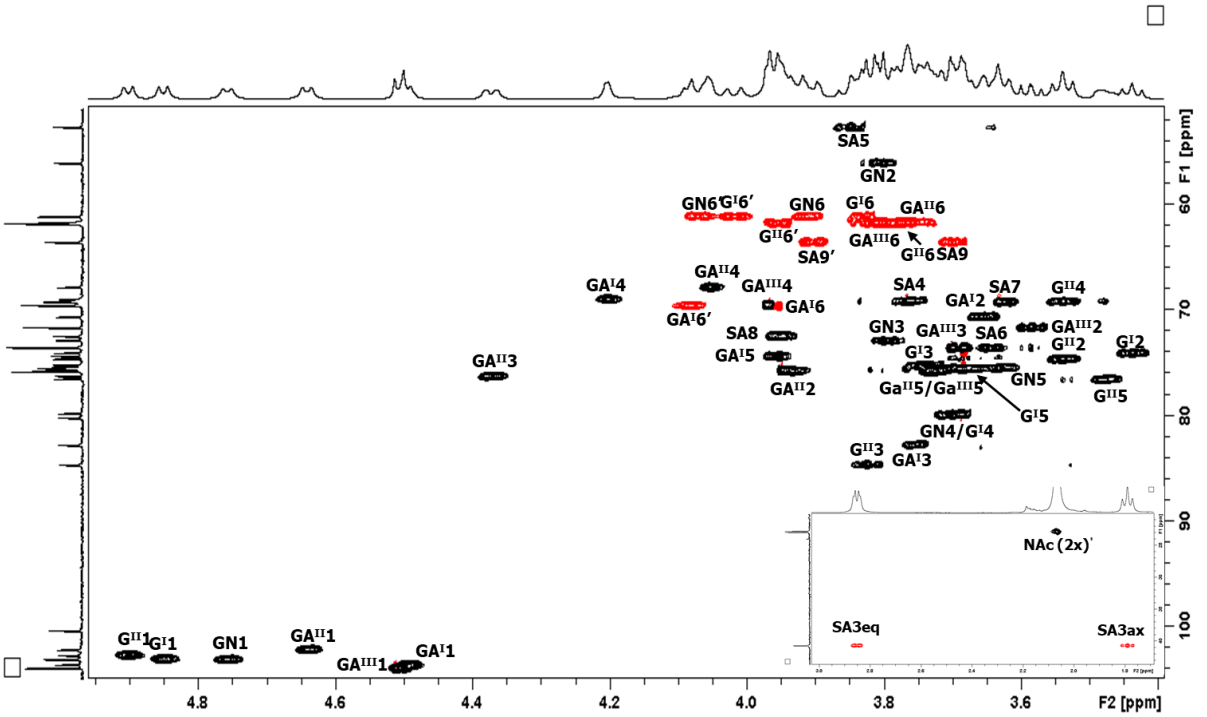


Figure 5.22: Fully labelled HSQC-DEPT spectrum of the GBS II CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

Table 5.1: Full NMR assignment of the GBS II CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→3)-β-D-Glc*p*-(1→	4.90	3.54	3.82	3.54	3.48	3.78/3.95	-	-	-
G^{II}	102.71	74.55	<u>84.61</u>	69.13	76.45	61.67	-	-	-
	(5.87)	(-0.65)	(7.85)	(-1.58)	(-0.31)	(-0.17)			
→4)-β-D-Glc*p*-(1→	4.85	3.44	3.75	3.69	3.68	3.83/4.02	-	-	-
G^{I}	103.08	73.98	75.11	<u>79.73</u>	75.44	61.03	-	-	-
	(6.24)	(-1.22)	(-1.65)	(9.02)	(-1.32)	(-0.81)			
→4)-β-D-Glc*p*NAc-(1→	4.76	3.80	3.80	3.70	3.63	3.91/4.07	-	2.06	-
GN	103.10	55.94	72.77	<u>80.14</u>	75.36	61.03	175.37	22.92	-
	(7.25)	(-1.92)	(-2.04)	(9.08)	(-1.46)	(-0.82)			
→2,3)-β-D-Gal*p*-(1→	4.64	3.93	4.37	4.05	3.74	3.74	-	-	-
GA^{II}	102.18	<u>75.71</u>	<u>76.17</u>	67.75	75.46	61.58	-	-	-
	(4.81)	(2.75)	(2.39)	(-1.94)	(-0.47)	(-0.26)			
→3,6)-β-D-Gal*p*-(1→	4.50	3.66	3.76	4.20	3.95	3.95/ 4.08	-	-	-
GA^{I}	103.60	70.55	<u>82.73</u>	68.88	74.27	<u>69.48</u>	-	-	-
	(6.23)	(-2.41)	(8.95)	(-0.81)	(-1.66)	(7.64)			
β-D-Gal*p*-(1→	4.51	3.59	3.69	3.97	3.72	3.82	-	-	-
GA^{III}	103.93	71.55	73.47	69.41	75.74	61.72	-	-	-
	(6.56)	(-1.41)	(-0.31)	(-0.28)	(-0.19)	(-0.12)			
α-D-Neu*p*NAc-(2→	-	-	1.79/ 2.85	3.76	3.85	3.64	3.62	3.95	3.70/ 3.90
SA[a]	174.17	100.36	40.68	69.03	52.52	73.47	69.16	72.37	63.46
	(-1.76)	(2.34)	(-0.98)	(-0.41)	(-0.48)	(0.00)	(-0.24)	(-0.06)	(-0.58)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.07, 22.74 ppm, C=O at 175.66 ppm.

The GBS IV assignments (Figure 5.23 and Table 5.2) agree with previously published anomeric chemical shift values recorded at 298 K [83] and anomeric and backbone residue signals at 310 K [157] with differences in chemical shift attributed to the difference in experiment temperature.

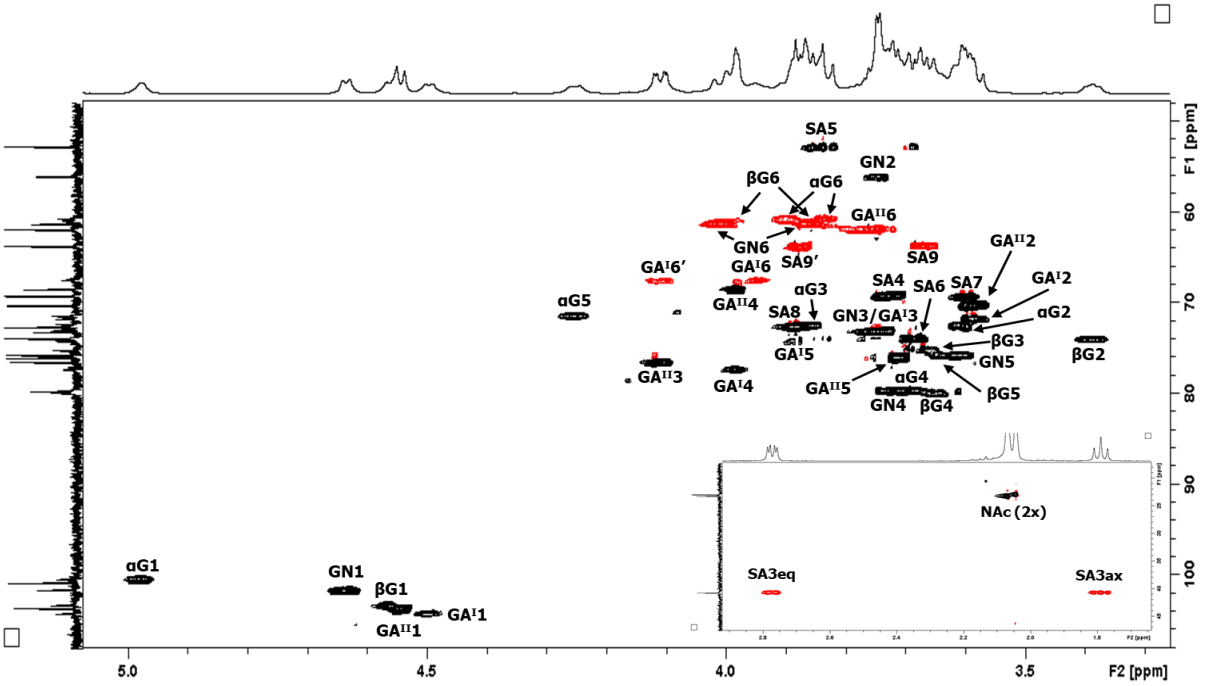


Figure 5.23: Fully labelled HSQC-DEPT spectrum of the GBS IV CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

Table 5.2: Full NMR assignment of the GBS IV CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→4)-α-D-Glc*p*-(1→	4.98	3.60	3.85	3.68	4.25	3.83/ 3.90	-	-	-
αG	100.43	72.30	72.52	<u>79.58</u>	71.41	60.68	-	-	-
	(7.44)	(-0.17)	(-1.26)	(8.87)	(-0.96)	(-1.16)			
→4)-β-D-Glc*p*-(1→	4.56	3.39	3.66	3.66	3.64	3.85/ 3.99	-	-	-
βG	103.31	73.94	75.17	<u>79.90</u>	75.76	60.95	-	-	-
	(6.47)	(-1.26)	(-1.59)	(9.19)	(-1.00)	(-0.89)			
→4)-β-D-Glc*p*NAc-(1→	4.64	3.74	3.74	3.71	3.61	3.86/ 4.00	-	2.07	-
GN	101.62	56.03	73.15	<u>79.74</u>	75.72	61.26	175.20	23.23	-
	(5.77)	(-1.83)	(-1.66)	(8.68)	(-1.10)	(-0.59)			
→3)-β-D-Gal*p*-(1→	4.54	3.58	4.11	3.98	3.71	3.76	-	-	-
GA^{II}	103.63	70.26	<u>76.50</u>	68.50	76.03	61.85	-	-	-
	(6.26)	(-2.70)	(2.72)	(-1.19)	(0.10)	(0.01)			
→4,6)-β-D-Gal*p*-(1→	4.50	3.59	3.73	3.98	3.88	3.95/ 4.11	-	-	-
GA^{I}	104.23	71.74	72.91	<u>77.30</u>	74.16	<u>67.45</u>	-	-	-
	(6.86)	(-1.22)	(-0.87)	(7.61)	(-1.77)	(5.61)			
α-D-Neu*p*NAc-(2→	-	-	1.79/ 2.77	3.73	3.84	3.68	3.60	3.88	3.67/ 3.88
SA[a]	174.46	100.84	40.68	69.13	52.73	73.84	69.25	72.62	63.70
	(-1.47)	(2.82)	(-0.98)	(-0.31)	(-0.27)	(0.37)	(-0.15)	(0.19)	(-0.34)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.04, 22.93 ppm, C=O at 175.91 ppm.

The GBS V assignments (Figure 5.24 and Table 5.3) partly agree with previously published anomeric chemical shift values recorded at 298 K [83] and anomeric and backbone residue signals at 300 K [134] with some slight differences in chemical shift attributed to the difference in experiment temperature. Importantly, however, they reverse the assignments of the GlcNAc and backbone β-Glc anomeric signals, which were unambiguously assigned by a combination of HSQC and HSQC-TOCSY experiments. They also agree somewhat with published spectra of a GBS type V CPS monomer [151], but with significant differences since the monomer is just a fragment and has additional terminal signals which are in a different magnetic environment compared to the intact repeating unit characterised here.

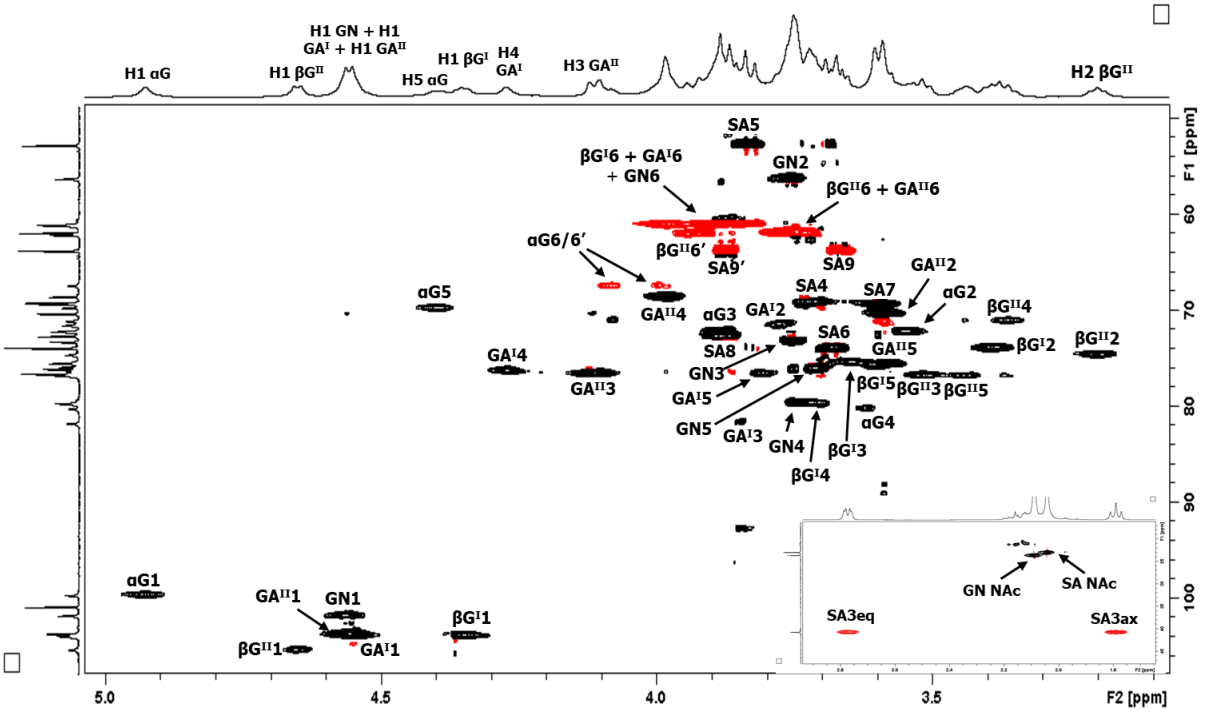


Figure 5.24: Fully labelled HSQC-DEPT spectrum of the GBS V CPS recorded at 600 MHz and 343 K with upfield signals in the insert.

Table 5.3: Full NMR assignment of the GBS V CPS recorded at 600 MHz and 343 K with the calculated carbon glycosylation shifts in brackets. Linkage positions underlined.

Sugar residue	$^{1}H/^{13}C$ (glycosylation shift)								
	1	2	3	4	5	6	7	8	9
→4,6)-α-D-Glc*p*-(1→	4.93	3.55	3.89	3.62	4.39	3.99/ 4.09	-	-	-
αG	99.58 (6.59)	72.12 (-0.35)	72.12 (-1.66)	80.14 (9.43)	69.64 (-2.73)	67.32 (5.48)	-	-	-
β-D-Glc*p*-(1→	4.65	3.20	3.52	3.36	3.44	3.73/ 3.93	-	-	-
βG^{II}	105.30 (8.46)	74.55 (-0.65)	76.69 (-0.07)	71.00 (0.29)	76.74 (-0.02)	62.04 (0.20)	-	-	-
→4)-β-D-Glc*p*NAc-(1→	4.56	3.76	3.75	3.75	3.72	3.83/ 3.90	-	2.09	-
GN	101.72 (5.87)	56.21 (-1.65)	73.16 (-1.65)	79.55 (8.49)	76.01 (-0.81)	61.05 (-0.80)	174.75	23.56	-
→3,4)-β-D-Gal*p*-(1→	4.55	3.77	3.85	4.27	3.81	3.89	-	-	-
GA^{I}	103.82 (6.45)	71.49 (-1.47)	81.67 (7.89)	76.20 (6.51)	76.50 (0.57)	61.05 (-0.79)	-	-	-
→3)-β-D-Gal*p*-(1→	4.54	3.58	4.11	3.98	3.71	3.76	-	-	-
GA^{II}	103.63 (6.26)	70.26 (-2.70)	76.50 (2.72)	68.50 (-1.19)	76.03 (0.10)	61.85 (0.01)	-	-	-
→4)-β-D-Glc*p*-(1→	4.35	3.39	3.65	3.71	3.60	3.90/ 3.98	-	-	-
βG^{I}	103.93 (7.09)	73.82 (-1.38)	75.33 (-1.43)	79.68 (8.97)	75.80 (-0.96)	61.05 (-0.79)	-	-	-
α-D-Neu*p*NAc-(2→	-	-	1.79/ 2.77	3.72	3.84	3.68	3.60	3.88	3.67/ 3.87
SA[a]	174.45 (-1.48)	100.81 (2.79)	40.68 (-0.98)	69.12 (-0.32)	52.71 (-0.29)	73.82 (0.35)	69.23 (-0.17)	72.61 (0.18)	63.69 (-0.35)

[a] The ^{1}H and ^{13}C assignments for NAc of Neu*p*NAc are CH_3 at 2.04, 22.91 ppm, C=O at 175.90 ppm.

Thus, using the NMR approach and methodology described in Chapter 3 for serotype III, the full NMR characterisation of the GBS II, IV and V capsular polysaccharides has been achieved.

Chapter 6. Applications to GBS vaccine development

Having fully assigned the ^{1}H and ^{13}C spectra of the GBS capsular polysaccharide serotypes Ia, Ib, II, III, IV and V in the previous three chapters, this chapter will cover some of the potential applications of these data in the realm of GBS vaccine development.

6.1 Uses of a 1D and 2D NMR database of GBS

As mentioned in Chapter 3, one the benefits of the NMR data recorded in this study is that it was recorded at 343 K instead of the usual temperatures between 298 and 323 K, resulting in sharper, more defined peaks (Figure 6.1).

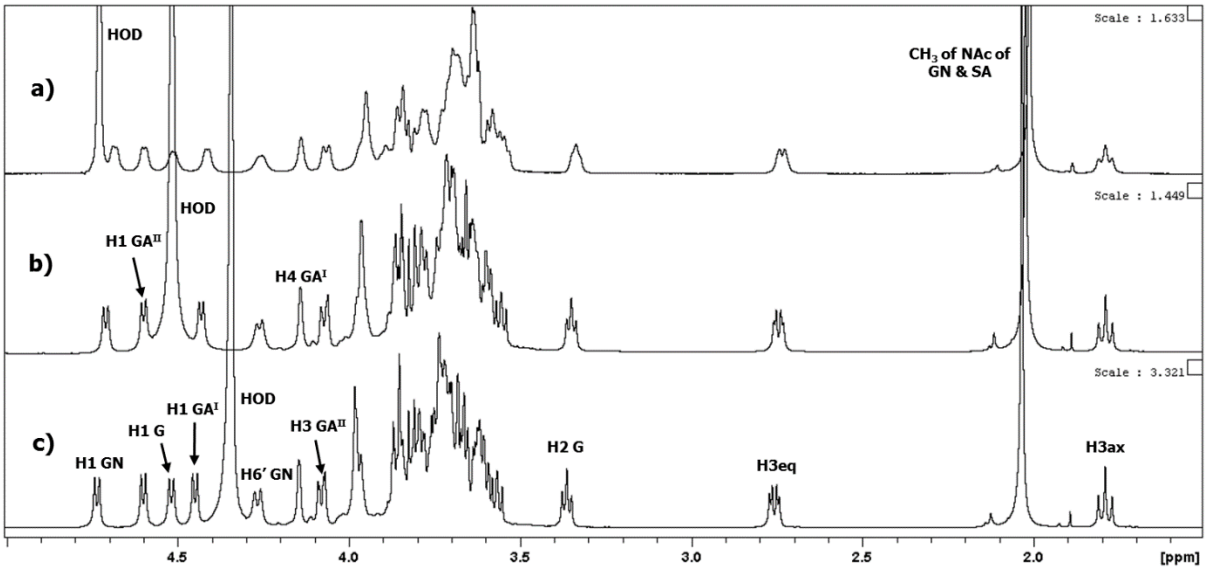


Figure 6.1: Overlay of 1D ^{1}H spectra of the GBS III CPS recorded at 298 (a), 323 (b) and 343 K (c) with diagnostic peaks labelled.

This means that not only is it easier to visually separate peaks that are near each other and that would overlap if they were broader, but it also allows one to better discern signal multiplicity, which is another valuable feature of signal assignment and spectral fingerprinting. Additionally, recording at a higher temperature moves the highly temperature-sensitive HOD peak upfield, away from the anomeric peaks that are commonly used as diagnostic peaks for polysaccharide repeating units. Considering that, at neutral pH, this increased temperature doesn't degrade the polysaccharide, there is a good reason to record NMR experiments of GBS polysaccharides at 343 K especially if the spectra are being used for identity and integrity testing. Having all the NMR data in a database, such as one used for identity and integrity testing, being recorded at the same temperature is also important since the chemical shifts of NMR signals are inherently temperature sensitive to varying degrees,

meaning experiments recorded at different temperatures will result in different chemical shift values, and possibly even misassignments.

Outside of the uses expanded upon in the rest of this chapter, the 1D ^{1}H, ^{13}C, DOSY, TOCSY and NOESY spectra, 2D COSY, TOCSY, NOESY, HSQC and HMBC spectra, as well as the complete proton and carbon chemical shift assignment data presented in this study and their inclusion in a database would be an important reference resource for future efforts in the space of GBS capsular polysaccharide research both in- and outside the realm of vaccine development.

6.1.1 GBS identity testing

The data presented in this study represents a good start on the creation of a consolidated database of GBS capsular polysaccharide NMR data, considering Chapters 3, 4 and 5 cover the complete assignment and fingerprinting of the CPSs of GBS serotypes Ia, Ib, II, III, IV and V which, together, are responsible for more than 95% of global GBS disease cases [7]. The NMR data presented includes 1D ^{1}H and DOSY spectra, which are ideal for fingerprinting since they are experiments that can be performed quickly and thus at little cost. Presented below are overlays of the full 1D ^{1}H spectra (Figure 6.2) and expanded overlays of the anomeric and ring regions of 1D DOSY spectra (Figure 6.3) for the six serotypes covered in this study.

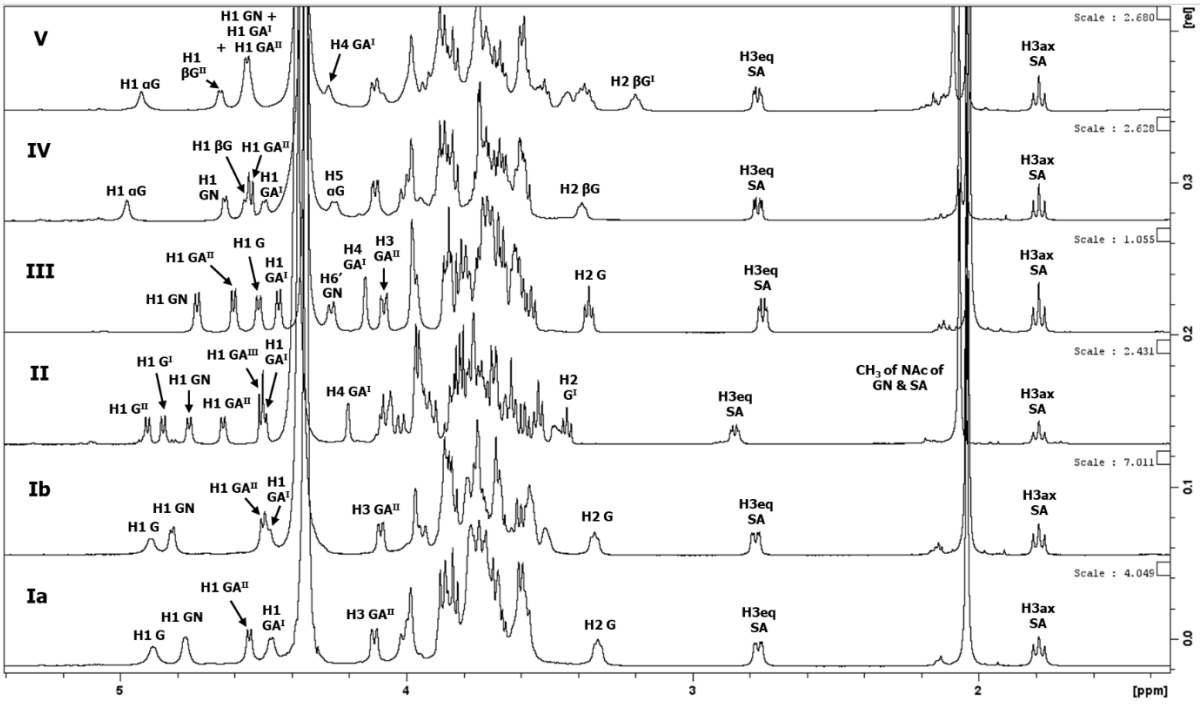


Figure 6.2: Overlay of the 1D ^{1}H spectra of the GBS Ia, Ib, II, III, IV and V CPSs recorded at 600 MHz and 343 K with diagnostic peaks labelled.

The most important signals for confirming the identity and integrity of a GBS polysaccharide are the individual signals that lie outside the crowded ring region between 3.5 and 4.1 ppm. These include the anomeric signals that appear above 4.2 ppm, as well as the outlying signals below 3.4 ppm, which include H3ax of sialic acid; the signal by which each spectrum is referenced, since it remains relatively unaffected by the structural differences of each CPS serotype. The chemical shift values of all of these signals are presented in Chapters 3, 4 and 5.

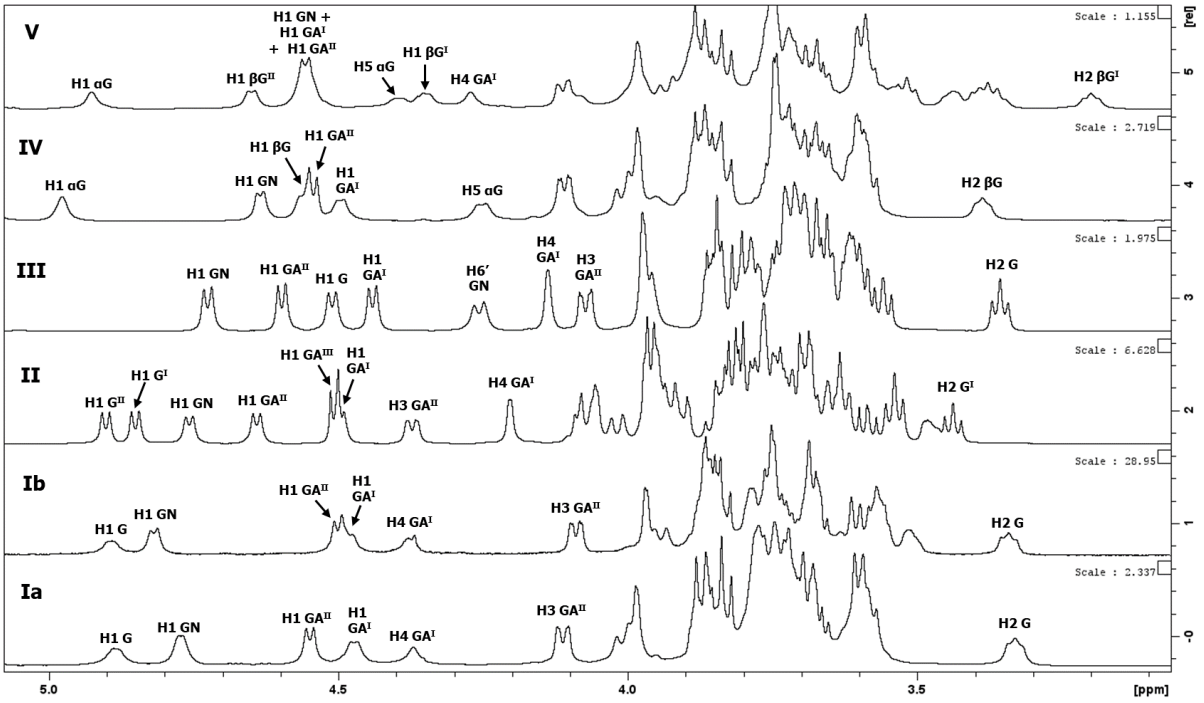


Figure 6.3: Expanded overlay of the anomeric and ring regions of 1D DOSY spectra of the GBS Ia, Ib, II, III, IV and V CPSs with diagnostic peaks labelled.

Figure 6.3 shows the DOSY spectrum which provides one major benefit over a conventional ^{1}H NMR spectrum. The DOSY experiment eliminates signals from low molecular weight compounds, so the large water peak that appears at approximately 4.35 ppm at 343 K, partly overlapped with the anomeric region, is removed. This, as well as the elimination of any other low molecular weight residual peaks, facilitates fingerprinting of the GBS CPS signals, especially in the anomeric region. Thus, Figure 6.3 shows that the anomeric region of each serotype is unique and noticeably different to that of every other serotype, meaning these 1D NMR spectra are diagnostic for each serotype at 343 K. This allows for identity confirmation since these unique configurations of anomeric proton signals can not only

confirm that a given sample being analysed is a GBS capsular polysaccharide but can also identify which specific serotype is being analysed.

6.1.2 Confirmation of polysaccharide N-acetylation status

Another, related, application of the NMR data is integrity testing, i.e., confirming that the polysaccharide preparation has been successful. For isolation involving a base treatment, for example, the final step in the preparation of the GBS CPS is re-N-acetylation of the amino groups of the GlcNAc and sialic acid residues, so a good example of integrity testing is evaluating the degree of N-acetylation in the CPS. The overlay of a number of GBS III CPSs in varying states of de-N-acetylation is presented below (Figure 6.4), and from it a number of peaks indicative of incomplete N-acetylation can be identified.

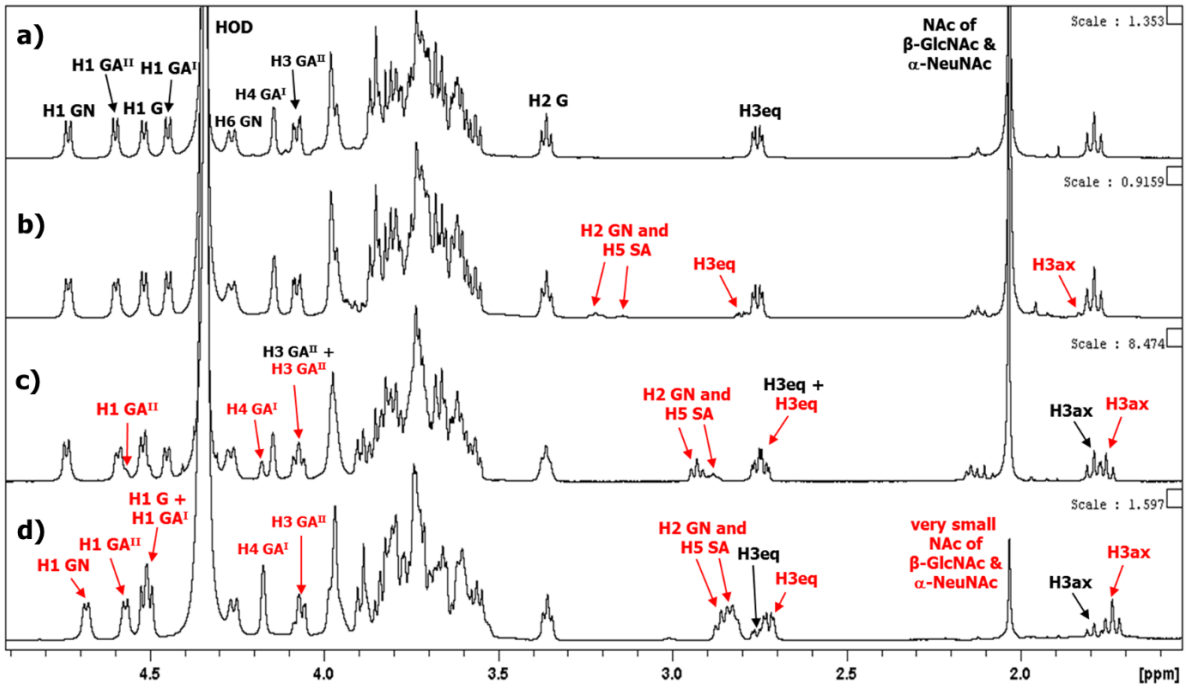


Figure 6.4: Overlay of 1D ^{1}H NMR spectra of fully N-acetylated (a), slightly de-N-acetylated (b), half de-N-acetylated (c) and mostly de-N-acetylated (d) GBS III CPS samples. Peaks arising from de-N-acetylated GBS III are labelled in red.

The most easily identifiable peaks resulting from incomplete N-acetylation include the C<u>H</u>NH$_2$ signals of GlcNAc and SA, H3ax/H3eq signals of SA, H4 GA^{I}, H3 GA^{II} and upfield/downfield shifts of the anomeric signals. Because these signals result from proximity to an amine group, many of them are pH sensitive, most noticeably the C<u>H</u>NH$_2$ signals, which can resonate anywhere in the range of 2.8 to 3.3 ppm and the H3 SA signals which have been observed upfield, overlapped with, or downfield of their respective N-acetylated signals. The extent of

N-acetylation can be estimated by integration of a sufficiently isolated CPS signal, such as H2 of Glc (which is given a value of 1.00) that can be compared to the area of the N-acetyl signal, which should then integrate for approximately 6.00 if fully N-acetylated.

6.1.3 Evaluation of polysaccharide purity

NMR data can also be used to detect any NMR active contaminants in a given sample. This is understandably extremely important in the vaccine development process, to ensure that the capsular polysaccharide sample is pure before conjugation. Figure 6.5 below shows two example samples with contaminants/residuals present in the purified CPS.

An acetate peak can be seen at 1.90 ppm in the GBS Ia spectrum in Figure 6.5, which is derived from the N-acetylation step of CPS preparation. In the GBS III spectrum, trace ethanol can be identified by a triplet at 1.17 ppm, likely from solvent used during preparation, while some broad protein or nucleic acid signals can be seen between 1.3 and 1.5 ppm. Additionally, a small peak can be seen near 1.95 ppm due to some unidentified trace contaminant. For unknown extra peaks, 2D experiments such as HSQC or TOCSY can help elucidate what they may be by determining both the proton and carbon chemical shift of that contaminant/residual signal and identifying its associated spin system.

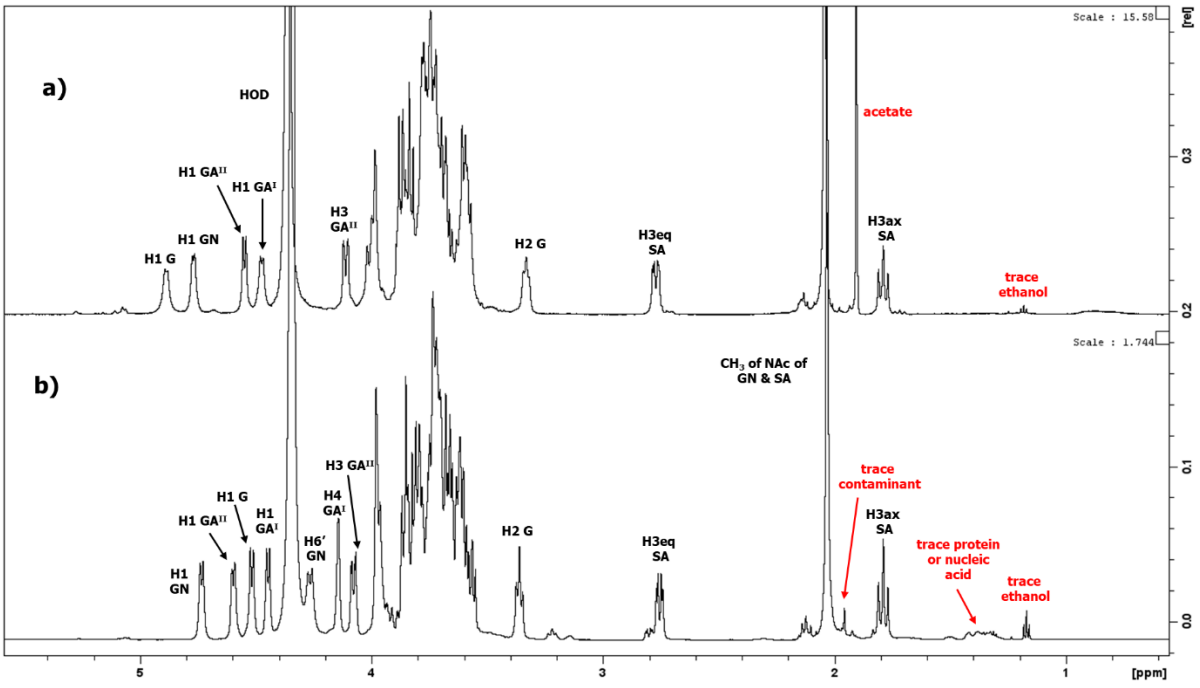


Figure 6.5: Overlay of 1D ^{1}H spectra of GBS Ia (a) and GBS III (b) CPS samples with diagnostic serotype and contaminant/residual peaks labelled in black and red respectively.

6.2 Evaluation of GBS conjugate batches

Similar to capsular polysaccharide preparation, the process of polysaccharide-protein conjugation can also make use of the integrity and purity testing capabilities of NMR to determine whether the structure of the GBS CPS that has been conjugated to the protein is intact and to determine the presence of residuals that may be left over from the conjugation process.

6.2.1 Evaluation of GBS conjugate integrity and purity

NMR spectra of conjugate samples are slightly more challenging to analyse, since they contain broad protein signals in all regions of the spectrum, but fingerprinting can still reliably be performed if the GBS CPS peaks are sufficiently resolved. Figures 6.6 and 6.7 demonstrate this, wherein a batch of GBS III-TT (tetanus toxoid) conjugate is compared to the GBS III CPS batch it was prepared from. This glycoconjugate, as well as the GBS II-TT conjugate presented in Figure 6.8 were prepared by random activation of the polysaccharide using CDAP and conjugation to TT or ADH-derivatised TT.

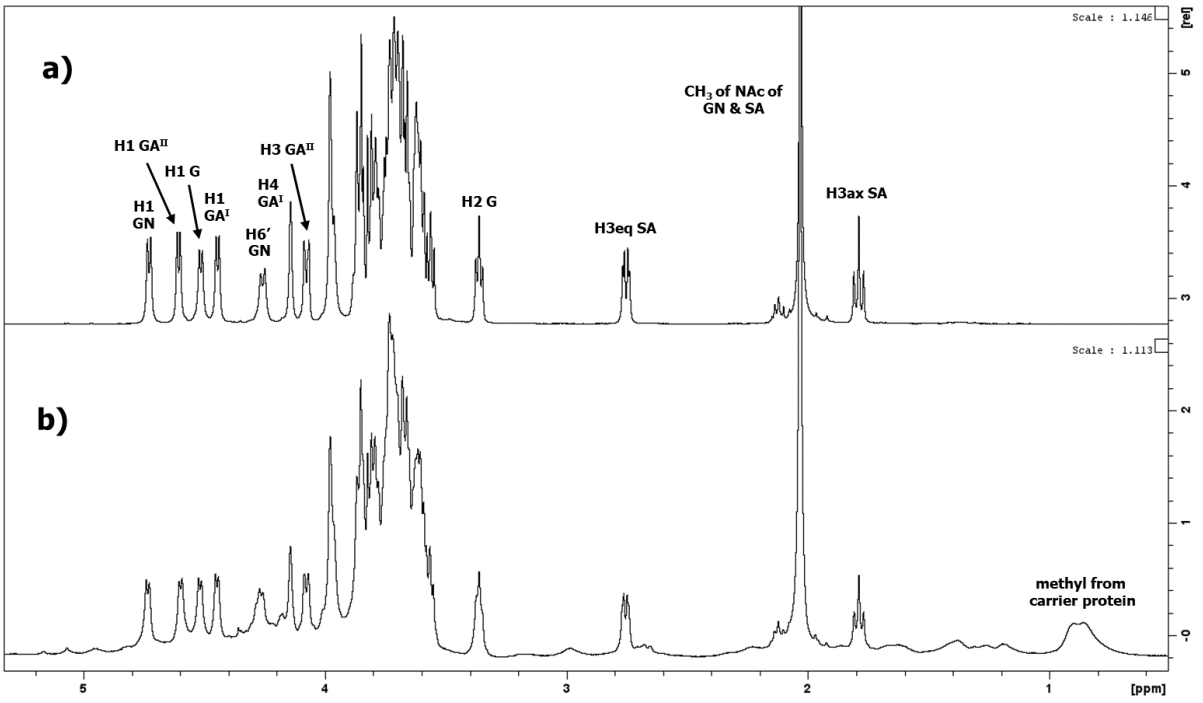


Figure 6.6: Overlay of 1D DOSY spectra of a sample of GBS III CPS (a) and the GBS III-TT conjugate prepared from it (b) with diagnostic peaks labelled.

Even with the broad protein peaks, the GBS III anomeric fingerprint can be identified in the 1D spectrum confirming, at least preliminarily, that the GBS III CPS has remained intact through the conjugation process. A more thorough proof, however, is the 2D DOSY-TOCSY,

wherein the correlation patterns of the anomeric signals provide a very robust fingerprinting method and indeed, in Figure 6.7, the correlation patterns of the four anomeric peaks in the conjugate batch are identical to those of the polysaccharide batch, proving that no degradation of the polysaccharide has occurred during the conjugation process.

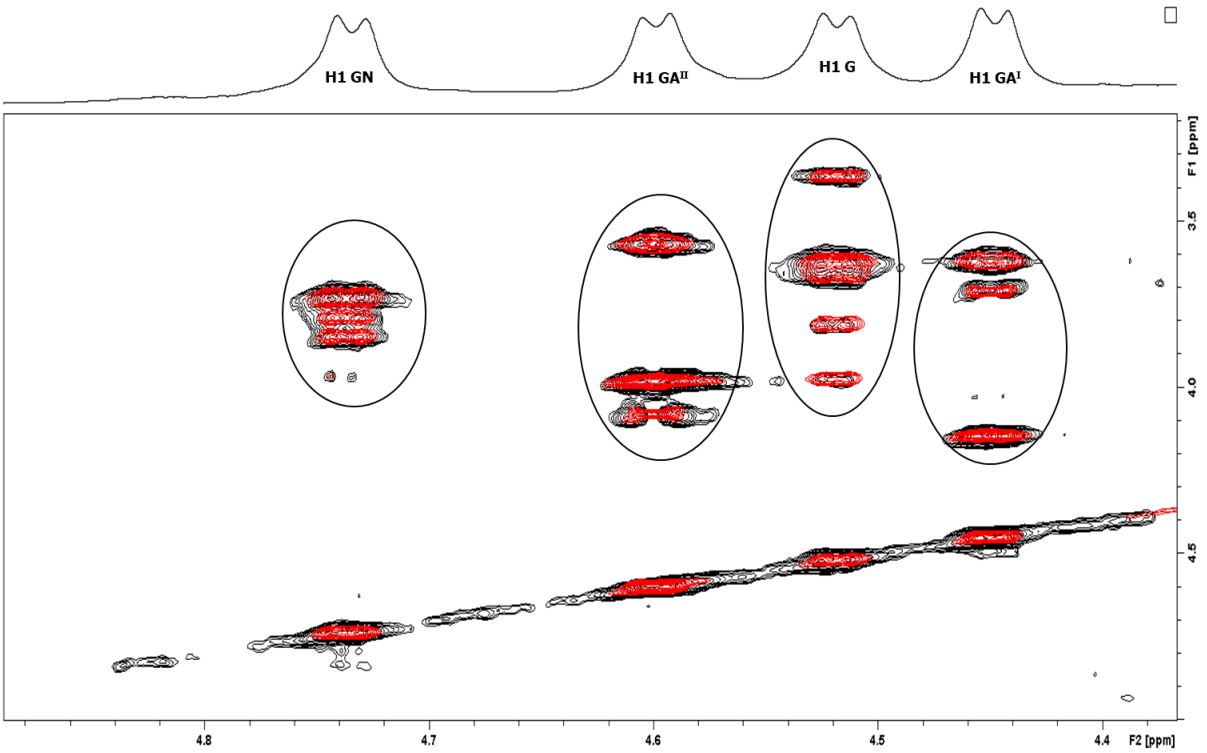


Figure 6.7: Overlaid expanded anomeric region of the 2D DOSY-TOCSY of a sample of GBS III CPS (red) and the GBS III-TT conjugate prepared from it (black) with identical signals circled.

Finally, NMR can be used to identify residual contaminants remaining in a conjugate sample after conjugation. In Figure 6.8 a GBS II-TT conjugate spectrum is shown that has a number of contaminants/residuals, most notably an ethanol triplet at approx. 1.2 ppm and a formaldehyde peak at about 4.86 ppm, but there are also smaller unidentified trace contaminants appearing between 1.2 and 3.3 ppm. To determine the identity of these residuals, like with purity testing for the polysaccharide, HSQC and 2D TOCSY spectra can provide further information such as ^{13}C assignments and correlations that can lead to their identification.

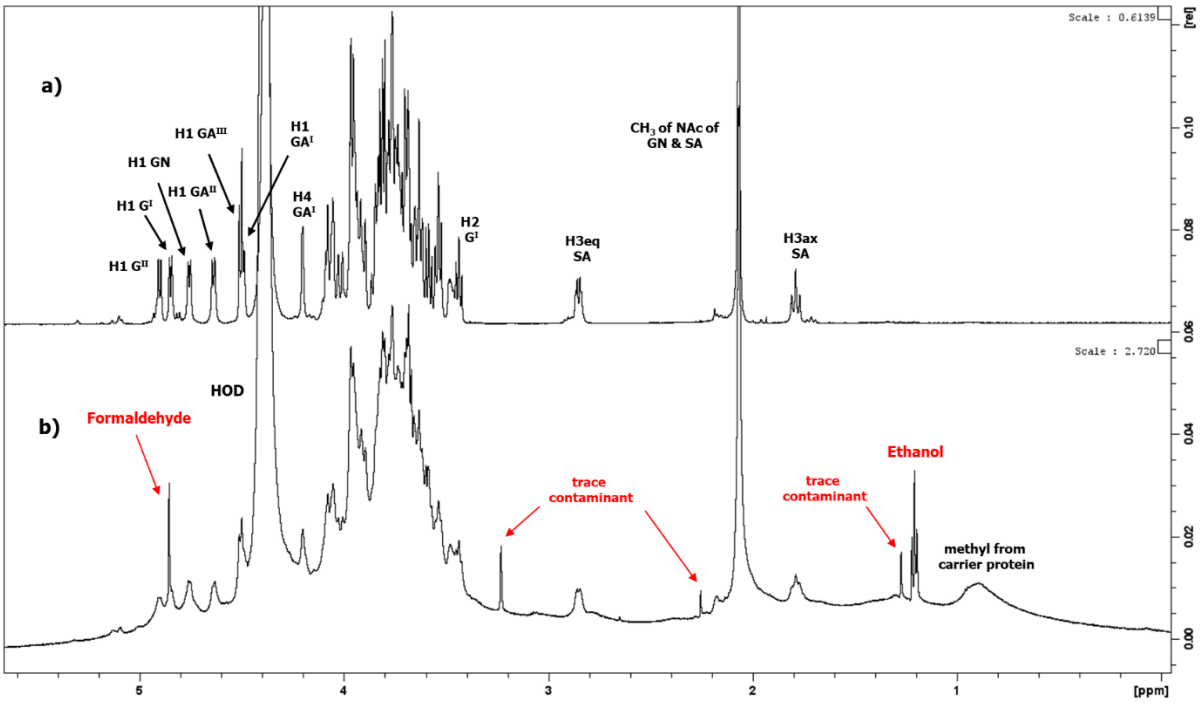


Figure 6.8: Overlay of 1D ^{1}H spectra of a GBS II reference spectrum (a) and a GBS II-TT conjugate with a number of contaminants (b) with diagnostic and contaminant peaks labelled.

Thus, several examples of how the NMR data presented in this study can be applied to analyse polysaccharide and the derived conjugate batches for identity, purity and structural integrity have been described. This indicates the potential of using these NMR data to prepare a standardised database of GBS CPS NMR data to aid and expedite the process of GBS vaccine development.

Chapter 7. Conclusions

This study employed a wide array of nuclear magnetic resonance experiments to perform full NMR characterisation of the capsular polysaccharide repeating units of Group B *Streptococcus* serotypes Ia, Ib, II, III, IV and V to aid in the development of glycoconjugate GBS vaccines. The spectra presented in this work describe the NMR assignment and chemical shift data of every proton and carbon in the six serotypes covered as well as examples of how these data can be used during the preparation and process development of the polysaccharide and the derived conjugate vaccine. These full characterisations of GBS capsular polysaccharides thus aim to serve as reference spectra and chemical shift assignments for use in identity, integrity and purity testing during the preparation and control of future glycoconjugate vaccines against GBS.

NMR spectroscopy has been established as an extremely useful and robust method for tracking the industrial manufacturing process of polysaccharide vaccines from bulk antigen to intermediates and through to the final formulation. Additionally, with advancements in NMR technology, it now possible to fully elucidate the structure of even complex polysaccharides using solely NMR methods. This is because NMR can provide, depending on experiment, all the information that traditionally would be provided by chemical experiments. This includes monosaccharide composition and ratios, linkage patterns, sequences and configurations, as well as the position and extent of non-carbohydrate substituents. The biggest advantages of NMR methods over chemical ones are that they are non-destructive, meaning that this structural information can be obtained by experiments performed on only one sample and requires a relatively small quantity of sample. Currently, the main disadvantage of NMR is that it naturally requires an NMR spectrometer, for which the infrastructure and upkeep is significantly more expensive than that demanded by chemical methods.

The NMR databases and spectra published for vaccines against *Haemophilus influenzae* type b, *Neisseria meningitidis* and *Streptococcus pneumoniae* have facilitated their development, licensure and control testing. Similarly, to perform the identity, integrity and purity testing of GBS polysaccharides during vaccine development, reference NMR spectra and proton/carbon assignments of the specific serotype polysaccharides in question are necessary. While partial proton assignments, primarily of the anomeric protons, and full carbon assignments have been published of some of the GBS CPS serotypes, full proton and carbon assignments have

only been published for GBS type III. However, NMR data currently in the literature was generally recorded in the temperature range of 290 – 323 K. For example, the 1D ^{1}H profiles of the majority of these antigens have been published, however, the identity spectra were recorded at 298 K, resulting in broad peaks and overlap of the large water signal with diagnostic GBS signals in the anomeric region. This work shows that recording the NMR data of the GBS capsular polysaccharides at a temperature of 343 K presents several advantages. Thus, there is a need for reference spectra and proton/carbon assignments of the various GBS CPS serotypes at the temperature of 343 K to facilitate the role of NMR in the area of GBS vaccine development and control.

To begin, an in-depth NMR analysis was performed on the GBS III CPS (chapter 3), since GBS III is the serotype responsible for the greatest number of disease cases, thus it is the most widely studied and characterised by NMR spectroscopy. To establish an approach and the methodology for the NMR characterisation, a wide variety of NMR experiments (1D, DOSY, TOCSY, NOESY, 2D, HSQC, HMBC among others) were sequentially performed on the GBS III CPS. As previously highlighted, this NMR data was gathered at 343 K which presented several advantages. The higher temperature resulted in sharper and more defined peaks that were well resolved. This improvement was key to deconvoluting the data, especially in the heavily overlapped ring region and further, resulted in improved crosspeaks in the 2D correlation experiments. The movement of the temperature-sensitive water peak upfield, away from the anomeric region also facilitated the analysis of the diagnostic anomeric peaks by 2D homonuclear and selective 1D TOCSY and NOESY experiments during characterisation. No degradation of polysaccharide samples due to this increased temperature at neutral pH was observed.

The NMR characterisation began with 1D ^{1}H spectra to assign peaks lying outside the ring region by inspection from previous literature. Additionally, DOSY experiments were also performed since they eliminate signals from low molecular weight compounds, removing the large water peak that appears partly overlapped with the anomeric region at approximately 4.35 ppm at 343 K. To preliminarily determine the chemical shift values of the ring protons of each sugar residue in the CPS, 2D homonuclear correlation experiments were used in conjunction with 1D selective TOCSY spectra. 1D selective and 2D NOESY spectra were used to assign ring signals not assigned using scalar coupling experiments, as well as confirm the

published connectivities between the sugar residues of the repeating unit through inter-residue correlations.

Once nearly every proton signal had been assigned, 2D heteronuclear experiments such as HSQC-DEPT, HSQC-TOCSY, HSQC-NOESY and HMBC were used to perform the ^{13}C assignments as well as the confirmation of all the ^{1}H assignments. The majority of these assignments were achieved using overlays of the selective 1D TOCSY spectra of the anomeric signals over the 2D HSQC-DEPT and HSQC-TOCSY spectra. It also was possible to assign some peaks using a similar overlay of the 1D NOESY, 2D HSQC-DEPT and HSQC-NOESY spectra. The repeating unit connectivity was yet further confirmed by the HMBC spectrum, which also allowed for the assignment of the signals from carbons that do not have any attached protons. Once all of the protons and carbons had been assigned, the HSQC-DEPT spectrum can serve as the 2D identity map of the GBS III CPS, showing the chemical shift of every proton/carbon correlation in the CPS repeating unit. The chemical shift data was also presented in tabulated form and glycosylation shifts relative to unsubstituted monosaccharides were calculated in order to further validate the linkage positions.

The analytical workflow that was established during the characterisation of GBS serotype III described in chapter 3 was then applied to the two GBS serotypes Ia and Ib in chapter 4. This pair of serotypes are structurally similar, differing by only one linkage, namely the Gal-GlcNAc linkage in the branch of the repeating unit, with the GBS Ia CPS having a (1→4) linkage and the GBS Ib CPS having a (1→3) linkage. An additional experiment, HSQC-NOESY was performed on GBS Ia to establish correlations confirming the GN-(1→3)-GA^{I} and G-(1→4)-GA^{I} linkages, unfortunately only showing a correlation between H1 GN and C3 GA^{I}, leaving the G-(1→4)-GA^{I} linkage only assigned by the 1D and 2D homonuclear NOESY experiments

Once they had been fully assigned, a comparison between their NMR spectra was presented, to highlight the differences between their similar NMR profiles. Compared to GBS Ia, GBS Ib displays significant upfield shifts of the H1 GA^{II} and H5 GN proton signals as well as the separation of the two methyl signals in the 1D ^{1}H spectrum. Additionally, the GN3 and GN4 signals of GBS Ib in the HSQC spectrum show significant downfield and upfield shifts respectively as compared to GBS Ia.

The previously established methodology was then applied to the CPSs of the remaining three GBS serotypes covered in this study, GBS II, IV and V in chapter 5. The assignments of these three serotypes (and indeed all six serotypes covered) largely agree with previously published NMR data for these serotypes. The exception to this was GBS V, where HSQC-TOCSY data presented in chapter 5.1.3. shows that the anomeric peaks of GlcNAc and the backbone β-Glc are reversed relative to their assignments in the current literature.

Once these serotypes had been fully elucidated, the data was used to perform identity, integrity and purity testing on a number of polysaccharide and CPS-TT conjugate samples in chapter 6. It allowed identification of each serotype by its diagnostic anomeric peaks, confirmed the structural integrity of the polysaccharide both before and after conjugation and could aid in the detection of impurities such as residuals.

This means that the goal of providing a comprehensive NMR characterisation of the CPS repeating units of GBS serotypes Ia, Ib and II – V, including complete proton and carbon assignments at 343 K has been successfully achieved. Additionally, in Chapter 6 the data was used to perform identity, integrity and purity testing on a number of polysaccharide and CPS-TT conjugate samples, achieving the aim of showing that the data can be used in or as a reference database for future GBS vaccine development.

The continuation of this work would involve further validating the assignments presented in this work with higher quality data, especially for some of the HSQC-TOCSY and HMBC spectra, as well as applying the same analysis performed in this work to the four remaining GBS serotypes (VI, VII, VIII, IX), since the ultimate goal of the GBS vaccine development effort is to produce a universally effective vaccine which likely means it will have to be decavalent to provide full coverage. To facilitate the development of vaccines against those serotypes, reference spectra and assignments such as the ones presented in this study will be necessary.

Another avenue of future work could include molecular dynamics simulations alongside the NMR characterisation. This would not only provide 3D conformational data to accompany the NMR characterisation but could also contextualise and be corroborated by the NMR data such as NOESY correlations confirming predicted linkage angles.

Furthermore, in view of a broader application, this approach of NMR characterisation can be applied to other bacterial capsular polysaccharides, including ones with O-acetylation, as

prior studies conducted on O-acetylated capsular polysaccharides show no loss of O-acetylation, since the loss of O-acetyl groups is dependent on pH, not temperature (ref).

Overall, this work presents a general approach and methodology for the full NMR characterisation of a bacterial capsular polysaccharide. Specifically, this work presents this approach and methodology as applied to the capsular polysaccharides of GBS serotypes Ia, Ib, II – V. Ultimately, the NMR characterisation and data presented here successfully provide a powerful reference resource for use in the development, preparation and control testing of future GBS glycoconjugate vaccines.

References

[1] P. T. Heath, “An update on vaccination against group B streptococcus,” *Expert Rev. Vaccines*, vol. 10, no. 5, pp. 685–694, 2011.

[2] S. Shabayek and B. Spellerberg, “Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology,” *Front. Microbiol.*, vol. 9, no. MAR, pp. 1–14, 2018.

[3] V. N. Raabe and A. L. Shane, “Group B Streptococcus (Streptococcus agalactiae),” *Microbiol. Spectr.*, vol. 7, no. 2, pp. 228–238, 2019.

[4] T. H. Skoff *et al.*, “Increasing Burden of Invasive Group B Streptococcal Disease in Nonpregnant Adults, 1990-2007,” *Clin. Infect. Dis.*, vol. 49, no. 1, pp. 85–92, 2009.

[5] M. J. Cieslewicz *et al.*, “Structural and Genetic Diversity of Group B Streptococcus Capsular Polysaccharides,” vol. 73, no. 5, pp. 3096–3103, 2005, doi: 10.1128/IAI.73.5.3096.

[6] H. C. Slotved, F. Kong, L. Lambertsen, S. Sauer, and G. L. Gilbert, “Serotype IX, a Proposed New Streptococcus agalactiae Serotype,” *J. Clin. Microbiol.*, vol. 45, no. 9, pp. 2929–2936, 2007.

[7] L. Madrid *et al.*, “Infant Group B Streptococcal Disease Incidence and Serotypes Worldwide: Systematic Review and Meta-analyses,” *Clin. Infect. Dis.*, vol. 65, no. March, pp. S160–S172, 2017.

[8] R. C. Lancefield, “A Serological Differentiation of Specific Types of Bovine Hemolytic Streptococci (Group B),” 1934.

[9] R. M. Fry, “Fatal Infections By Hæmolytic Streptococcus Group B,” *Lancet*, vol. 231, no. 5969, pp. 199–201, 1938.

[10] M. Hood, A. Janney, and G. Dameron, “Beta hemolytic streptococcus group B associated with problems of the perinatal period,” *Am. J. Obstet. Gynecol.*, vol. 82, no. 4, pp. 809–818, 1961.

[11] B. F. Anthony and D. M. Okada, “The Emergence of Group B Streptococci in Infections of the Newborn Infant,” *Annu. Rev. Med.*, vol. 28, pp. 355–369, 1977.

[12] R. E. Behrman, “Group B Streptococci: The New Challenge in Neonatal Infections,” *J. Pediatr.*, vol. 82, no. 4, pp. 703–706, 1973.

[13] A. Schuchat, “Epidemiology of Group B Streptococcal Disease in the United States: Shifting Paradigms,” *Clin. Microbiol. Rev.*, vol. 11, no. 3, pp. 497–513, 1998.

[14] A. C. Seale *et al.*, “Stillbirth with Group B Streptococcus Disease Worldwide: Systematic

Review and Meta-analyses," *Clin. Infect. Dis.*, vol. 65, no. Suppl 2, pp. S125–S132, 2017.

[15] J. M. Martin, "Probable Community Acquisition of Group B Streptococcus in an Infant With Late-Onset Disease: Demonstration Using Field Inversion Gel Electrophoresis," *Arch. Pediatr. Adolesc. Med.*, vol. 150, no. 7, pp. 766–768, 1996.

[16] S. J. Schrag *et al.*, "Epidemiology of Invasive Early-Onset Neonatal Sepsis, 2005 to 2014," *Pediatrics*, vol. 138, no. 6, pp. 1–9, 2016.

[17] A. Berardi *et al.*, "Group B Streptococcus Late-Onset Disease: 2003-2010," *Pediatrics*, vol. 131, no. 2, pp. e361–e368, 2013.

[18] R. Libster *et al.*, "Long-Term Outcomes of Group B Streptococcal Meningitis," *Pediatrics*, vol. 130, no. 1, pp. e8–e15, 2012.

[19] Z. Dangor *et al.*, "Burden of Invasive Group B Streptococcus Disease and Early Neurological Sequelae in South African Infants," *PLoS One*, vol. 10, no. 4, pp. 1–13, 2015.

[20] M. Deutscher, M. Lewis, E. R. Zell, T. H. Taylor, C. Van Beneden, and S. Schrag, "Incidence and Severity of Invasive Streptococcus pneumoniae, Group A Streptococcus, and Group B Streptococcus Infections Among Pregnant and Postpartum Women," *Clin. Infect. Dis.*, vol. 53, no. 2, pp. 114–123, 2011.

[21] L. A. Jackson *et al.*, "Risk Factors for Group B Streptococcal Disease in Adults," *Ann. Intern. Med.*, vol. 123, no. 6, pp. 415–420, 1995.

[22] M. Fabbrini *et al.*, "The Protective Value of Maternal Group B Streptococcus Antibodies: Quantitative and Functional Analysis of Naturally Acquired Responses to Capsular Polysaccharides and Pilus Proteins in European Maternal Sera," *Clin. Infect. Dis.*, vol. 63, no. 6, pp. 746–753, 2016.

[23] P. Melin and A. Efstratiou, "Group B Streptococcal Epidemiology and Vaccine Needs in Developed Countries," *Vaccine*, vol. 31, no. S4, pp. D31–D42, 2013.

[24] A. Amin, Y. M. Abdulrazzaq, and S. Uduman, "Group B Streptococcal Serotype Distribution of Isolates from Colonized Pregnant Women at the Time of Delivery in United Arab Emirates," *J. Infect.*, vol. 45, no. 1, pp. 42–46, 2002.

[25] S. Shabayek, S. Abdalla, and A. M. Abouzeid, "Serotype and Surface Protein Gene Distribution of Colonizing Group B Streptococcus in Women in Egypt," *Epidemiol. Infect.*, vol. 142, no. 1, pp. 208–210, 2014.

[26] K. Matsubara, K. Katayama, K. Baba, H. Nigami, H. Harigaya, and M. Sugiyama,

"Seroepidemiologic Studies of Serotype VIII Group B Streptococcus in Japan," *J. Infect. Dis.*, vol. 186, no. 6, pp. 855–858, 2002.

[27] A. Alhhazmi, D. Hurteau, and G. J. Tyrrell, "Epidemiology of Invasive Group B Streptococcal Disease in Alberta, Canada, from 2003 to 2013," *J. Clin. Microbiol.*, vol. 54, no. 7, pp. 1774–1781, 2016.

[28] S. J. Schrag *et al.*, "Group B Streptococcal Disease in the Era of Intrapartum Antibiotic Prophylaxis," *N. Engl. J. Med.*, vol. 342, no. 1, pp. 15–20, 2000.

[29] K. M. Edmond *et al.*, "Group B streptococcal disease in infants aged younger than 3 months: Systematic review and meta-analysis," *Lancet*, vol. 379, no. 9815, pp. 547–556, 2012, doi: 10.1016/S0140-6736(11)61651-6.

[30] S. A. Madhi *et al.*, "High Burden of Invasive Streptococcus agalactiae Disease in South African Infants," *Ann. Trop. Paediatr.*, vol. 23, no. 1, pp. 15–23, 2003.

[31] M. R. Jacobs, H. J. Koornhof, and H. Stein, "Group B Streptococcal Infections in Neonates and Infants," *South African Med. J.*, vol. 54, no. July, pp. 154–158, 1978.

[32] A. F. Hallett, P. Govender, S. S. Pillay, and R. Cooper, "Group B Streptococci in Blacks," *South African Med. J.*, vol. 55, no. February, pp. 157–159, 1979.

[33] I. E. Haffejee, R. H. Bhana, Y. M. Coovadia, A. A. Hoosen, A. V. Marajh, and E. Gouws, "Neonatal Group B Streptococcal Infections in Indian (Asian) babies in South Africa," *J. Infect.*, vol. 22, pp. 225–231, 1991.

[34] C. L. Cutland *et al.*, "Maternal HIV Infection and Vertical Transmission of Pathogenic Bacteria," *Pediatrics*, vol. 130, no. 3, pp. e581–e590, 2012.

[35] C. L. Cutland *et al.*, "Increased Risk for Group B Streptococcus Sepsis in Young Infants Exposed to HIV, Soweto, South Africa, 2004–2008," *Emerg. Infect. Dis.*, vol. 21, no. 4, pp. 638–645, 2015.

[36] V. Quan *et al.*, "Invasive Group B Streptococcal Disease in South Africa: Importance of Surveillance Methodology," *PLoS One*, vol. 11, no. 4, pp. 1–10, 2016.

[37] "https://www.cdc.gov/drugresistance/pdf/threats-report/gbs-508.pdf." .

[38] Y. P. Li, C. M. Kuok, S. Y. Lin, W. S. Hsieh, and M. K. Shyu, "Group B streptococcus antimicrobial resistance in neonates born to group B streptococcus-colonized mothers: Single-center survey," *J. Obstet. Gynaecol. Res.*, vol. 42, no. 11, pp. 1471–1475, 2016, doi: 10.1111/jog.13082.

[39] S. J. Schrag *et al.*, “Risk Factors for Neonatal Sepsis and Perinatal Death Among Infants Enrolled in the Prevention of Perinatal Sepsis Trial, Soweto, South Africa,” *Pediatr. Infect. Dis. J.*, vol. 31, no. 8, pp. 821–826, 2012.

[40] E. R. Moxon and J. S. Kroll, “The Role of Bacterial Polysaccharide Capsules as Virulence Factors,” *Curr. Top. Microbiol. Immunol.*, vol. 150, pp. 65–85, 1990.

[41] J. W. Costerton and R. T. Irvin, “The Bacterial Glycocalyx in Nature and Disease,” *Annu. Rev. Microbiol.*, vol. 35, no. 1, pp. 299–324, 1981.

[42] J. W. Costerton *et al.*, “Bacterial Biofilms in Nature and Disease,” *Ann. Rev. Microbiol*, vol. 41, no. 1, pp. 435–464, 1987.

[43] I. S. Roberts, F. K. Saunders, and G. J. Boulnois, “Bacterial Capsules and Interactions with Complement and Phagocytes,” *Biochem. Soc. Trans.*, vol. 17, no. 3, pp. 462–464, 1989.

[44] I. C. Sutcliffe, G. W. Black, and D. J. Harrington, “Bioinformatic insights into the biosynthesis of the Group B carbohydrate in Streptococcus agalactiae,” *Microbiology*, vol. 154, no. 5, pp. 1354–1363, 2008, doi: 10.1099/mic.0.2007/014522-0.

[45] D. N. Greenberg, D. P. Ascher, B. A. Yoder, D. M. Hensley, H. S. Heiman, and J. F. Keith, “Sensitivity and Specificity of Rapid Diagnostic Tests for Detection of Group B Streptococcal Antigen in Bacteremic Neonates,” *J. Clin. Microbiol.*, vol. 33, no. 1, pp. 193–198, 1995.

[46] M. S. Edwards, D. L. Kasper, H. J. Jennings, C. J. Baker, and A. Nicholson-Weller, “Capsular Sialic Acid Prevents Activation of the Altenative Complement Pathway by Type III, Group B Streptococci,” *J. Immunol.*, vol. 128, no. 3, pp. 1278–1283, 1982.

[47] M. R. Wessels, C. E. Rubens, V. J. Benedi, and D. L. Kasper, “Definition of a Bacterial Virulence Factor: Sialylation of the Group B Streptococcal Capsule,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 22, pp. 8983–8987, 1989.

[48] M. J. Severin and J. L. Wiley, “Changes in Susceptibility of Group B Streptococci to Penicillin G from 1968 Through 1975,” *Antimicrob. Agents Chemother.*, vol. 10, no. 2, pp. 380–381, 1976.

[49] C. S. F. Easmon, M. J. G. Hastings, J. Deeley, B. Bloxham, R. P. A. Rivers, and R. Marwood, “The Effect of Intrapartum Chemoprophylaxis on the Vertical Transmission of Group B Streptococci,” *BJOG An Int. J. Obstet. Gynaecol.*, vol. 90, no. 7, pp. 633–635, 1983.

[50] P. Yagupsky, M. A. Menegus, and K. R. Powell, “The Changing Spectrum of Group B Streptococcal Disease in Infants: an Eleven-Year Experience in a Tertiary Care Hospital,” *Pediatr. Infect. Dis. J.*, vol. 10, no. 11, pp. 801–808, 1991.

[51] V. L. Katz, M.-K. Moos, R. C. Cefalo, J. M. Thorp, W. A. Bowes, and S. D. Wells, “Group B Streptococci: Results of a Protocol of Antepartum Screening and Intrapartum Treatment,” *Am. J. Obstet. Gynecol.*, vol. 170, no. 2, pp. 521–526, 1994.

[52] CDC, “Adoption of Perinatal Group B Streptococcal Disease Prevention Recommendations by Prenatal-Care Providers- Conneticut and Minnesota, 1998,” *JAMA J. Am. Med. Assoc.*, vol. 283, no. 18, pp. 2384–2385, 2000.

[53] R. C. Lancefield, “Two Serological Types of Group B Hemolytic Streptococci with Related, but not Identical, Type-Specific Substances,” *J. Exp. Med.*, vol. 67, no. 1, pp. 25–40, 1938.

[54] C. J. Baker and D. L. Kasper, “Correlation of Maternal Antibody Deficiency with Susceptibility to Neonatal Group B Streptococcal Infection,” *N. Engl. J. Med.*, vol. 294, no. 14, pp. 753–756, 1976.

[55] L. C. Madoff, L. C. Paoletti, J. Y. Tai, and D. L. Kasper, “Maternal Immunization of Mice with Group B Streptococcal Type III Polysaccharide-Beta C Protein Conjugate Elicits Protective Antibody to Multiple Serotypes,” *J. Clin. Invest.*, vol. 94, no. 1, pp. 286–292, 1994.

[56] G. Kwatra, P. V. Adrian, T. Shiri, E. J. Buchmann, C. L. Cutland, and S. A. Madhi, “Natural Acquired Humoral Immunity Against Serotype-Specific Group B Streptococcus Rectovaginal Colonization Acquisition in Pregnant Women,” *Clin. Microbiol. Infect.*, vol. 21, no. 6, pp. 568.e13-568.e21, 2015.

[57] Z. Dangor *et al.*, “Correlates of Protection of Serotype-Specific Capsular Antibody and Invasive Group B Streptococcus Disease in South African Infants,” *Vaccine*, vol. 33, no. 48, pp. 6793–6799, 2015.

[58] C. J. Baker, M. S. Edwards, and D. L. Kasper, “Immunogenicity of Polysaccharides from Type III, Group B Streptococcus,” *J. Clin. Invest.*, vol. 61, no. 4, pp. 1107–1110, 1978.

[59] C. J. Baker and D. L. Kasper, “Group B Streptococcal Vaccines,” *Rev. Infect. Dis.*, vol. 7, no. 4, pp. 458–467, 1985.

[60] T. Lagergard, J. Shiloach, J. B. Robbins, and R. Schneerson, “Synthesis and Immunological Properties of Conjugates Composed of Group B Streptococcus Type III Capsular Polysaccharide Covalently Bound to Tetanus Toxoid,” *Infect. Immun.*, vol. 58, no. 3, pp. 687–694, 1990.

[61] M. R. Wessels *et al.*, “Immunogenicity in Animals of a Polysaccharide-Protein Conjugate Vaccine against Type III Group B Streptococcus,” *J. Clin. Invest.*, vol. 86, no. 5, pp. 1428–1433, 1990.

[62] L. C. Paoletti, M. R. Wessels, A. K. Rodewald, A. A. Shroff, H. J. Jennings, and D. L. Kasper, "Neonatal Mouse Protection against Infection with Multiple Group B Streptococcal (GBS) Serotypes by Maternal Immunization with a Tetravalent GBS Polysaccharide-Tetanus Toxoid Conjugate Vaccine," *Infect. Immun.*, vol. 62, no. 8, pp. 3236–3243, 1994.

[63] C. J. Baker, L. C. Paoletti, M. A. Rench, H. K. Guttormsen, M. S. Edwards, and D. L. Kasper, "Immune Response of Healthy Women to 2 Different Group B Streptococcal Type V Capsular Polysaccharide-Protein Conjugate Vaccines," *J. Infect. Dis.*, vol. 189, no. 6, pp. 1103–1112, 2004.

[64] E. Campisi *et al.*, "Serotype IV Streptococcus agalactiae ST-452 has Arisen from Large Genomic Recombination Events Between CC23 and the Hypervirulent CC17 Lineages," *Sci. Rep.*, vol. 6, no. June, pp. 1–11, 2016.

[65] S. A. Madhi *et al.*, "Safety and immunogenicity of an investigational maternal trivalent group B streptococcus vaccine in healthy women and their infants: a randomised phase 1b/2 trial," *Lancet Infect. Dis.*, vol. 16, no. 8, pp. 923–934, 2016, doi: 10.1016/S1473-3099(16)00152-3.

[66] E. T. Buurman *et al.*, "A Novel Hexavalent Capsular Polysaccharide Conjugate Vaccine (GBS6) for the Prevention of Neonatal Group B Streptococcal Infections by Maternal Immunization," *J. Infect. Dis.*, vol. 220, no. 1, pp. 105–115, 2019, doi: 10.1093/infdis/jiz062.

[67] J. Absalon *et al.*, "Safety and immunogenicity of a novel hexavalent group B streptococcus conjugate vaccine in healthy, non-pregnant adults: a phase 1/2, randomised, placebo-controlled, observer-blinded, dose-escalation trial," *Lancet Infect. Dis.*, vol. 21, no. 2, pp. 263–274, 2021, doi: 10.1016/S1473-3099(20)30478-3.

[68] F. Bloch, W. W. Hansen, and M. Packard, "Nuclear Induction," *Phys. Rev.*, vol. 69, no. 3–4, p. 127, 1946, doi: 10.1103/PhysRev.69.127.

[69] E. M. Purcell, H. C. Torrey, and R. V. Pound, "Resonance Absorbtion by Nuclear Magnetic Moments in a Solid," *Phys. Rev.*, vol. 69, no. 1–2, pp. 37–38, 1946, doi: 10.1103/PhysRev.69.37.

[70] J. W. Emsley and J. Feeney, "Forty years of Progress in Nuclear Magnetic Resonance Spectroscopy," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 50, no. 4, pp. 179–198, 2007, doi: 10.1016/j.pnmrs.2007.01.002.

[71] J. N. Shoolery, "The development of experimental and analytical high resolution NMR," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 28, no. 1, pp. 37–52, 1995, doi: 10.1016/0079-6565(95)01019-X.

[72] W. P. Aue, E. Bartholdi, and R. R. Ernst, “Two-dimensional spectroscopy. Application to nuclear magnetic resonance,” *J. Chem. Phys.*, vol. 64, no. 5, pp. 2229–2246, 1976.

[73] M. Mobli and J. C. Hoch, “Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 83, pp. 21–41, 2014, doi: 10.1016/j.pnmrs.2014.09.002.Nonuniform.

[74] K. F. Morris and C. S. Johnson, “Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy,” *Am. Chem. Soc.*, vol. 114, pp. 3141–3142, 1992, [Online]. Available: papers://3a44b00c-3356-4f71-8e28-59fa753ece5f/Paper/p1551.

[75] Y. Brummer and S. W. Cui, “Understanding Carbohydrate Analysis.,” in *Food Carbohydrates: Chemistry, Physical Properties, and Applications*, 1st ed., S. W. Cui, Ed. Florida: CRC Press, 2005, pp. 67–99.

[76] I. Ciucanu and F. Kerek, “A simple and rapid method for the permethylation of carbohydrates,” *Carbohydr. Res.*, vol. 131, no. 2, pp. 209–217, Jan. 1984.

[77] H. N. Englyst and J. H. Cummings, “Simplified method for the measurement of total non-starch polysaccharides by gas - liquid chromatography of constituent sugars as alditol acetates,” *Analyst*, vol. 109, no. 7, pp. 937–942, 1984, doi: 10.1039/an9840900937.

[78] B. Lindberg, J. Lönngren, and S. Svensson, “Specific degradation of polysaccharides,” *Adv. Carbohydr. Chem. Biochem.*, vol. 31, no. C, pp. 185–240, 1975, doi: 10.1016/S0065-2318(08)60297-8.

[79] J. N. BeMiller, *Methods in Carbohydrate Chemistry, Vol 10: Enzymic Methods*. Wiley-Interscience, 1995.

[80] X. Lemercinier and C. Jones, “Full 1H NMR assignment and detailed O-acetylation patterns of capsular polysaccharides from Neisseria meningitidis used in vaccine production,” *Carbohydr. Res.*, vol. 296, no. 1–4, pp. 83–96, 1996, doi: 10.1016/S0008-6215(96)00253-4.

[81] C. Jones, C. Whitley, and X. Lemercinier, “Full NMR assignment of the proton and carbon NMR spectra and revised structure for the capsular polysaccharide from Streptococcus pneumoniae type 17F,” *Carbohydr. Res.*, vol. 325, pp. 192–201, 2000.

[82] S. W. Cui, “Structural Analysis of Polysaccharides,” in *Food Carbohydrates: Chemistry, Physical Properties, and Applications*, 1st ed., S. W. Cui, Ed. Florida: CRC Press, 2005, pp. 105–160.

[83] V. Pinto and F. Berti, “Exploring the Group B Streptococcus capsular polysaccharides : The structural diversity provides the basis for development of NMR-based identity assays,” *J.*

Pharm. Biomed. Anal., vol. 98, pp. 9–15, 2014, doi: 10.1016/j.jpba.2014.05.004.

[84] J. R. Brisson *et al.*, "NMR and molecular dynamics studies of the conformational epitope of the type III group B Streptococcus capsular polysaccharide and derivatives," *Biochemistry*, vol. 36, no. 11, pp. 3278–3292, 1997, doi: 10.1021/bi961819l.

[85] A. S. Perlin and B. Casu, "Spectroscopic methods," in *The Polysaccharides*, Vol 1., G. O. Aspinall, Ed. New York: Academic Press, 1982, pp. 133–193.

[86] C. Jones and B. Mulloy, "The Application of Nuclear Magnetic Resonance to Structural Studies of Polysaccharides," in *Methods in Molecular Biology*, vol. 17, C. Jones, B. Mulloy, and A. H. Thomas, Eds. Totowa, NJ: Humana Press, 1993, pp. 149–167.

[87] N. Ravenscroft and F. Berti, "NMR characterization of bacterial glycans and glycoconjugate vaccines," in *Recent Trends in Carbohydrate Chemistry*, Vol. 2., A. P. Rauter, B. E. Christensen, L. Somsák, P. Kosma, and R. Adamo, Eds. Amsterdam: Elsevier, 2020, pp. 239–281.

[88] W. A. Bubb, "NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity," *Concepts Magn. Reson. Part A Bridg. Educ. Res.*, vol. 19, no. 1, pp. 1–19, 2003, doi: 10.1002/cmr.a.10080.

[89] J. F. G. Vliegenthart, L. Dorland, H. van Halbeek, and J. Haverkamp, "NMR Spectroscopy of Sialic Acids," in *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. Vienna and New York: Springer, 1982, pp. 127–172.

[90] D. M. Doddrell, D. T. Pegg, and M. R. Bendall, "Distortionless enhancement of NMR signals by polarization transfer," *J. Magn. Reson.*, vol. 48, no. 2, pp. 323–327, 1982, doi: 10.1016/0022-2364(82)90286-4.

[91] L. Lerner and A. Bax, "Application of new, high-sensitivity, 1H-13C-N.M.R.-spectral techniques to the study of oligosaccharides," *Carbohydr. Res.*, vol. 166, no. 1, pp. 35–46, 1987, doi: 10.1016/0008-6215(87)80042-3.

[92] M. Guéron, P. Plateau, and M. Decorps, "Solvent signal suppression in NMR," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 23, no. 2, pp. 135–209, 1991, doi: 10.1016/0079-6565(91)80007-O.

[93] C. S. Johnson, "Diffusion ordered nuclear magnetic resonance spectroscopy: Principles and applications," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 34, no. 3–4, pp. 203–256, 1999, doi: 10.1016/S0079-6565(99)00003-5.

[94] A. Bax and R. Freeman, "Investigation of complex networks of spin-spin coupling by two-

dimensional NMR," *J. Magn. Reson.*, vol. 44, no. 3, pp. 542–561, 1981, doi: 10.1016/0022-2364(81)90287-0.

[95] S. W. Homans, R. A. Dwek, J. Boyd, N. Soffe, and T. W. Rademacher, "A method for the rapid assignment of 1H NMR spectra of oligosaccharides using homonuclear Hartmann-Hahn spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, no. 5, pp. 1202–1205, 1987, doi: 10.1073/pnas.84.5.1202.

[96] C. Dalvit and G. Bovermann, "Pulsed field gradient one-dimensional NMR selective ROE and TOCSY experiments," *Magn. Reson. Chem.*, vol. 33, no. 2, pp. 156–159, 1995, doi: 10.1002/mrc.1260330214.

[97] S. W. Homans, R. A. Dwek, J. Boyd, M. Mahmoudian, W. G. Richards, and T. W. Rademacher, "Conformational Transitions in N-Linked Oligosaccharides," *Biochemistry*, vol. 25, no. 20, pp. 6342–6350, 1986, doi: 10.1021/bi00368a076.

[98] J. Stonehouse, P. Adell, A. J. Shaka, and J. Keeler, "Ultrahigh-Quality NOE Spectra," *J. Am. Chem. Soc.*, vol. 116, no. 13, pp. 6037–6038, 1994, doi: 10.1021/ja00092a092.

[99] N. H. Andersen, H. L. Eaton, and X. Lai, "Quantitative small molecule NOESY. A practical guide for derivation of cross-relaxation rates and internuclear distances," *Magn. Reson. Chem.*, vol. 27, no. 6, pp. 515–528, 1989, doi: 10.1002/mrc.1260270603.

[100] L. Lerner and A. Bax, "Sensitivity-enhanced two-dimensional heteronuclear relayed coherence transfer NMR spectroscopy," *J. Magn. Reson.*, vol. 69, no. 2, pp. 375–380, 1986, doi: 10.1016/0022-2364(86)90091-0.

[101] R. Wagner and S. Berger, "Heteronuclear edited gradient selected 1D and 2D NOE spectra: Determination of the NOE effect between chemically equivalent protons," *Magn. Reson. Chem.*, vol. 35, no. 3, pp. 199–202, 1997, doi: 10.1002/(SICI)1097-458X(199703)35:3<199::AID-OMR55>3.0.CO;2-1.

[102] X. Lemercinier and C. Jones, "An NMR spectroscopic identity test for the control of the capsular polysaccharide from Haemophilus influenzae type b," *Biologicals*, vol. 28, no. 3, pp. 175–183, 2000, doi: 10.1006/biol.2000.0255.

[103] X. Lemercinier, I. Martinez-Cabrera, and C. Jones, "Use and validation of an NMR test for the identity and O-acetyl content of the Salmonella typhi VI capsular polysaccharide vaccine," *Biologicals*, vol. 28, no. 1, pp. 17–24, 2000, doi: 10.1006/biol.1999.0238.

[104] C. Jones and X. Lemercinier, "Use and validation of NMR assays for the identity and O-acetyl content of capsular polysaccharides from Neisseria meningitidis used in vaccine

manufacture," *J. Pharm. Biomed. Anal.*, vol. 30, no. 4, pp. 1233–1247, 2002, doi: 10.1016/S0731-7085(02)00462-4.

[105] C. Abeygunawardana, T. C. Williams, J. S. Sumner, and J. P. Hennessey, "Development and validation of an NMR-based identity assay for bacterial polysaccharides," *Anal. Biochem.*, vol. 279, no. 2, pp. 226–240, 2000, doi: 10.1006/abio.1999.4470.

[106] A. Bardotti *et al.*, "Physicochemical characterisation of glycoconjugate vaccines for prevention of meningococcal diseases," *Vaccine*, vol. 26, no. 18, pp. 2284–2296, 2008, doi: 10.1016/j.vaccine.2008.01.022.

[107] WHO Expert Committee on Biological Standardization, "Recommendations for the production and control of pneumococcal conjugate vaccines," *WHO Tech. Rep. Ser.*, vol. 927, pp. 64–98, 2005, [Online]. Available: http://whqlibdoc.who.int/trs/WHO_TRS_927_eng.pdf.

[108] T. M. Rivers and L. A. Kohn, "The Biological and the Serological Reactions of Influenza Bacillia Broducing Meningitis," *J. Expermiental Med.*, vol. 34, no. 477–494, pp. 477–494, 1921.

[109] S. H. W. Sell, R. E. Merrill, E. O. Doyne, and E. P. Zimsky, "Long-term Sequelae of Hemophilus Influenzae Meningitis," *Pediatrics*, vol. 49, no. 2, pp. 206–211, 1972.

[110] J. B. Robbins, R. Schneerson, P. Anderson, and D. H. Smith, "Prevention of Systemic Infections, Especially Meningitis, Caused by Haemophilus influenzae Type b," *JAMA*, vol. 276, no. 14, pp. 1181–1185, 1996, doi: 10.1001/jama.1996.03540140069031.

[111] H. Peltola, H. Käyhty, M. Virtanen, and P. H. Mäkelä, "Prevention of Hemophilus influenzae Type B Bacteremic Infections with the Capsular Polysaccharide Vaccine," *N. Engl. J. Med.*, vol. 310, no. 24, pp. 1561–1566, 1984, doi: 10.1056/NEJM198406143102404.

[112] J. Eskola, "Use of conjugate vaccines to prevent meningitis caused by Haemophilus influenzae type b or Streptococcus pneumoniae," *J. Hosp. Infect.*, vol. 30, pp. 313–321, 1995.

[113] O. T. Avery and W. F. Goebel, "Chemo-immunological studies on conjugated carbohydrate-proteins: II. Immunological specificity of synthetic sugar-protein antigens," *J. Exp. Med.*, vol. 50, no. 4, pp. 533–550, 1929.

[114] A. W. Sturgess *et al.*, "Haemophilus influenzae type b conjugate vaccine stability: Catalytic depolymerization of PRP in the presence of aluminum hydroxide," *Vaccine*, vol. 17, no. 9–10, pp. 1169–1178, 1999, doi: 10.1016/S0264-410X(98)00337-5.

[115] Y. Aubin, C. Jones, and D. I. Freedberg, "Using NMR spectroscopy to obtain the higher order structure of biopharmaceutical products," *BioPharm Int.*, vol. 23, no. 8 SUPPL., pp. 28–34,

2010.

[116] B. Greenwood, “Meningococcal meningitis in Africa,” *Trans. R. Soc. Trop. Med. Hyg.*, vol. 93, pp. 341–353, 1999.

[117] F. F. Schwentker, S. Gelman, and P. H. Long, “The Treatment of Meningococcic Meningitis with Sulfanilamide: Preliminary Report,” *J. Am. Med. Assoc.*, vol. 108, no. 17, pp. 1407–1408, 1937.

[118] M. S. Artenstein, R. Gold, J. G. Zimmerly, F. A. Wyle, H. Schneider, and C. Harkins, “Prevention of Meningococcal Disease by Group C Polysaccharide Vaccine,” *N. Engl. J. Med.*, vol. 282, no. 8, pp. 417–420, 1970.

[119] I. Goldschneider, E. C. Gotschlich, and M. S. Artenstein, “Human Immunity to the Meningococcus II. Development of Natural Immunity,” *J. Exp. Med.*, vol. 129, no. 6, pp. 1327–1348, 1969.

[120] J. Souza de Morais, R. S. Munford, J. Baptista Risi, E. Antezana, and R. A. Feldman, “Epidemic disease due to serogroup C Neisseria meningitidis in Sao Paulo, Brazil,” *J. Infect. Dis.*, vol. 129, no. 5, pp. 568–571, 1974, doi: 10.1093/infdis/129.5.568.

[121] R. Garrido *et al.*, “Quantitative Proton Nuclear Magnetic Resonance evaluation and total assignment of the capsular polysaccharide Neisseria meningitidis serogroup X,” *J. Pharm. Biomed. Anal.*, vol. 70, pp. 295–300, 2012, doi: 10.1016/j.jpba.2012.07.014.

[122] G. M. Sternberg, “The Etiology of Croupous Pneumonia,” *Lancet*, vol. 133, pp. 370–371, 1889.

[123] A. E. Wright *et al.*, “Observations on Prophylactic Inoculation Against Pneumococcus Infections and on the Results which have been Achieved by it,” *Lancet*, 1914.

[124] R. Austrian, “Of Gold and Pneumococci: A History of Pneumococcal Vaccines in South Africa,” *Trans. Am. Clin. Climatol. Assoc.*, vol. 89, pp. 141–161, 1978, doi: 10.9783/9781512800135-010.

[125] J. Eskola and M. Antilla, “Pneumococcal Conjugate Vaccines,” *Pediatr. Infect. Dis. J.*, vol. 18, no. 6, pp. 543–551, 1999.

[126] K. M. Boyer and S. P. Gotoff, “Prevention of Early-Onset Neonatal Group B Streptococcal Disease With Selective Intrapartum Chemoprophylaxis,” *Pediatr. Infect. Dis. J.*, vol. 6, no. 1, p. 90, 1987, doi: 10.1097/00006454-198701000-00045.

[127] D. L. Kasper *et al.*, “Immune response to type III group B streptococcal polysaccharide-tetanus toxoid conjugate vaccine,” *J. Clin. Invest.*, vol. 98, no. 10, pp. 2308–2314, 1996, doi:

10.1172/JCI119042.

[128] C. J. Baker, M. A. Rench, M. Fernandez, L. C. Paoletti, D. L. Kasper, and M. S. Edwards, “Safety and immunogenicity of a bivalent group B streptococcal conjugate vaccine for serotypes II and III,” *J. Infect. Dis.*, vol. 188, no. 1, pp. 66–73, 2003, doi: 10.1086/375536.

[129] S. A. Madhi *et al.*, “Considerations for a phase-III trial to evaluate a group B Streptococcus polysaccharide-protein conjugate vaccine in pregnant women for the prevention of early- and late-onset invasive disease in young-infants,” *Vaccine*, vol. 31, no. S4, pp. D52–D57, 2013, doi: 10.1016/j.vaccine.2013.02.029.

[130] H. J. Jennings, E. Katzenellenbogen, C. Lugowski, and D. L. Kasper, “Structure of Native Polysaccharide Antigens of Type Ia and Type Ib Group B Streptococcus,” *Biochemistry*, vol. 22, no. 5, pp. 1258–1264, 1983, doi: 10.1021/bi00274a042.

[131] H. J. Jennings, K.-G. Rosell, E. Katzenellenbogen, and D. L. Kasper, “Structural Determination of the Capsular Polysaccharide Antigen of Type II Group B Streptococcus,” *J. Biol.*, vol. 258, no. 3, pp. 1793–1798, 1983.

[132] H. J. Jennings, K. G. Rosell, and D. L. Kasper, “Structural determination and serology of the native polysaccharide antigen of type-III group B Streptococcus,” *Can. J. Biochem.*, vol. 58, no. 2, pp. 112–120, 1980, doi: 10.1139/o80-016.

[133] J. L. Di Fabio *et al.*, “Structure of the capsular polysaccharide antigen of type IV group B Streptococcus,” *Can. J. Chem.*, vol. 67, no. 5, pp. 877–882, May 1989, doi: 10.1139/v89-136.

[134] M. R. Wessels *et al.*, “Structural determination and immunochemical characterization of the type V group B Streptococcus capsular polysaccharide,” *J. Biol. Chem.*, vol. 266, no. 11, pp. 6714–6719, 1991, doi: 10.1074/jbc.271.15.8786.

[135] G. Kogan *et al.*, “Structure of the type VI group B Streptococcus capsular polysaccharide determined by high resolution NMR spectroscopy,” *J. Carbohydr. Chem.*, vol. 13, no. 8, pp. 1071–1078, 1994, doi: 10.1080/07328309408011849.

[136] G. Kogan, J. R. Brisson, D. L. Kasper, C. von Hunolstein, G. Orefici, and H. J. Jennings, “Structural elucidation of the novel type VII group B Streptococcus capsular polysaccharide by high resolution NMR spectroscopy,” *Carbohydr. Res.*, vol. 277, no. 1, pp. 1–9, 1995, doi: 10.1016/0008-6215(95)00195-Y.

[137] G. Kogan *et al.*, “Structural and immunochemical characterization of the type VIII group B Streptococcus capsular polysaccharide,” *J. Biol. Chem.*, vol. 271, no. 15, pp. 8786–8790, 1996, doi: 10.1074/jbc.271.15.8786.

[138] F. Berti *et al.*, “Structure of the type IX Group B Streptococcus capsular polysaccharide and its evolutionary relationship with types V and VII,” *J. Biol. Chem.*, vol. 289, no. 34, pp. 23437–23448, 2014, doi: 10.1074/jbc.M114.567974.

[139] L. Deng, D. L. Kasper, T. P. Krick, and M. R. Wessels, “Characterization of the linkage between the type III capsular polysaccharide and the bacterial cell wall of group B streptococcus,” *J. Biol. Chem.*, vol. 275, no. 11, pp. 7497–7504, 2000, doi: 10.1074/jbc.275.11.7497.

[140] S. Mehta and D. M. Whitfield, “Polymer-supported synthesis of a branched trisaccharide of the type IA Group B Streptococcus capsular polysaccharide: 3-Iodo-4-methoxybenzyl as a new O-protecting group,” *Tetrahedron*, vol. 56, no. 35, pp. 6415–6425, 2000, doi: 10.1016/S0040-4020(00)00612-8.

[141] P. K. Mondal, G. Liao, M. A. Mondal, and Z. Guo, “Chemical synthesis of the repeating unit of type Ia group B streptococcus capsular polysaccharide,” *Org. Lett.*, vol. 17, no. 5, pp. 1102–1105, 2015, doi: 10.1021/ol5036563.

[142] H. Zhang, S. Zhou, Y. Zhao, and J. Gao, “Chemical synthesis of the dimeric repeating unit of type Ia group B: Streptococcus capsular polysaccharide,” *Org. Biomol. Chem.*, vol. 17, no. 23, pp. 5839–5848, 2019, doi: 10.1039/c9ob01024f.

[143] L. Del Bino *et al.*, “Regioselective Glycosylation Strategies for the Synthesis of Group Ia and Ib Streptococcus Related Glycans Enable Elucidating Unique Conformations of the Capsular Polysaccharides,” *Chem. - A Eur. J.*, vol. 25, no. 71, pp. 16277–16287, 2019, doi: 10.1002/chem.201903527.

[144] L. Shao *et al.*, “Chemical Synthesis of the Repeating Unit of Type II Group B Streptococcus Capsular Polysaccharide,” *J. Org. Chem.*, vol. 83, no. 11, pp. 5920–5930, 2018, doi: 10.1021/acs.joc.8b00396.

[145] M. R. Wessels, V. Pozsgay, D. L. Kasper, and H. J. Jennings, “Structure and Immunochemistry of an Oligosaccharide Repeating Unit of the Capsular Polysaccharide of Type III Group B Streptococcus,” *J. Biol. Chem.*, vol. 262, no. 17, pp. 8262–8267, 1987.

[146] V. Pozsgay, J.-R. Brisson, and H. J. Jennings, “Synthesis of a tri- and a tetra-saccharide fragment of the capsular polysaccharide of type III group B Streptococcus,” *Carbohydr. Res.*, vol. 205, no. C, pp. 133–146, 1990, doi: 10.1016/0008-6215(90)80134-O.

[147] V. Pozsgay, J. Gaudino, J. C. Paulson, and H. J. Jennings, “Chemo-enzymatic synthesis of a branching decasaccharide fragment of the capsular polysaccharide of type III group B Streptococcus,” *Bioorg. Med. Chem. Lett.*, vol. 1, no. 8, pp. 391–394, 1991.

[148] V. Pozsgay, J.-R. Brisson, H. J. Jennings, S. Allen, and J. C. Paulson, “Combined Chemical and Enzymatic Synthesis of a Pentasaccharide Repeating Unit of the Capsular Polysaccharide of Type III Group B Streptococcus and One- and Two-Dimensional NMR Spectroscopic Studies,” *J. Org. Chem.*, vol. 56, no. 10, pp. 3377–3385, 1991, doi: 10.1021/jo00010a037.

[149] W. Zou, J.-R. Brisson, Q.-L. Yang, M. Van Der Zwan, and H. J. Jennings, “Synthesis and NMR assignment of two repeating units (decasaccharide) of the type III group B Streptococcus capsular polysaccharide and its 13C-labeled and N-propionyl substituted sialic acid analogues,” *Carbohydr. Res.*, vol. 295, no. 39525, pp. 209–228, 1996, doi: 10.1016/S0008-6215(96)00236-4.

[150] V. Cattaneo *et al.*, “Synthesis of Group B Streptococcus type III polysaccharide fragments for evaluation of their interactions with monoclonal antibodies,” *Pure Appl. Chem.*, vol. 89, no. 7, pp. 855–875, 2017, doi: 10.1515/pac-2016-0918.

[151] J. Gao and Z. Guo, “Chemical Synthesis of the Repeating Unit of Type V Group B Streptococcus Capsular Polysaccharide,” *Org. Lett.*, vol. 18, no. 21, pp. 5552–5555, 2016, doi: 10.1021/acs.orglett.6b02796.

[152] M. Lundborg and G. Widmalm, “Structural analysis of glycans by NMR chemical shift prediction,” *Anal. Chem.*, vol. 83, no. 5, pp. 1514–1517, 2011, doi: 10.1021/ac1032534.

[153] S. Beri, D. Gandhi, and N. Ravenscroft, “Use of NMR as an analytical tool in the process development of conjugate vaccines against Haemophilus influenzae type b (Hib) and meningococcal serogroup A (MenA),” *Biologicals*, vol. 62, no. August, pp. 102–106, 2019, doi: 10.1016/j.biologicals.2019.10.005.

[154] F. Berti and N. Ravenscroft, “Characterization of Carbohydrate Vaccines by NMR Spectroscopy,” in *Carbohydrate-Based Vaccines*, B. Lepenies, Ed. New York: Humana Press, 2015, pp. 189–209.

[155] H. J. Jennings, K. G. Rosell, and D. L. Kasper, “Structure and serology of the native polysaccharide antigen of type Ia group B streptococcus,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 77, no. 5 I, pp. 2931–2935, 1980, doi: 10.1073/pnas.77.5.2931.

[156] E. Campisi *et al.*, “Group B Streptococcus chimeric capsular polysaccharides as novel multivalent vaccine candidates,” *Glycoconj. J.*, vol. 38, no. 4, pp. 447–457, 2021, doi: 10.1007/s10719-021-10000-4.

[157] M. R. Wessels, V.-J. Benedi, H. J. Jennings, F. Michon, J. L. DiFabio, and D. L. Kasper, “Isolation and characterization of type IV group B Streptococcus capsular polysaccharide,” *Infect.*

Immun., vol. 57, no. 4, pp. 1089–1094, 1989, doi: 10.1128/iai.57.4.1089-1094.1989.