Analysis of Complex Polysaccharides of Biological and Industrial Relevance via Nuclear Magnetic Resonance Spectroscopy: Structure, Modification and Protein Interactions.

Thesis submitted in accordance with the requirements of the University of Liverpool

for the degree of Doctor of Philosophy by:

James Andrew London

March 2022

Acknowledgements

Firstly, I would like to thank my University of Liverpool supervisors, Dr. Ed Yates and Dr. Igor Barsukov for their never-ending support and patience throughout my PhD, and particularly in these last few months. You are both always there to help if ever I ask for it, no matter what it was or how busy you are. I really couldn't have asked for two better supervisors!

Secondly, to my Unilever supervisors, Dr. Jane Whittaker, Dr. Sue Rogers, and Dr. Craig Fairgrieve, thank you for all you have done for me, for your guidance, and for giving me the opportunity to experience working in industry, for which I will be forever grateful.

As well as my supervisors I have to thank two friends and colleagues, Emily Wang and Dr. Rudi Grossman, without whom I could never have completed this project. Emily, beyond all of your help in the lab and advice on protein production, your support and friendship has been, and remains, appreciated more than you know. Rudi, thank you for all the script making, NMR discussions, and help in the lab, without it I would have struggled massively. But mostly, thank you for the advice and encouragement you've given me throughout this project and beyond. I am glad I've been able to spend the past four and a half years working with both of you!!

To everyone in the NMR Centre and Lab C thank you for everything; your help and support has always been and will remain greatly appreciated! In particular, to Prof. Lu-Yun Lian and Dr. Marie Phelan for offering to help with NMR if ever I asked you. Also to Zain Ghanameh and Michele Fresneda Alarcon for our conversations during long days of data analysis which helped me get through this (relatively) sane.

Finally, to my parents, brother, family, and friends for your support and encouragement which both allowed me to start this PhD and kept me going throughout it. I owe you all massively!

Abstract

Carbohydrates, as one of the principal classes of biological molecule, are ubiquitous, with natural roles and biopharmaceutical, commercial, and industrial applications resulting from their structures. Variations to these structures have the potential to generate vastly differing biological, physico-chemical, and interaction properties as required. One such structural variation, sulfation, has been observed to affect the ability of multiple carbohydrates to bind proteins whilst also having the potential to alter their commercially relevant physico-chemical properties. The characterisation of carbohydrate structures including any differences present between samples from different sources, and the effect that modifications have is an important asset during the development or analysis of novel samples. The research detailed in this thesis, therefore, was targeted towards the optimisation of NMR-based methods for the analysis of, both starting material and modified, carbohydrate structures. Additional targets included the optimisation of a pyridine sulfur trioxide based sulfation method in recyclable solvents, and the analysis of resulting samples in terms of their interactions with proteins.

Improvements were, therefore, made to the nuclear magnetic resonance based protocol for carbohydrate structure characterisation with an increase in resolution approaching, and in some cases at, the natural linewidth. These spectrum resolution improvements generated a greater number of peaks than was previously possible, which were assigned specifically to the individual CH groups present, in 2D ^{1}H,^{13}C-HSQC spectra, for the repeating units of a xyloglucan test-case sample.

Application of this protocol to a series of sulfated samples provided a means by which modification could be, indirectly, monitored more specifically for individual hydroxyl groups. The development of an improved sulfation method, suitable for gel forming and water insoluble carbohydrates, resulted in the production of a range of samples that could be used to test the NMR protocol developed. The amended pyridine sulfur trioxide sulfation method developed, with solvation in 1-butyl-3-methylimidazolium chloride instead of the traditional organic solvents, generated a series of variably sulfated samples. In some instances these reactions followed the expected order of hydroxyl group reactivity, i.e. primary -OHs > equatorial secondary -OHs > axial secondary -OHs, however in other reactions modifications did not follow this order. Those reactions that differed from the expected order of hydroxyl group reactivity highlight the potential applicability of this method for the generation of carbohydrates containing novel sulfation patterns.

Carbohydrate-protein interactions are involved in many of the biological processes throughout all forms of life. A selection of the sulfated samples produced and analysed were therefore screened to determine their interactions with four proteins, Fibroblast Growth Factors (FGFs) 1 and 2 as well as Cyclophilins (Cyps) B and D, using differential scanning fluorimetry. Of the combinations screened, examples that were shown to form interactions, were analysed via isothermal titration calorimetry to provide thermodynamic data and corroboration of the DSF.

The research carried out in this project, therefore, could be used to analyse carbohydrate samples in greater detail than was previously possible, in a manner proven by the analysis of samples sourced from different suppliers, as well as those produced by chemical modification. Equally, the research presented in this thesis has identified a method by which gel forming and water insoluble samples could be modified without requiring a preparation step, that has been shown to produce novel sulfation pattern samples. Screening the samples produced; they have subsequently been shown to bind proteins in interactions that have been previously identified as well as those that haven’t. These could form the basis of any further work that could be carried out.

Table of Contents

List of Abbreviations

AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride
Ara	Arabinose
APS	Ammonium Persulfate
ATR	Attenuated Total Reflectance
$BMImBF_4$	1-butyl-3-methylimidazolium tetrafluoroborate
BMImCl	1-butyl-3-methylimidazolium chloride
COSY	Correlation Spectrscopy
CPD	Composite Pulse Decoupling
Cyp	Cyclophilin
CS	Compressed Sensing
DEPT	Distortionless Enhancement by Polarisation Transfer
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DOSY	Diffusion Ordered Spectroscopy
DSS	Sodium trimethylsilylpropanesulfonate
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ELSD	Evaporative Light Scattering Detection
ESI	Electrospray Ionisation
FGF	Fibroblast Growth Factor
FGFR	Fibroblast Growth Factor Receptor
FPLC	Fast Protein Liquid Chromatography
FTIR	Fourier Transform Infrared Spectroscopy
Gal	Galactose
Glc	Glucose
GlcA	Glucuronic Acid
HILIC	Hydrophobic Interaction Liquid Chromatography
HMBC	Heteronuclear Multiple Bond Correlation
HPAEC	High Performance Anion Exchange Chromatography
HPLC	High Performance Liquid Chromatography
HPSEC	High Performance Size Exclusion Chromatography
HSQC	Heteronuclear Single Quantum Correlation
INEPT	Insensitive Nuclei Enhancement by Polarisation Transfer
ITC	Isothermal Titration Calorimetry
MDD	Multidimensional Decomposition
MS	Mass Spectrometry
MWCO	Molecular Weight Cut Off
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
$Pyr{\cdot}SO_3$	Pyridine Sulfur Trioxide
SEC	Size Exclusion Chromatography
SDS	Sodium Dodecyl Sulfate
PAD	Pulsed Amperometric Detection
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction

ppm	parts per million
RID	Refractive Index Detection
ROESY	Rotating-frame Overhauser Effect Spectroscopy
Ser	Serine
TEMED	N,N,N’,N’-tetramethylethane-1,2-diamine
TEV	Tobacco Etch Virus
TOCSY	Total Correlation Spectroscopy
UV	Ultraviolet
Xyl	Xylose

Chapter 1 – General Introduction

1.1 – Overview:

1.1.1 – History:

Humans have used carbohydrates since prehistory, when honey and sugarcane were widely applied as sweeteners, cancer treatments and antimicrobial agents (Cilliers and Retief, 2008; Karpozilos and Pavlidis, 2004; Namias, 2003). The development of novel applications for carbohydrates continued throughout the ancient and medieval periods, to include roles as texturing agents in cosmetics and during the production of papyrus, as well as its successor – paper (Castillo-Valdivia et al., 2014; Righini, 1998; Evershed et al., 2004; Wiedemann and Bayer, 1983; Corsaro et al., 2013). Applications for crystalline sugars in the form of sweeteners developed from around A.D.300 spreading to China and Egypt in the following centuries and onwards to North Africa, Europe, and far later the Americas (Pigman, 1957).

Increasingly widespread applications were developed down to the late 18th Century resulting in the extraction of beet and sugarcane's constituent monosaccharides. The discovery that identical residues were present initiated industrial carbohydrate extraction with methods utilising both starch hydrolysis and monosaccharide isolation from natural products, initially glucose isolated from honey (Fischer, 1989; Lentze, 2018; Pigman, 1957). During this time fledgling structural analysis of the isolated carbohydrates focussed on their ability to rotate polarised light, where glucose was shown to polarise light to the right, in contrast to fructose where light was polarised to the left.

Carbohydrate characterisation method development advanced during the 19th Century to provide greater detail on both structure and stereochemistry. These advances enabled the identification of the first glucose epimer, mannose, via phenylhydrazine reactions while the relationship between L-arabinose, D-glucose, D-galactose, and L-fructose residues were simultaneously determined via application of cyanohydrin synthesis reactions. Subsequent application of these reactions characterised the complete series of glucose enantiomers as aldoses, sugar alcohols and sugar acids alongside those for ketoses (Lichtenthaler, 1992; Kunz, 2002).

1.1.2 – Naturally Occurring Carbohydrate Sources:

As one of the four principal classes of biological molecules, carbohydrates occur in myriad plant, microbial and mammalian cells beyond the beet and sugarcane initially used for isolation. Plants contain many mono-, di-, and oligosaccharides, while their cell walls represent the greatest naturally abundant source of polysaccharides. Currently only a small proportion of this vast resource is being utilised in some form of application (Pauly and Keegstra, 2008). Plant cell wall polysaccharides differ between species with those of land-based plants principally comprising cellulose complexed to hemicellulosic xyloglucan, pectins, arabinans and arabinogalactans (Cosgrove, 2005; Charrier et al., 2019). Cellulose and the hemicelluloses are also present in algal cell walls; however these matrices are supplemented with different polysaccharides to land-based plants, one notable example being the carrageenan family members (Charrier et al., 2019; Usov, 1998; Jiao et al., 2011).

As with plants carbohydrates are present in microbes with those derived from bacteria principally located on the cell surface whilst viral carbohydrates typically exist in the form of glycans attached to spike proteins (Imperiali, 2019; Raman et al., 2016). Bacterial carbohydrates such as the commercially relevant xanthan gum, located on the cytoplasmic membrane, as part of the cell wall, and extracellularly within the glycocalyx (Ahmad et al., 2015; Rosalam and England, 2006), are a principal target for antibacterial agents. Modifications to these polysaccharides have been seen to affect the mammalian immune response to infection (Imperiali, 2019; Avci and Kasper, 2020). Similarly, viral spike protein linked glycan carbohydrates are involved in their entry into target host cells whilst reducing the efficiency, or enabling the evasion, of the host's immune response, although they do provide a potential target for neutralising antibodies and carbohydrate based vaccines (Raman et al., 2016; Wintjens et al., 2020; Binley et al., 2010).

Finally, there are several forms of mammalian carbohydrate, such as the mono-, di-, and oligosaccharides used during glycolysis, as ABO blood group antigens, and as glycoproteins and glycolipids (Lunt et al., 2011; Milland and Sandrin, 2006; Reily et al., 2019; Hanafusa et al., 2020). Similarly, ribose and deoxyribose residues are constituent in ribonucleic and deoxyribonucleic acids respectively, governing the biological processes required in the mammalian, plant, and microbial sources of carbohydrates (Watson and Crick, 1953; Travers and Muskhelishvili, 2015). More complex polysaccharides also have functional roles in mammals such as the glycosaminoglycan family of proteoglycans, with members present on most, if not all, cell types (Rabenstein, 2002). Examples of such roles include within the immune response where glycosaminoglycans have a number of activities, including, regulating cell adhesion, modulating cytokine and chemokine functions, sensing tissue injuries, and acting as a barrier to leucocyte migration (Simon-Davis and Parish, 2013; Kamhi et al., 2013). These roles are largely facilitated by a number of protein interactions such as those with complement factors, interferons and interleukins, and lectins, to name a few examples (Blaum et al., 2010; Kamhi et al., 2013; Ori et al., 2008; Capila and Linhardt, 2002; Migliorini et al., 2015).

1.1.3 – Carbohydrate Applications:

The number and variety of carbohydrate applications has expanded dramatically from those used historically to incorporate wide ranging biopharmaceutical and commercial products in addition to their natural roles. Of these natural roles those related to structure and storage, as well as cell signalling and immunity have been most widely investigated alongside applications related the interactions between carbohydrates and proteins (Iijima and Hashizume, 2015; Kanungo et al., 2018; Cobb and Kasper, 2005; Huang et al, 2019). Alternatively, carbohydrates can be present on the surface of proteins, as N- or O-glycans, bound to the asparagine or serine and threonine residues to mediate their host protein's natural effects (Krasnova and Wong, 2016).

The commercial and biopharmaceutical applications developed from the properties of these natural carbohydrates include during drug delivery or as vaccines, anticoagulants, antitumour and antimicrobial agents (Astronomo and Burton, 2010; Poupard et al., 2017; Sousa et al., 2009). As biopharmaceuticals, carbohydrates are present in either their modified or native forms, one notable example being heparin which is applied as both its natural structure and as modified variants for diverse pharmaceutical applications (Oduah et al., 2016; Cassinelli and Naggi, 2016; Laurienzo, 2010). Similarly, carbohydrates' gelation, thickening and emulsion properties enable their incorporation in food, home, and personal care products such as the bacterial guar and xanthan gums or plant derived carrageenans with some restrictions depending on the type of product into which they are being applied (Laurienzo, 2010; Petri, 2015).

The diversification of carbohydrate uses throughout thousands of years of development, coupled with the myriad starting materials offered by mono-, di-, oligo- and polysaccharides found in all forms of life make carbohydrates an ideal starting point for the development of novel applications. This is further benefited by the advancement of methods used for their characterisation over the past 150 years which, whilst occurring more slowly than was the case for the other principal classes of biological molecules, are continually improving.

1.2 – Carbohydrate Structure and Nomenclature:

Commercial and industrial products that contain carbohydrates to mediate their physicochemical and protein interaction properties cover a range of applications and their effects are derived from the constituent carbohydrate's structure. The most basic form of structural variation exhibited by carbohydrates is the constituent monosaccharides from which they are comprised, and the differences exhibited between these residues. These variations, which generate a complex variety of conformationally different residues, include carbonyl carbon position, hydroxyl and final chiral atom stereochemistry, anomeric proton position, ring structure and conformation, and functional groups present. These are supplemented by the carbohydrate's chain length which can be subdivided into mono-, di-, oligo- and polysaccharides comprising one, two, three to ten, and more than ten constituent residues, respectively. This complexity is exacerbated during oligo- and polysaccharide formation, where very extensive structural variety is theoretically possible, and is increased further by post-polymerisation functional group addition, and higher order structural variation.

Carbohydrates exist as 3D structures comprising a series of stereoisomers, the simplest two of which – those for glyceraldehyde, are shown in Figure 1. The majority of the figures which follow, that are representative of carbohydrates will be drawn, for simplicity, in a 2D manner despite this ignoring the stereochemical relationship between functional groups present in the genuine 3D structure. Depicting the carbohydrates in this manner could lead to the impression that two stereoisomers give an identical structure if one is flipped by 180°, such as the lower depiction in Figure 1. This is not the case, as seen from the upper structures of the figure, where flipping the L-isomer by 180°, such that the hydroxyl groups of both compounds are pointing right, would leave the hydrogen atom behind the plane (represented by the aldehyde and hydroxyl groups) and the CH_2OH group in front of it, i.e. the opposite of that which occurs for the D-isomer, as seen in the diagram present in the figure.

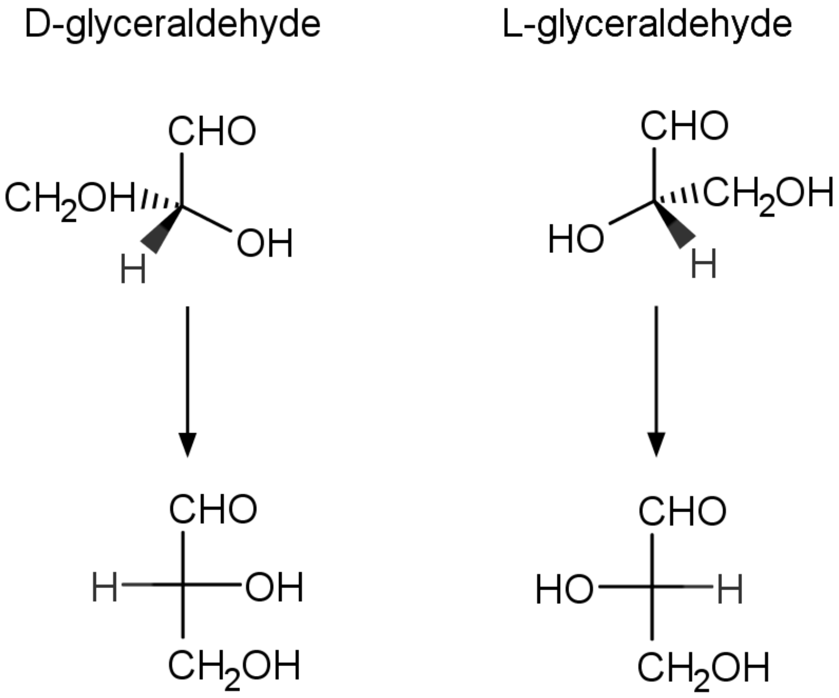


Figure 1: Showing one of the smallest carbohydrate chains, 3D and 2D glyceraldehyde structures in their D- and L- forms.

1.2.1 – Aldoses and Ketoses:

The acyclic carbohydrate chain's carbonyl carbon position is one of the principal methods by which the resulting cyclic structure differs. Incorporation of this carbonyl carbon at position C-1, as part of an aldehyde functional group, results in a class of monosaccharides called aldoses, Figure 2A. Conversely, the incorporation of this carbonyl as a ketone functional group, most commonly at position C-2 – although other locations are possible, generates the ketose class of residues, Figure 2B. These structures can then be further varied by the number of carbon atoms present within the acyclic chain as well as the stereochemistry of the secondary hydroxyl groups present (Lindhorst, 2003; Robyt, 1998).

The aldoses are a family of polyhydroxylated aldehydes, with the simplest – shown in Figure 1 and Figure 2A – being the triose glyceraldehyde, which is comprised of four functional groups, two of which contain hydroxyls, in a tetrahedral structure with a single chiral centre. Increasing chain length increases the potential stereoisomeric variants due to the greater number of chiral centres. Tetroses have two asymmetric centres generating four possible stereoisomeric variants rising to eight variants in pentoses, with three asymmetric centres, and sixteen variants in hexoses, with four asymmetric centres. Higher order aldoses such as heptoses and octoses exist, however they are much less common than those of chains comprised of between three and six carbon atoms (Lindhorst, 2003; Robyt, 1998).

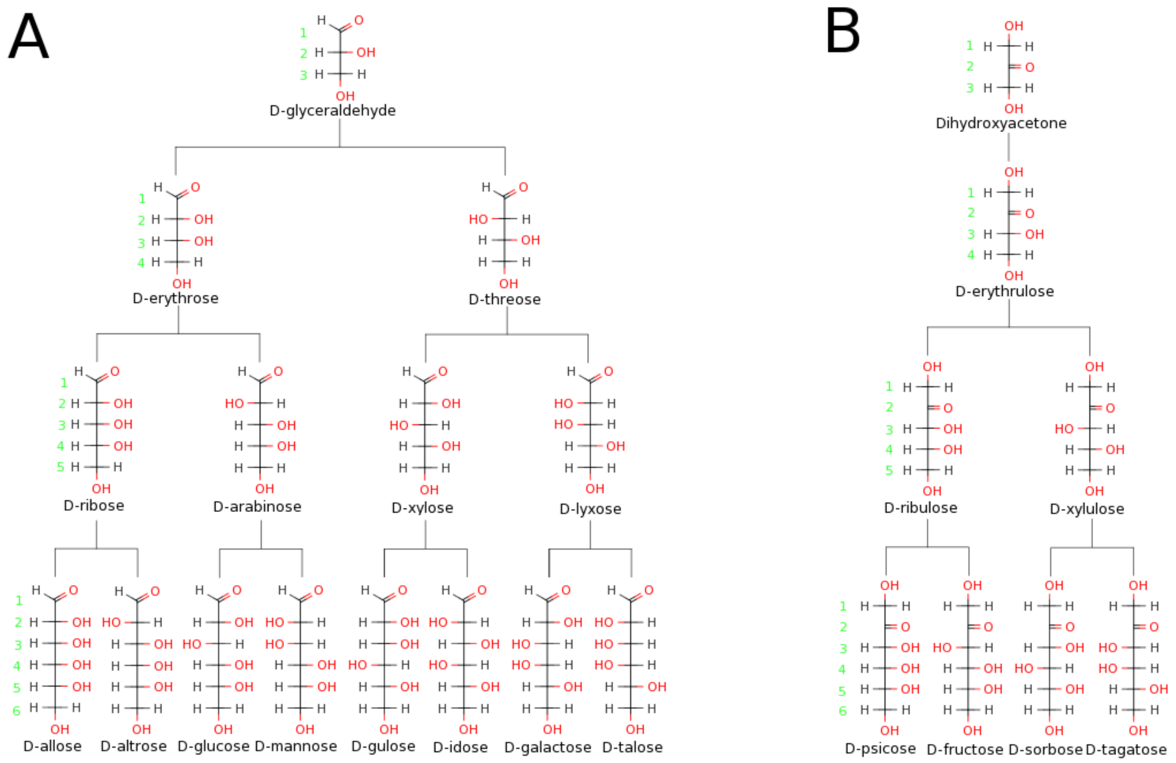


Figure 2: Fisher projections for acyclic aldose (A) and 2-ketose (B) variants between three and six carbon chain lengths.

Similarly, polyhydroxylated ketones form the ketose family of residues – shown in Figure 2B – comprised of multiple monosaccharide variants. These variants form in part due to the carbonyl carbon location which, whilst most common at C-2, can be placed on any of the acyclic chain's central carbon atoms. As with the aldoses, increasing chain length increases the number of chiral centres and consequently the number of potential stereoisomeric variants. However, having the carbonyl carbon present in the chain as a ketone rather than an aldehyde removes one of these centres, therefore a ketose of a specific chain length will have half the number of variants that the corresponding chain length aldose does, i.e. there are one triose, two tetroses, four pentoses and eight hexoses. Aldose and ketose residue variants of the same chain length cannot be distinguished solely based on their near identical molecular weight, for example the twenty-four hexose variants from the aldose and ketose families all have molecular weights around 180g/mol (Lindhorst, 2003; Robyt, 1998).

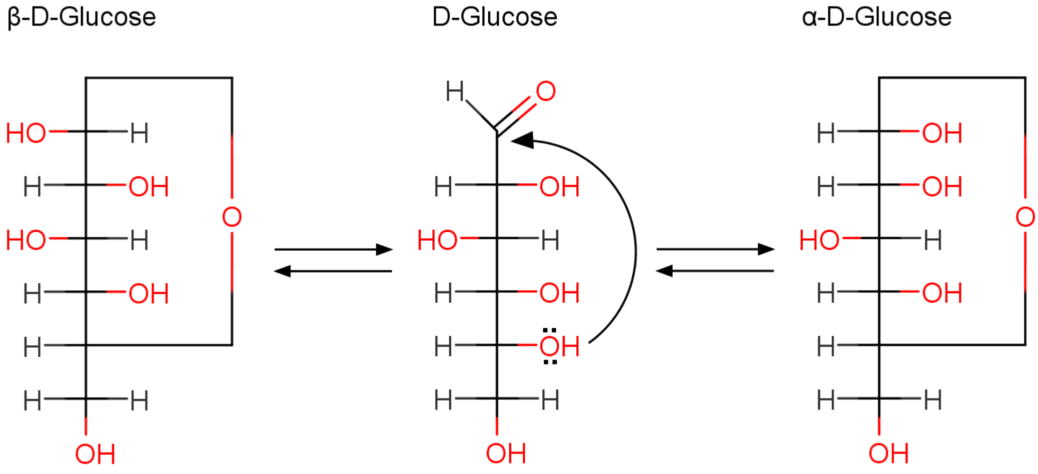


Figure 3: Fischer and Fischer-Tollens projections showing the cyclisation of D-glucose into the alpha and beta forms.

Aldose and ketose residue variants exist preferentially as cyclic structures, forming hemiacetals and hemiketals respectively via the nucleophilic attack of a hydroxyl group oxygen on the carbonyl carbon – Figure 3. For aldose residues the hydroxyl group oxygen which attacks the carbonyl carbon dictates the resulting ring structure, i.e. attack from the C-4 linked oxygen forms a furanose and the C-5 linked oxygen a pyranose ring conformation. Ketose residues form into hemiketal structures in the same manner except that a second primary hydroxyl will be present at C-1 prior to the ring forming secondary hydroxyl, assuming formation from a 2-ketose acyclic chain (Lindhorst, 2003).

1.2.2 – Final Chiral Atom Stereochemistry and Anomeric Configuration:

As mentioned in section 1.2.1, multiple chiral centres present on monosaccharide residues generate a series of stereoisomers. One of these centres, the final chiral carbon, results in a series of mirror image structures. Eight such enantiomeric pairs exist for hexose chain length aldoses, with the enantiomers termed D- and L- for the rightward and leftward pointing hydroxyls respectively – Figure 4 (Robyt, 1998). Figure 1 and its accompanying text indicates why a 2D format has been used for Figure 4 and how the arrangement around a chiral centre prevents the superimposition of one enantiomer over the other when one has been flipped through 180°. The D- and L- designations used for the final chiral centre orientation are independent of the (d)- and (l)-, now more commonly (+)- and (-)-, designations used to indicate a residue's ability to rotate polarised light in a dextrorotatory (rightward) or laevorotatory (leftward) manner.

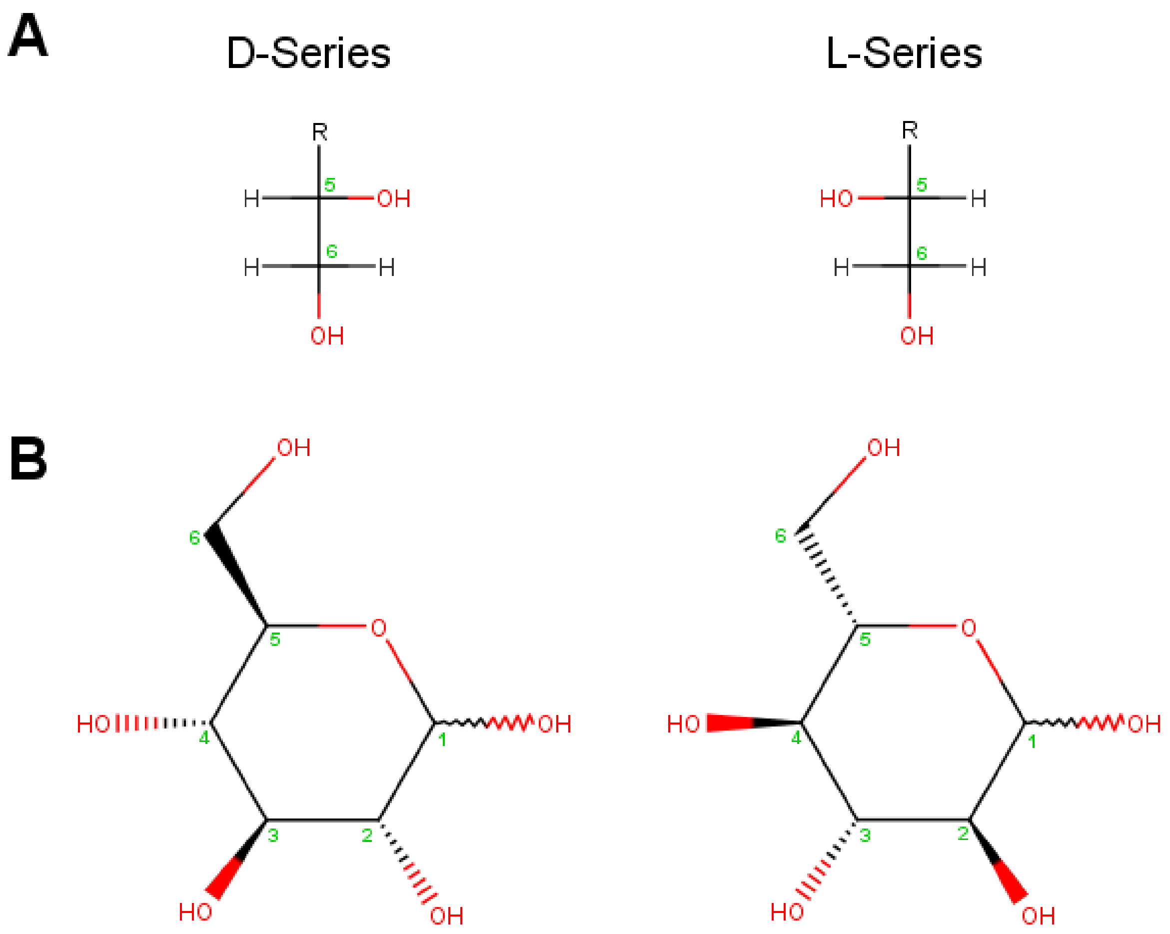


Figure 4: A) Partial Fischer projections showing the D- and L- enantiomers present in both aldose and ketose residues alongside B) exemplar D- and L- monosaccharides.

Hemiacetal and hemiketal ring formation produces a further chiral centre on the anomeric carbon with two further stereoisomeric variants resulting. These new stereoisomers, termed anomers, are designated α- and β- for variants where the hydroxyl group is oriented in the same and alternate direction to that of the final chiral centre in a Fischer representation – Figure 3 (Lindhorst, 2003; Robyt, 1998). Combination of the notation systems for the final chiral atom and anomeric stereocentres therefore give four possible combinations, α-D-, α-L-, β-D-, and β-L-. These four conformations differ in the orientation of the final chiral centre, with D- enantiomer CH_2OH groups present above the plane of the ring, while for L- enantiomers it is below this plane, as shown in Figure 4B. Further to this, α-anomeric conformations will have this hydroxyl group oriented axially on the opposite side of the ring plane than the CH_2OH group, while β-anomeric hydroxyl groups are positioned equatorially. The anomers, as with all non-anomeric stereoisomeric variants, generate distinct effects on the cyclic residue's stability owing to the steric hinderance, electrostatic interactions and torsion strain effects exhibited. For the anomeric chiral centre these are supplemented by two further factors, the anomeric and exo-anomeric effects (Stick and Williams, 2009; Lindhorst, 2003).

The anomeric effect occurs contrary to the general principal that secondary hydroxyl groups oriented equatorially, i.e. in plane, are sterically favourable to those oriented axially, i.e. perpendicular to the plane of the ring, with most anomeric hydroxyls present in the axial form at a higher proportion than would otherwise be expected (Lindhorst, 2003; Stick and Williams, 2009; Sinnott, 2013). This effect has been explained as either resulting from the interaction of lone-pair electrons or dipoles in the sequence C-5–O-5–C-1–O-1 or resulting from the comparative orientation of the O-5 dipoles with respect to the C-1–O-1 bond orbitals (Stick and Williams, 2009; Kuszmann, 2006). The size of the anomeric effect, and consequently the ratio of α- and β-anomers, can be altered by a number of factors including C-2 hydroxyl orientation, which when equatorial generates a weaker anomeric effect (Lindhorst, 2003). Additionally, the polarity of the solvent in which the sample is dissolved alters the strength of the anomeric effect, which is reduced in more polar solvents (Fuchs et al., 1986; Wiberg et al., 2018). The electronegative potential of the C-1 bound group also alters the size of the anomeric effect, with greater electronegativity generating a larger effect, and thus a greater proportion of the α-anomer (Stick and Williams, 2009).

Similarly, the exo-anomeric effect, which dictates the preference of hydroxyl groups to adopt a gauche, 60° bond angle, rather than trans, 180° bond angle, configuration can be explained in two ways. The first is the interaction of dipoles in the sequence O-5–C-1–O-1–R, while the second is as a result of the comparable orientation of the O-1 dipole with respect to the C-1–O-5 bond. Whilst both of these effects have been introduced for C–Y–C–X or Y–C–X–R atom chains, where both X and Y are oxygen, these can be substituted for other atoms such as nitrogen and sulfur at both X and Y or most halogens at X (Stick and Williams, 2009; Kuszmann, 2006; Sinnott, 2013).

1.2.3 – Mutarotation:

The hemiacetal and hemiketal residue structures generated by cyclisation – such as D-glucose in Figure 5 – are not fixed when in solution, instead an equilibrium forms between the acyclic polyhydroxylated aldehyde, or ketone, and four cyclic variants over a time frame between minutes and hours (Silva et al., 2006; de la Concepción et al., 2020). The α and β variants of both pyranose and furanose residues between which this mutarotation occurs can be monitored in a number of ways. Three principal methods are used for this: ^{1}H NMR spectroscopy, chromatography protocols, and optical polarimetry protocols, exhibiting the two anomers' specific optical rotations. During optical polarimetry analyses, these rotations are averaged out in solution and can be used to determine the proportion of each residue conformation present (Lin et al., 2010; Pazourek, 2010). The rate of equilibrium formation, and as such the rate of mutarotation, can be determined using changes to optical rotation values calculated, as can the number of conformations between which mutarotation is occurring (Andersen and Grønlund, 1979; Lin et al., 2010).

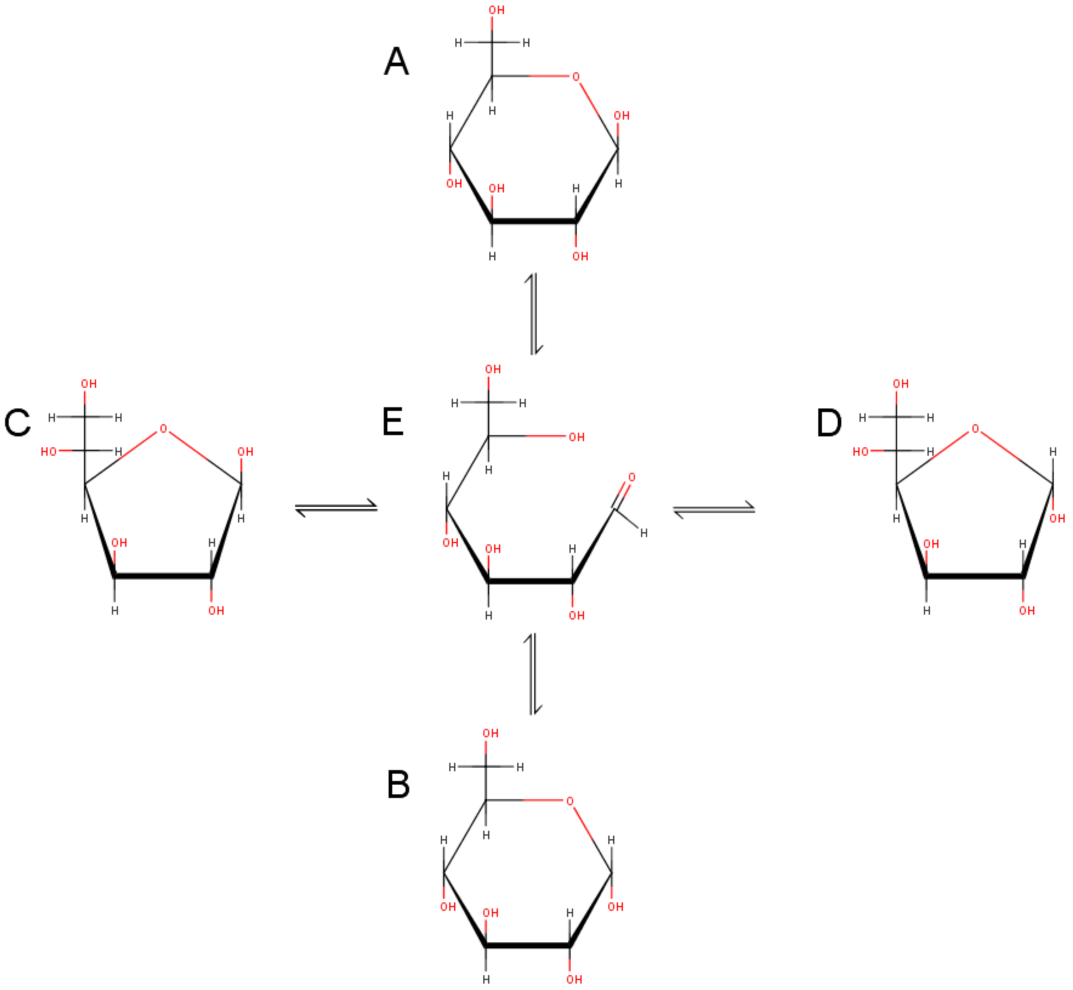


Figure 5: Mutarotation between the acyclic polyhydroxylated aldehyde (E) and cyclic β-D-glucopyranose (A), α-D-glucopyranose (B), β-D-glucofuranose (C), and α-D-glucofuranose (D).

1.2.4 – Ring structure and conformation:

The tetrahedral nature of the carbon atoms comprising cyclic monosaccharide residues ensures that, like cyclohexane, they are non-planar and therefore exist as a number of conformations capable of interconversion (Lindhorst, 2003). Consequently, a series of energetically more or less favourable conformations are possible for both the six-membered pyranose – Figure 6A – and five-membered furanose – Figure 6B – ring structures. Energetic favourability, important in determining the likelihood of a conformation, derives from preference to avoid bond torsion strain, steric hinderance, van der Waals forces, electrostatic stereochemistry, and intramolecular hydrogen bonding which can result in a high energy cost for the maintenance of certain cyclic conformations (Lindhorst, 2003; Robyt, 1998; Kuszmann, 2006). Notable examples of this are the structures of β-D-glucose in $^{4}C_{1}$ and $^{1}C_{4}$ chair conformations with all hydroxyls present as, the more favourable, equatorial and, less favourable, axial bonds, respectively. This results in the $^{1}C_{4}$ conformation having a higher relative energy making it less favourable, and less common, as a result (Alonso et al., 2014). While residues such as β-D-glucose have one predominant conformation others such as α-L-iduronic acid, found in the polysaccharide heparin, have multiple potentially extant variants, with slightly differing energies, which for the iduronic acid are the $^{4}C_{1}$, and $^{1}C_{4}$ chair, and the $^{2}S_{O}$ skew boat conformations (Sattelle et al., 2010; Hsieh et al., 2016).

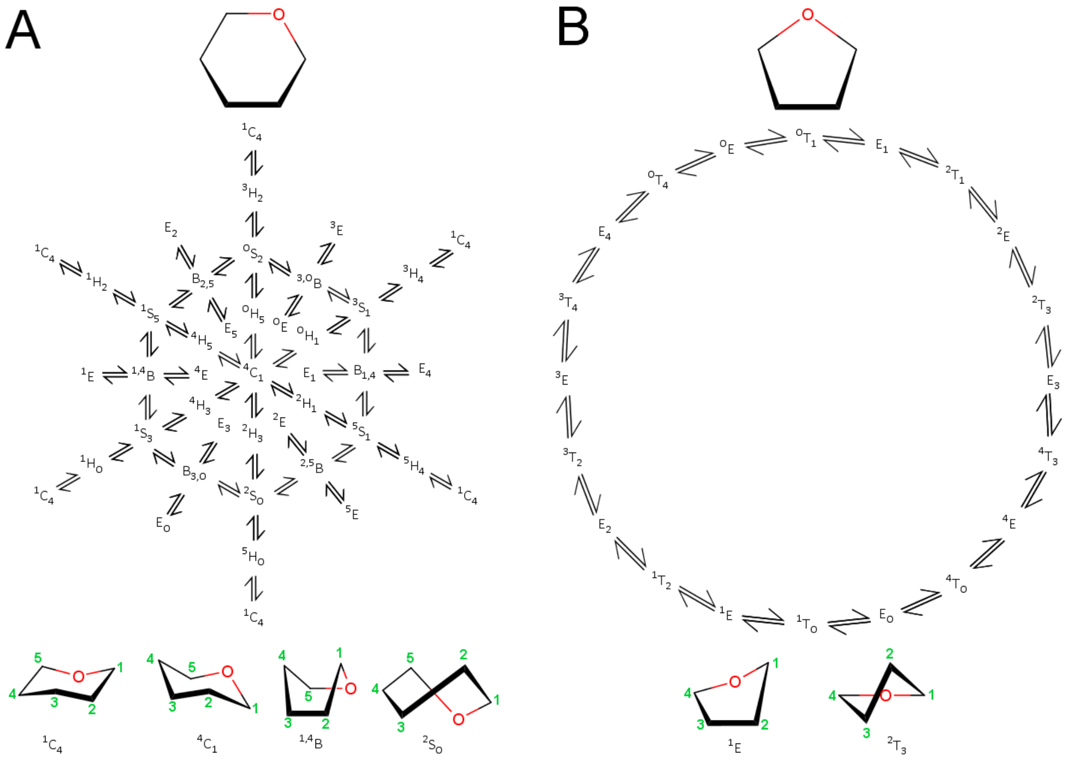


Figure 6: The pyranose (A) and furanose (B) carbohydrate ring structures, the interconversion between different conformations and ring structure diagrams of these different conformations (adapted from Kuszmann, 2006).

Pyranose rings exist as five conformations formed from the carbon and oxygen atoms present within the monosaccharide ring, these are the chair (C), half-chair (H), skew boat (S), boat (B), and envelope (E) variants, examples of some of these variants are shown in Figure 6A (Kuszmann, 2006; Sinnott, 2013). Each conformation has a different number of atoms forming the plane of the ring and a different pattern for those atoms sitting either above or below this plane. The first of these conformations, and most common in pyranoses, the chair, has a four membered plane comprised of three carbon atoms and the ring oxygen with C-1 and C-4 being non-planar. Two chair conformation variants exist, which are $^4C_1$, where C-4 is above the plane and C-1 is below, and $^1C_4$, which is the opposite. Similarly, the half-chair and boat conformations have a four membered plane but in the half-chair the non-planar atoms are adjacent to each other with one sitting above the plane and the other below and in the boat both non-planar atoms are present on the same side of the plane, either above or below, such as C-1 and C-4 in the $^{1,4}B$ structure shown in Figure 6A. The envelope conformation contains a fifth atom within the plane with the remaining atom either above or below it. Contrasting these four conformations, the skew boat conformation's plane is not comprised of wholly adjacent atoms. There are three adjacent and the fourth surrounded by one atom above the plane and the other below it, these are the C-2 and O atoms in the ^2S_O conformation shown in Figure 6A (Kuszmann, 2006).

A complex relationship exists for interconversion between these five conformations, as can be seen in Figure 6A. As expected from the names, half chair conformations are intermediates in the transition between chair and skew boat variants. Similarly, the boat conformations provide a middle ground via which skew boat structures can be converted from one to another, with the close relationship between these two conformations suggested by the close names and structures exhibited. The final conformational variant, the envelope, sits alone connecting only to boat conformations with the clear structural similarity, where one of the boats' non-planar atoms only needs to move in plane to create the envelope structure (Kuszmann, 2006).

By contrast, furanose ring structures have fewer potential conformations either being envelope (E) or twist (T) variants – Figure 6B (Kuszmann, 2006; Sinnott, 2013). Furanose envelope conformations have a plane of four ring atoms and the final carbon, or oxygen, located either above or below this plane. Alternatively, the twist conformations have three atoms present as the plane and the remaining two adjacent atoms sit one above and the other below it (Kuszmann, 2006). The ^1E and $^2T_3$ exemplar conformations, shown in Figure 6B, have the ^1E C-1 and $^2T_3$ C-2 above their respective planes whilst C-3 of $^2T_3$ is below the plane for that residue. The furanose ring conformations form as a connected loop – Figure 6B – where the envelope and twist variants alternate in a manner where one conformation is the intermediate for conversion between the two surrounding variants of the other (Kuszmann, 2006).

1.2.5 – Functional Group Variants:

The array of monosaccharide residues generated by aldose or ketose chain stereochemical, and conformational variation is supplemented by natural derivatisation of the hydroxyl and/or carbonyl groups present. The derivatives formed are broadly divided into sugar alcohols, sugar acids, deoxy-sugars, and amino-sugars due to the type of functional group added to, or lost from, the acyclic chain. The example in Figure 7 shows the alcohol, three acids, deoxy-sugar and amino-sugar which can be derived from a glucose residue.

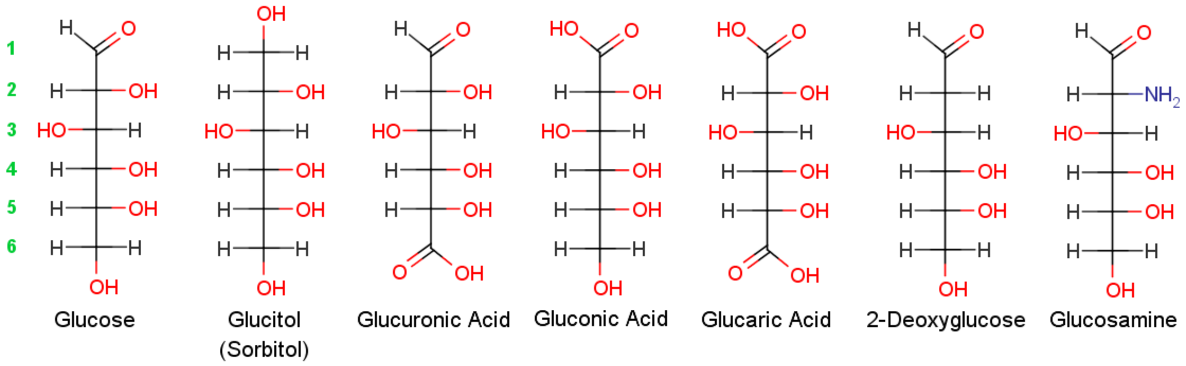


Figure 7: Fisher projections for the D-glucose monosaccharide residue's natural variants.

Derivatisation of sugars to sugar alcohols occurs via the reduction of the aldehyde group carbonyl to a hydroxyl preventing cyclisation of these residues (Grembecka, 2015). Conversely sugar acids form via oxidation, which can occur on either the primary hydroxyl to give an 'uronic acid', the aldehyde carbonyl to give an 'onic acid', or to both the primary hydroxyl and aldehyde carbonyl to give an 'aric acid' (Mehtiö et al., 2016). The final two derivatives, deoxy- and amino-sugars, form when a hydroxyl group is substituted for either a hydrogen atom or an amine group which, whilst possible for all hydroxyls, remains uncommon in nature (Marino and Bordoni, 2022; Skarbek and Milewska, 2016). The examples shown in Figure 7 are substituted at carbon-2 by the addition of a hydrogen atom and amine group, respectively.

1.2.6 – Relative Hydroxyl Group Reactivity:

Whilst different monosaccharides can exhibit different patterns of relative hydroxyl group reactivity and certain modifying groups target specific hydroxyls more readily, a typical pattern exists for carbohydrate residues. The reduced steric hinderance of primary hydroxyl groups results in their preferential modification when free secondary or tertiary hydroxyls are present. Steric hinderance also predisposes equatorial secondary hydroxyls to modification in the presence of free axial alternatives. However, equatorial hydroxyls surrounded by two other equatorial hydroxyl groups are less reactive than when there is one adjacent axial hydroxyl group. Thus, the order of reactivity for β-D-Glucose, with all equatorials, would be C-6 > 1 > 4 > 2 & 3 whereas for β-D-galactose with an axial C-4 secondary hydroxyl it would be C-6 > 1 > 2 & 3 > 4 (Dimakos and Taylor, 2018).

1.2.7 – Disaccharide and Polysaccharide formation:

Combinations of monosaccharide residues generated via glycosidic linkages produce multiple residue carbohydrates, the simplest of which are the disaccharides, such as maltose and lactose. These glycosidic linkages occur between the hemiacetal or hemiketal anomeric hydroxyls from one residue with a non-anomeric hydroxyl from a second. The result of this is that one residue, with an un-bound anomeric hydroxyl, forms a reducing end with mutarotation exhibited, while the second residue, with a bound anomeric hydroxyl, forms a non-reducing end. These are still reducing sugars. A wide range of disaccharides are possible owing to the potential monomer combinations and the linkages between these residues (Gray et al., 2017). Conversely, glycosidic linkages can form between two anomeric hydroxyls generating a non-reducing sugar such as in sucrose (Huang et al., 2021).

The further addition of residues results in oligo- and polysaccharide formation as either linear or branched chains. Linear chains comprise residues linked to up to two further monosaccharides such as the carrageenan family of red algal polysaccharides or the mammalian glycosaminoglycans (van de Velde et al., 2004; Yates et al., 1996). By contrast, branched polysaccharides' residues have linkages to a greater number of further monosaccharides or longer chains, for example three residues are bound to certain backbone residues of amylopectin and tamarind seed xyloglucans (Bertoft, 2013; Tuomivaara et al., 2015).

The geometry of linkages between residues is a key factor in their structural characterisation as well as their consequent physical-chemical properties and protein interacting ability owing to the effect it has on the polysaccharides' higher order structure. These geometries differ for linkages between two secondary hydroxyls, governed by the dihedral angles phi (ϕ) and psi (ψ), and those including a primary hydroxyl where a third angle, omega (ω), must also be considered. Phi and psi are used to define the angles between positions O_5 - C_1 - O – C_x' or C_1 - O - C_x' - C_{x-1}', whilst phi, psi, and omega are used for those between positions O - C_6' - C_5' – C_4' respectively (Wormald et al., 2002).

1.2.8 – Natural Modifications of Carbohydrates:

Whilst monosaccharide residue functional group variants have been introduced, further functional groups can naturally be added to all types of carbohydrate ranging from individual monosaccharide residues to polysaccharides. This occurs by substitution of hydroxyl or amine groups, amongst others, and is important in developing the biological, physicochemical and protein interaction properties of these molecules. Many such substitutions have been identified as naturally occurring including the addition of both charged and non-charged functional groups, amongst others (Yu and Chen, 2007).

Many biologically important roles originate from the addition of charged functional groups to carbohydrates. Sulfation is a prevalent example of this, with modification of the glycosaminoglycan family member heparin having been extensively studied given its wide ranging application as a biopharmaceutical. In this case, the addition or removal of sulfate groups drastically alters interactions with proteins such as antithrombin, fibroblast growth factors and cyclophilin B amongst others, as well as the biological activities resulting from these interactions (Hricovíni et al., 2001; Xu et al., 2012; Vanpouille et al., 2007). Other naturally sulfated polysaccharides, including from the carrageenan and glycosaminoglycan families, have been studied in terms of the effect that modification has on physicochemical properties and biological activity (de Araújo et al., 2013; Soares da Costa et al., 2017). These investigations have also highlighted the importance of substitution at specific sites on a polysaccharide chain. Charged functional group addition also has a major role for monosaccharides with important examples including glucose or ribose and deoxyribose phosphorylation during glycolysis or ribonucleic and deoxyribonucleic acid formation, respectively (Chandel, 2021; Banfalvi, 2021).

Similarly, the addition of non-charged functional groups can have a major impact on the physicochemical and protein interaction properties, as well as the biological activities of the modified carbohydrate. A principal example is the acetylation of amino groups, studied extensively for glycosaminoglycan family members, where heparin de-N-sulfation and N-acetylation has been shown to drastically reduce the level of anticoagulant activity, however, when examining the effect of N-acetylation on inhibition of complement system activation no difference was seen between the N-sulfated and N-acetylated heparin samples (Wessel et al., 1989; Weiler et al., 1992). Acetylation is also common in sialic acids, nine-carbon chain residues which commonly act as the final carbohydrate in glycolipids, both sulfated and unsulfated N- and O-linked glycans as well as in bacterial carbohydrates and during the evasion of the immune response (Yu and Chen, 2007; Varki, 2007; Park, 2019). One of these acids, N-Acetyl-Neuraminic acid, has been shown to be further acetylated and methylated, amongst others (Schauer, 2000). Such O-acetylation is also seen in plant based polysaccharides such as during xyloglucan biosynthesis in certain plant species (Jia et al., 2005; Schultink et al., 2014).

The structure of carbohydrates, including natural modifications of mono-, di-, oligo-, and polysaccharides plays a leading role in determining their resulting biological activities, protein interactions and physicochemical properties. Therefore an in-depth knowledge of the key principles surrounding carbohydrate structure, their behaviour in solution, natural variants, naturally occurring modifications, and predicted reaction chemistry is a prerequisite for the development of novel samples and novel reaction mechanisms. The information introduced regarding monosaccharide chemistry and the formation of oligo- and polysaccharides is therefore important in terms of its application to the understanding of higher order carbohydrate structures, to be introduced in section 1.3, and the modification methods which can be applied to target carbohydrates, introduced in section 1.4.

1.3 – Polysaccharides:

Carbohydrates as one of the four principal classes of biological molecule, alongside proteins, lipids and nucleic acids, are widespread throughout all forms of life. Their diversity has resulted in their use in myriad applications, both natural and synthetic, owing to their structures and consequent physico-chemical and/or protein interaction characteristics. Polysaccharides are the most abundant form, corresponding to 90% of the total mass of carbohydrates, with activities either alone or when coupled to the other principal classes of biological molecule such as proteins, to form glycoproteins (BeMiller, 2019).

1.3.1 – Biosynthesis:

Carbohydrate biosynthesis employs a series of specific enzymes acting to generate a polymeric chain of glycosidically linked residues. Within eukaryotic cells, of both mammalian and plant-based varieties, the Golgi apparatus is the site of this biosynthesis whereas corresponding biosynthesis in prokaryotic cells takes place at the inner membrane (Esko and Lindahl, 2001; Zabotina et al., 2021; Schmid and Sieber, 2015; Whitfield and Paiment, 2003). The diversity of extant carbohydrate polysaccharides necessitates an equally diverse array of biosynthetic enzymes required both for the addition of residues, such as glucan synthase, galactosyltransferase, and fucosyltransferase, which add glucose, galactose, and fucose respectively during xyloglucan synthesis, as well as for the subsequent modification of the added residues (Schultink et al., 2014; Pauly and Keegstra, 2016; Zabotina, 2012). The variety of potential modifications is as diverse as that during residue addition, leading to greater variability of the resulting polysaccharide. A number of examples of this modification include acetylation or deacetylation as well as a sulfation and epimerisation, each of which occur during synthesis, for example of either xyloglucan or heparin (Schultink et al., 2014; Esko and Lindahl, 2001).

Polysaccharide biosynthetic enzymes commonly function in a specific order such as during the formation of the mammalian proteoglycan heparin, in the Golgi apparatus, in two distinct phases of synthesis. This first, chain elongation, is initiated by xylosyltransferase addition of the nucleotide sugar UDP-xylose to a serine residue on the core protein. The four residue linkage region, β-D-GlcA-1,3-β-D-Gal-1,3-β-D-Gal-1,4-β-D-Xyl-1-Ser, is then completed upon the addition of two galactose and a glucuronic acid residue by their respective glycosyltransferase enzymes. Subsequent addition of a N-acetyl-glucosamine by N-acetyl-glucosamine transferase differentiates the glycosaminoglycan being formed as either heparin or heparan sulfate, as opposed to chondroitin or dermatan sulfates which would contain a N-acetyl-D-galactosamine residue in this position. Chain elongation then occurs upon the alternating addition of D-glucuronic acid and N-acetyl-D-glucosamine residues by D-glucuronic acid and N-acetyl-D-glucosamine transferase enzymes respectively. Following chain elongation the second stage of heparin biosynthesis is the modification of individual residues within both the linker region and the polysaccharide itself. Within the polysaccharide, simultaneous N-acetyl-D-glucosamine deacetylase and N-sulfotransferase activity converts the glucosamine residues into their sulfated variant before glucuronic acid position-C5 epimerase converts the β-D-glucuronic acid residues to α-L-iduronic acids. Subsequent 2-O-sulfotransferase, then 6-O-sulfotransferase and finally 3-O-sulfotransferase enzymes generate the characteristic heparin sulfation pattern (Esko and Lindahl, 2001; Rudd and Yates, 2012; Chappell and Liu, 2013).

1.3.2 – Structure:

As has been introduced, the structure of polysaccharides can be highly heterogeneous owing to the different monosaccharides from which they are comprised, the ring conformations of these residues, their multiple hydroxyl groups capable of forming glycosidic linkages, and the myriad derivatisations that can occur to each of these hydroxyl groups. However, in many cases, polysaccharides exhibit a repeating unit structure with residues, linkages, and derivatisation patterns similar, if not identical, across the repeating units within the polysaccharide. This is the case in the linear glycosaminoglycan and carrageenan family members as well as for the branched xanthan gum (Pomin, 2014; van de Velde et al., 2004; Brunchi et al., 2019). Where multiple repeating units are possible, as is the case for hemicellulosic xyloglucan, typically a reduced number will be present when the polysaccharide is sourced from the same species. For example, tamarind seed xyloglucan has only four repeating unit variants out of the myriad total possibilities. Therefore, whilst the number of variants present for samples from the same species are reduced, some remains, as does the variability of the constituent repeating units present between samples from different species (Tuomivaara et al., 2015; Hsieh and Harris, 2009).

As there is such an array of possibilities for the structure of polysaccharides, which could not be comprehensively introduced here, the key structural characteristics of the carbohydrates used in this research will form the basis for the remainder of this introduction section. Each of the five polysaccharides used in this project, heparin, κ-carrageenan, ι-carrageenan, cellulose and xyloglucan, have a repeating unit structure, albeit simple for cellulose, comprising repeating instances of 1,4-linked β-D-glucopyranose residues with no substitution of the hydroxyl groups present either as further residues or anionic functional groups (Park et al., 2010). Whilst more complex than the repeating unit of cellulose, that of heparin remains straightforward with a most common underlying repeating disaccharide structure composed of 1,4-linked α-L-iduronic acid and 1,4-linked α-D-glucosamine residues, both in a pyranose ring conformation. Both the amide and the position six hydroxyl groups of the glucosamine residue are sulfated along with the position two hydroxyl group of the iduronic acid residue, with its carboxylate group adding a further charge to the repeating unit (Yates et al., 1996). Whilst both of the carrageenan polysaccharides are composed of a repeating disaccharide of 1,3-linked β-D-galactopyranose and 1,4-linked α-D-galactopyranose residues with one, β-D-galactopyranose position four, and two, β-D-galactopyranose position four and α-D-galactopyranose position two, hydroxyl groups sulfated for the κ- and ι- variants respectively, a further structural feature of interest is derived from the 3,6-anhydrous ring which can form under certain conditions (van de Velde et al., 2004). Unlike the four carbohydrate samples already introduced, xyloglucan does not have a linear repeating mono- or disaccharide structure, instead this polysaccharide has four repeating unit variants composed of seven, eight or nine residues for the *T.indica* species used in this research. The only substitution of hydroxyl groups present in these repeating units are the presence of further monosaccharide residues, where each of the four repeating units has the same four 1,4-linked β-D-glucopyranose residues forming a cellulosic backbone chain. α-D-xylopyranose residues branch from the first three of these glucose residues with a 1,6-linkage connecting the two residue types. Finally, the presence of a β-D-galactopyranose residue 1,2-linked to either the second, third, or both second and third xylose residues completes the repeating unit structures of *T.indica* xyloglucan (Tuomivaara et al., 2015). These repeating unit structures can be seen in Figure 8, below.

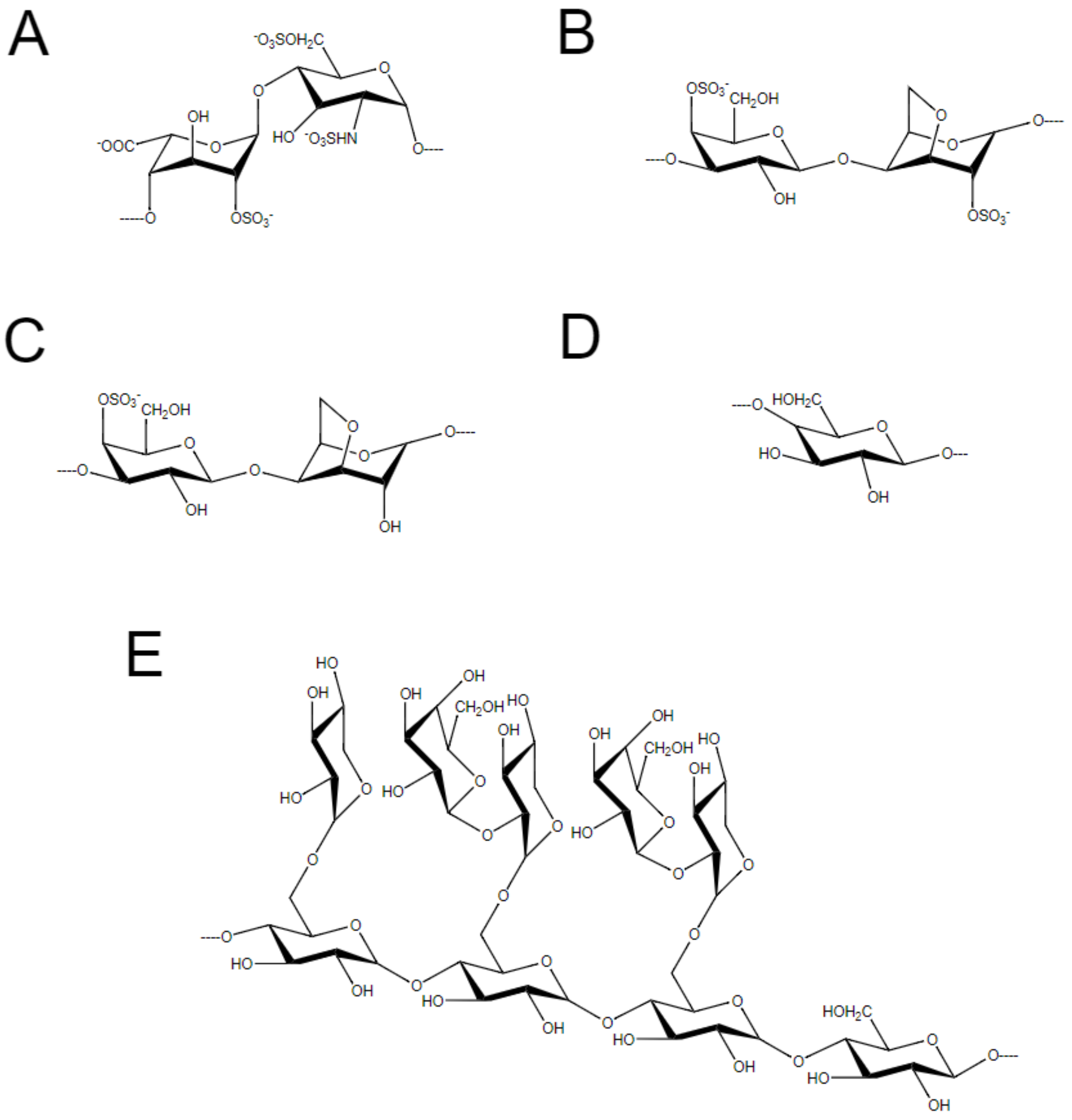


Figure 8: The repeating unit structures of heparin (A), ι-carrageenan (B), κ-carrageenan (C), cellulose (D), and one example from *T.indica* xyloglucan (E).

As well as the lower order structure of these carbohydrates determined by the monosaccharide residues, and the linkages between these residues, from which they are composed alongside any hydroxyl or amide group substitutions present, higher order structures have also been observed. The higher order structure of cellulose is formed from parallel polysaccharide chains with hydrogen bonds forming a network of interactions both within individual chains and between the chains resulting in the formation of a sheet structure (Nishiyama et al., 2002; Nishiyama et al., 2003). By contrast to the linear chains of cellulose polysaccharides which form a sheet structure, the higher order structure of the carrageenan family members forms into coils in a double helix. These coiled chains have been shown to aggregate into bundles of helices (Yuguchi et al., 2002).

The structures of carbohydrates, both the lower order residue composition, linkages, and hydroxyl group substitution and higher order parallel sheet and double helix conformations, have an effect on their function. A clear example of a structure-function relationship are the carrageenans, where both the lower and higher order structures affect their ability to form gels. Both κ- and ι-carrageenan samples have been identified as being gel-forming due to the presence of 3,6-anhydrous rings on the α-D-galactopyranose residues. This causes the residues to be present in a $^1C_4$ conformation and consequently the polymer chains form into the helical conformations necessary for their gelation. λ-carrageenan, which does not contain a 3,6-anhydrous ring, on the other hand, cannot undergo any of the processes described and as such gel formation does not occur (Yuguchi et al., 2002; Campo et al., 2009).

1.3.3 – Interactions:

The structures of carbohydrates, introduced above, determine their ability to interact with other biological molecules as a result of their residue composition, the linkages which connect these residues, and the natural modifications which can occur. These interactions form with all classes of biological molecules including other carbohydrates, lipids, and proteins in both advantageous and detrimental capacities (de la Fuente and Penadés, 2004; Liu et al., 2012; Zeng et al., 2012). Harnessing these interactions has been a focus of both pharmaceutical and commercial carbohydrate applications, notably in the case of heparin, which has a marked impact on coagulation when interacting with proteins of the coagulation factor cascade such as antithrombin and thrombin (Hricovíni et al., 2001; Jin et al, 1997).

Of the interactions formed by carbohydrates, those with proteins are particularly common and will be covered in greater detail below. These interactions have roles within a number of biological activities including during the formation of tyrosine kinase receptor complexes, leading to the stimulation of multiple downstream signalling pathways, as well as upon interaction with members of the lectin family, leading to signalling pathway activation and responses including limiting tumour cell proliferation and T-cell activity during auto immune diseases, and when bound to cyclophilin B, with the potential to affect the roles in inflammation and viral infection performed by this protein (Ornitz and Itoh, 2015; Roy et al., 2016; Hanoulle et al., 2007; Vanpouille et al., 2007).

Having observed natural interactions between carbohydrates and proteins, the modification, or identification, of novel carbohydrates with the potential to interact with commercially or medically significant binding partners has been a focus of research. This research has aimed to determine whether these samples can derive the same interactions, and downstream biological activities, as that of the natural interacting partner with examples including one project where the systematic sulfation of carrageenan family members was used to determine whether the anticoagulant effects of heparin could be mimicked (de Araújo et al., 2013). Alternatively, a chondroitin sulfate sample identified as having a unique sulfation pattern was subjected to a number of assays in order to characterise its ability as both an anticoagulant and an anti-inflammatory agent (Palhares et al., 2019).

Beyond the interactions of carbohydrates with proteins, many interactions have been reported between two carbohydrates. One of these examples, derived from the polysaccharides which form plant cell walls, is the interaction between the hemicellulosic polymer xyloglucan and polysaccharide cellulose in a manner which, whilst strong, was shown to the be the result of non-covalent interactions (Park and Cosgrove, 2015). Further studies have suggested that the carbohydrate-carbohydrate interactions formed during cellular recognition involve higher order structures, with multimer formation possible, to produce the affinity required for an interaction, although once formed these interactions have been shown to be as strong as those of two interacting proteins (Bucior and Burger, 2004). The interaction of Lewis^X carbohydrate antigens with one another is a further example of carbohydrate-carbohydrate interactions that has been shown to allow for the compaction of embryonic cells in a manner dependent on the presence of Ca^{2+} ions. This cellular compaction during embryonic development is critical for the process to continue (Bucior and Burger, 2004).

Carbohydrate interactions with non-carbohydrate small molecules have also been observed, with the formation of hydrogen bonds between the carbohydrates and lipid polar head groups indicated as the cause of interactions with these compounds, although hydrophobicity based mechanisms cannot be ruled out (Liu et al., 2016). Such interactions between carbohydrates and lipids can lead to increased uptake of the lipid cholesterol amongst other effects (Lairon et al., 2007). Non-carbohydrate small molecules other than the lipids introduced above have also been observed to interact with carbohydrate samples such as polyphenols with the resulting interactions playing significant roles in the human body. Specifically, interactions with cellulose, the cell wall like pectin, and dietary fibres have been investigated at length with increasing degrees of polyphenol polymerisation seen to correlate with increased interaction. As with the lipid samples, it has been postulated that these interactions have been caused by the presence of hydrophobic interactions and hydrogen bonds, particularly with the oxygen atoms of the glycosidic linkage, although covalent bonds can form between these two classes of compound (Jakobek, 2015).

The properties of polysaccharides, formed by one of many biosynthesis routes, are each a major factor in the natural roles and man-made applications of carbohydrates. Continued biosynthesis of carbohydrates with identical structure from chain to chain has enabled their use across multiple fields of research with the security that conclusions made can be matched to a specific structure or structural feature, such as that of xanthan gum and its physicochemical properties for use in commercial applications. Consequently, it is possible to match the desirable properties of a carbohydrate sample to a specific structure which, once identified in natural samples, can be isolated in batches of increased scale for use in industrially relevant applications or further researched to optimise the properties. With potential carbohydrate applications reliant on properties ranging from those of a physical-chemical nature to interactions with other classes of biological molecule, the number of desired structures would be immense. Therefore, modifications are required to generate not naturally occurring carbohydrate structures, and many methods are available which can be used to tailor them, including those introduced below.

1.4 – Carbohydrate Preparation and Modification:

The structure of carbohydrates plays a major role in determining the sample's physical properties, their interactions with other molecules, and their biological activities. Modifications made to these structures can therefore diversify their properties, further increasing the applications to which carbohydrates can be applied. The modifications performed can be split into a small number of groups depending on what functional groups, if any, are being added, which atom the modification is being performed on, and the manner in which the modification is being performed.

1.4.1 – Sites of Modification:

A number of functional groups are present within the carbohydrates' structures which can readily be subjected to derivatisation. One of the most commonly modified functional groups are the hydroxyls present at multiple sites across a carbohydrate's constituent monosaccharides (Dimakos and Taylor, 2018; Cumpstey, 2013). Other than these, carbohydrates have been modified on both amide and carboxyl functional groups, where these are present, such that the heparin glucosamine residue amine groups have been modified via both sulfation and acetylation for example. Similarly, iduronic acid carboxyl groups have been derivatised by both esterification methods and the addition of further amide functional groups (Fernández et al., 2006; reviewed in Palhares et al., 2021). As well as these, the carbon atoms present within carbohydrate residues can be modified, with the addition of functional groups including sulfonates and halides (Cumpstey, 2013).

1.4.2 – Enzymatic vs Chemical Modifications:

Carbohydrate modification is carried out via two principal routes, one utilising chemical compounds, typically in organic solvents, and the other comprising enzymatic modification, typically, performed in an aqueous buffer. In many cases, modifications can be performed via both methods, as is the case with heparin, where a series of enzymes from the biosynthesis process have been used for modification (Carlsson et al., 2008; Byrne et al., 2018). Equally, an array of chemical sulfation, desulfation, and acetylation protocols have been employed that are targeted towards the generation of heparin repeating units containing specific sulfation patterns (Yates et al., 1996; Yates et al., 2000). Whilst enzymes have been applied to a vast array of carbohydrate modification methods they did not form a focus of the research carried out for this thesis and, therefore, will not be introduced in detail here. Chemical carbohydrate modification, however, has been employed during this research and is described below.

1.4.3 – Protection and Chemical Modification Methods:

Whilst natural carbohydrates contain a vast array of potential residue combinations and linkage positions, amongst a number of other factors, the diversity of these samples can be further increased by the application of modification reactions. The ability to perform modifications by both enzymatic and chemical means with protecting group addition and hydrophobic salt preparation steps for chemical reactions increases the likelihood of being able to produce “designer” carbohydrate samples or a least to modify their properties for a range of applications. However whilst the implementation of different protecting group strategies can improve the selectivity with which hydroxyl, or other, functional groups are modified, the additional steps required can increase the proportion of the starting sample that is lost during synthesis.

A feature of chemical modifications, less commonly encountered with those performed using enzymes, is a reliance on preparatory methods such as functional group protection strategies and hydrophobic salt formation to aid solvent compatibility (Volbeda et al., 2019; Liu et al., 1994). Protecting group strategies are required when the target functional group for modification would not be prioritised according to the residue’s relative order of reactivities.

Protecting group strategies can be used to leave solely the target functional group free to react with the desired modifying agent. As primary hydroxyls are the most likely to be substituted during chemical modifications they are one of the principal targets for protecting groups which then allows reaction of the other, secondary hydroxyl groups (Volbeda et al., 2019). The type of protecting group selected however can affect the reactivity of the carbohydrate, an example of which is the acyl protecting groups, which limit the ability of a glycosyl donor to interact with the protected carbohydrate consequently preventing any further chain length increase (Volbeda et al., 2019).

Alongside protecting group strategies, hydrophobic salt formation is a means of preparing samples prior to modification to increase the solubility of the carbohydrate in organic solvents. Two such examples of this preparation come from the reactions required to modify the heparin sulfation pattern through conversion using a H^+ form cation exchange resin and neutralisation to either its tetrabutylammonium or pyridinium salt (Yates et al., 1996). Using such a method to prepare samples for modification can generate problems however, particularly when the cation salt is being generated across an exchange resin (Liu et al., 1994). For example, in cases where hydrophobic salt production is being attempted for a gel forming carbohydrate it can lead to the exchange resin, contained within a column, becoming blocked.

Having produced adequately protected samples in a hydrophobic salt form, there are a wider range of modifications that can then be performed. These include the addition of charged functional groups such as sulfates and phosphates in a manner which has the potential to alter the carbohydrate's ability to interact with a number of proteins (Xu et al., 2016; Dadhich et al., 2015). Alternatively, however, when varying the number of charged substituent groups is not required, modifications based on the addition of acyl or alkyl chains can be performed to potentially increase hydrophobicity (Fox et al., 2011). Alongside these modifications, others such as amination reactions can be carried out which alter the functional groups present on the carbohydrate structure but retain a means by which further modifications can be performed (Simi and Abraham, 2010; Zhang et al., 2017). The final type of chemical modification to be introduced here are glycosylation reactions which can be used to create "designer" carbohydrates either for protein interaction, rheological property generation, or biological activity. However, to create a truly designer carbohydrate, these reactions would need to be combined with a protecting group strategy and considerable further research into protocol optimisation would be required (Galonić and Gin, 2007).

In addition to modifications which have a direct effect on the structure of carbohydrate chains, there are those, such as cation variation, which can have a major effect on the carbohydrate chain's conformation without the addition of covalently linked functional groups. One such example is the exchange of sodium cations for those of calcium, leading to a stiffening of the heparin chain structure with a consequent alteration of the its ability to bind certain proteins (Hughes et al, 2017). Such conformation changes, resulting from the coordinating cation present, could therefore have the potential to alter the relative reactivities of the carbohydrate's functional groups. This is exemplified by the alkylation of myo-inositol when sodium or lithium ions are present, which had an observable effect on the regioselectivity of alkylation reactions performed (Devaraj et al., 2009). Prior modifications would likely be required to increase the number of anionic functional groups present on the carbohydrate, leading to potentially increased solubility and decreased viscosity as well as ionic interactions with the cations present. However, the same issues regarding passing gel forming carbohydrates across a column based exchange medium would remain.

Modifying natural carbohydrate samples can also be used as a means of mimicking biological properties. One key example of this is the chemical sulfation of κ-carrageenan, converting a carbohydrate which cannot naturally bind antithrombin into a range of sulfation pattern variants that could subsequently interact with this protein and exhibit various effects on coagulation time (de Araújo et al., 2013).

1.5 – Carbohydrate-Protein Interactions:

Carbohydrates form complexes with a number of other biological compounds but those with proteins are of particular importance. The effects that these carbohydrate-protein interactions have can be profound with their impacts felt across a range of biological and medical processes. Therefore, a selection of the interactions between a number of exemplar carbohydrates and their protein interaction partners will be introduced below. Whilst this introduction section focuses on the interactions between carbohydrates and proteins during biological processes, interactions between enzymes and the carbohydrates they modify also occur, however, as enzymatic carbohydrate modifications were not performed in the research contained in this thesis they will not be discussed further.

1.5.1 – Roles of Carbohydrate-Protein Interactions in Health and Disease:

The prime example of a carbohydrate in a medical capacity that utilises its protein interactions is that of heparin interacting with proteins from the coagulation cascade, amongst which, the interactions with antithrombin have been researched extensively (Hricovíni et al., 2001; Jin et al, 1997; Choay et al., 1983; Toida et al., 1996; Chen et al., 2017). However, whilst this interaction has proved particularly important for the development of pharmaceutical heparin and its derivatives, it is not the only interaction to involve heparin as a binding partner, a vast array of which have been determined (Ori et al., 2008; Peysselon and Ricard-Blum, 2014; Capila and Linhardt, 2002). Amongst these there are a number of interactions which have provided a particular research focus including those with FGF family members, amyloid-β, α-synuclein, and the amyloid precursor protein, complement factors, and cyclophilin B (Ornitz and Itoh, 2015; Madine et al., 2009;

Scholefield et al., 2003; Blaum et al., 2010; Vanpouille et al., 2007). Each of these interactions has been shown to have an effect which could prove significant if it were to alter the natural functions of the interacting proteins. Given the natural functions of these interacting proteins include roles in muscle development and wound healing, amongst others, for FGF family members and roles in the development of neurodegenerative disorders for amyloid-β, α-synuclein, and the amyloid precursor protein, the potential detrimental effect on health at all stages of life could be profound (Ornitz and Itoh, 2015; March and Blurton-Jones, 2012; Irwin and Hurtig, 2018; Scholefield et al., 2003). Equally important, owing to their potential effects on health, are the natural functions of cyclophilin B and complement factor H, which are important in T-cell adhesion and the complement system of innate immunity (Hanoulle et al., 2007; Ferreira et al., 2010).

In addition to the interactions involving heparin, the carrageenan family members have been shown to interact with proteins such as those of the coagulation cascade, β-casein and κ-casein present in milk, and β-lactoglobulin, and α-lactalbumin present in whey (de Araújo et al., 2013; Silva et al., 2010; Ozawa et al., 1984; Drohan et al., 1997; Stone and Nickerson, 2012; Liu et al., 2020). Whilst each of these interactions are not natural, they represent the effects of industrial applications. The interactions recorded with growth factors, and FGFs in particular, relate most closely to the research carried out in this project. These interactions, when previously reported, have been shown to result in the inhibition of growth factors, where ι-carrageenan proved to be the most effective FGF-2 agonist, in a manner which could be applied as an inhibitor of angiogenesis and tumour metastasis (Hoffman, 1993; Niu et al., 2015). As mentioned above the interactions between carrageenan family members are affected by modification reactions, such as the manipulation of carrageenan sulfation pattern which dramatically increased the κ-carrageenan starting material's ability to act as an anticoagulant (de Araújo et al., 2013).

Equally, xyloglucan-protein interactions have been identified including those of the carbohydrate with epidermal growth factor and its receptor, as well as integrin β4 which serve to inhibit the natural function of these complexes and which triggers skin cells to begin differentiation (Zacharski et al.,2015). Interactions between xyloglucan and FGF-18 have also been shown to include a significant hydrophobic contribution, enabling its use as a scaffold during soft tissue regeneration (Dispenza et al., 2017).

1.5.2 – Types of Carbohydrate-Protein Interaction:

Carbohydrate-protein interactions occur via a range of mechanisms, including those requiring electrostatic interactions, hydrophobic interactions, CH/π-interactions, and interactions formed as a result of cation bridging (Zeng et al., 2012; del Carmen Fernández-Alonso et al., 2012).

Electrostatic interactions form primarily between negatively charged carbohydrates and proteins' positively charged amino acid side chains. Carbohydrate samples which are highly charged, such as heparin with multiple sulfate groups as well as carboxyl groups, are therefore prone to the formation of these types of interactions as is the case for heparin with antithrombin, FGF-2 and complement factor H (Blaum et al., 2010; Seyrek and Dubin, 2010; Seyrek et al., 2007; Chiu et al., 2014). Further to this interaction mechanism, hydrogen bonds can form between the carbohydrate's hydroxyl group hydrogen atoms and the nitrogen and/or oxygen atoms present within the sidechains of certain amino acids including asparagine, glutamine and histidine, amongst others. The reciprocal arrangement is also true with the hydrogen atoms from the amino acids' side chain amide groups interacting with the hydroxyl group oxygen atoms (Weis and Drickamer, 1996; Gabius et al., 2011). Similar to these interactions are those which form between the carbohydrate and amino acid functional groups mentioned and cations, including Calcium (II) and Manganese (II), to form cation bridging interactions (del Carmen Fernández-Alonso et al., 2012; Weis and Drickamer, 1996; Gabius et al., 2011). Finally, there are CH/π-interactions which form when a cluster of three or more CH groups, such as those oriented axially to the plane of the ring in D-glucopyranose residues, form a hydrophobic patch which can interact with one formed by the aromatic regions of amino acids. These interactions can also form between carbohydrate and aromatic small molecules (del Carmen Fernández-Alonso et al., 2012; Spiwok, 2017; Hudson et al., 2015).

1.5.3 – The Effect of Cations of Carbohydrate-Protein Interactions:

As previously mentioned cations can affect carbohydrate-protein interactions, such as those of lectins with their carbohydrate binding partners, by enabling the formation of cation bridges as an additional means of binding (Gabius et al., 2011). However, other carbohydrate-protein interactions are affected by the presence of a cation, such as those of heparin with endostatin and heparin cofactor II which were both enhanced by the addition of divalent cations and specifically calcium ions, respectively (Ricard-Blum et al., 2004; Zhang et al., 2004). The presence of coordinating calcium ions interacting with heparin polysaccharide chains have been shown to increase the rigidity of the carbohydrate (Hughes et al, 2017). However, when the activity of FGF-2:2-O-Desulfated Heparin:FGFR complexes has been determined for both the sodium and calcium heparin salts, little difference was observed in signalling activity (Rudd et al., 2007). Despite this, cation variation can have a major effect on the ability of both FGF-2:2-O-Desulfated Heparin: FGFR and FGF-2:N-Acetylated Heparin:FGFR complexes to generate downstream signalling. Conversion of 2-O-Desulfated Heparin from its sodium salt to either the copper or potassium forms in a FGF-2:2-O-Desulfated Heparin:FGFR complex switched its activity from being inactive to being active (Rudd et al., 2007). Similarly, FGF-1:N-Acetylated Heparin:FGFR signalling complexes were seen to be active when the carbohydrate was present as its sodium salt, however, when present as its copper salt this complex was seen to be inhibitory (Guimond et al., 2009).

1.6 – Project Aims:

As described above, carbohydrates have wide ranging applications in natural processes, as well as within both pharmaceutical and commercial products. Many of these applications relate to the carbohydrate’s structure and its consequent physical, chemical, and interaction properties. Therefore, accurate structure characterisation protocols would be required for the analysis of starting material samples as well as during modification monitoring. One facet of the carbohydrates’ structure which generates the desired outcomes within commercial and biopharmaceutical products is their ability to interact with other biological compounds, notably proteins. These interactions with proteins can vary considerably as a result of structural modifications to the carbohydrate, with effects also exhibited in any consequent biological processes. Therefore, both carbohydrate modifications and interactions require accurate characterisation in terms of the sites and degree of modification as well as interacting partners and interaction kinetics, respectively.

As such, the aim for this research is to develop an NMR based carbohydrate structure characterisation method, for use on both starting materials and modified samples, with additional methods employed for the characterisation of carbohydrate-protein interactions. To achieve this aim a series of targeted objectives have been generated, these are:

- Development of an NMR-based method capable of collecting spectra with resolution at the peaks’ natural linewidth (Chapter 4).
- Application of this method to a series of samples from different suppliers (Chapter 4).
- Development of a pyridine sulfur trioxide based sulfation method using an ionic liquid as the only solvent present, i.e. the reaction mixture being wholly absent of traditional organic solvents (Chapter 5).
- Application of the carbohydrate structure characterisation method developed to a series of sulfated samples produced using the sulfation protocol developed (Chapter 5).
- Screening of the starting material (where possible) and sulfated carbohydrate samples produced for interactions with a select group of proteins (Chapter 6).
- Application of further biophysical interaction characterisation methods to a select group of the screened carbohydrate and protein combinations shown to interact (Chapter 6).

Development of an NMR based method with resolution at the peaks’ natural linewidth for carbohydrate structure characterisation is required due to their complexity. Coupled with this complexity, the repeating unit variants present within a polysaccharide can be near identical to one another, making them difficult to distinguish. NMR is the best method for carbohydrate structure characterisation particularly with regard distinguishing the structures of near identical repeating units. However, the current methods for carbohydrate structure characterisation do not separate signals to the extent required for a detailed analysis of structure composition or for modification monitoring. Therefore the development of a method capable of resolution at the peaks' natural linewidth is required, with test experiments important to gauge the effectiveness of the method developed with regards real world applications, such as testing samples from different suppliers.

The ability to produce and analyse a variety of modified carbohydrate samples is important for compound development in biopharmaceutical, commercial, and industrial applications. Therefore, the NMR-based carbohydrate structure characterisation method developed will be applied to a series of sulfated samples produced using a method developed wherein the sole solvent present is an ionic liquid. Traditional pyridine sulfur trioxide based carbohydrate sulfation methods require sample pre-treatment to convert them into a hydrophobic salt, which can prove very difficult for gel-forming or water insoluble samples, whilst potentially reducing final sample yield. Ionic liquids do not require such pre-treatment steps and can be used as a means to overcome solubility issues for native carbohydrate starting materials whilst retaining some of the anhydrous properties that are required for sulfation reactions.

Finally, proteins form the basis of many biological processes as a result of their interactions with carbohydrates. As mentioned above, previously published research has identified instances where a carbohydrate's ability to interact with a protein can be mimicked by other carbohydrates which have been suitably modified. Several of these examples relate to carbohydrate sulfation, therefore the sulfated samples produced and analysed will be screened for interactions with FGFs 1 and 2 as well as cyclophilins B and D. This should identify whether any of the modified carbohydrates produced can be used in place of previously characterised interaction partners.

Investigation of the overall aim, and the stated objectives, for this research project will necessitate the use of multiple methods to generate the data required. An outline of how these methods function and how they have been used for data collection, in similar projects, previously has been included as a "Technical Introduction" in the following chapter.

Chapter 2 – Technical Introduction

The analysis of carbohydrates in terms of their structure and interactions with other molecules is a multi-faceted process and many techniques are required. This chapter will introduce some of those that have been used to generate the data presented in the results chapters that follow. The carbohydrates studied have known applications as biopharmaceuticals, with modes of action reliant on interactions with proteins, or as gelling agents used in products across the home and personal care, and food and drink sectors. The methods used include extensive applications of Nuclear Magnetic Resonance (NMR) spectroscopy alongside other biophysical analysis techniques to determine the carbohydrates' structures, monitor modifications made to these structures, and to investigate carbohydrate-protein interactions.

2.1 – Nuclear Magnetic Resonance:

Nuclear magnetic resonance spectroscopy is a particularly powerful analytical tool owing to its ability to characterise compounds in a non-destructive manner, without being reliant on a unique mass or the presence of a chromophore for detection. Further to this the nature of NMR being sensitive to the chemical structure of the compounds measured means that information regarding functional groups, connectivity and conformation can be generated. Consequently, the applications of NMR are various including, but not limited to, the structural characterisation of small molecules, screening of small molecules, the investigation of both protein structure and their interactions, as well as in the field of metabolomics (Duus et al., 2000; Ciulli, 2013; Purslow et al., 2020; Barbar, 2004; Mielke and Krishnan, 2009; Bonvin et al., 2005; Kogelberg et al., 2003; Emwas et al., 2019). However, despite the advantages of this technique, there are also a number of disadvantages surrounding its relatively low sensitivity and, hence, significant sample requirements. Whilst there are two principal branches of NMR, solution and solid state, the research that follows was carried out in solution, therefore the remainder of this introduction will be tailored towards solution studies and further reference to NMR, beyond the general principles, specifically means solution NMR spectroscopy.

NMR spectroscopic techniques are dependent on the quantum mechanical property of spin, possessed by the nuclei of some of the constituent atoms present within a molecule. The value given to this nuclear spin, denoted I, differs depending on the atom present, increasing in increments of ½, with some of the spin values for biologically relevant atoms listed in Table 1 (Claridge, 2016):

Table 1: Spin values and natural abundances of select biologically relevant atoms (Harris et al., 2001; Hagan, 2006).

Atomic Isotope	Spin Value (*I*)	Natural Abundance (%)
^{1}H	1/2	99.9
^{2}H	1	0.01
^{12}C	0	98.93
^{13}C	1/2	1.07
^{14}N	1	99.63
^{15}N	1/2	0.37
^{17}O	5/2	0.04
^{19}F	1/2	100
^{31}P	1/2	100
^{33}S	3/2	0.76

Each of the nuclei which generates a spin also have an angular momentum that is capable of forming a magnetic moment in a manner dependent on the gyromagnetic ratio of the respective nucleus. Those which do not exhibit a spin, e.g., ^{12}C, are unobservable by NMR and are described as NMR silent (Claridge, 2016). Whilst some of the nuclei useful for biologically relevant NMR are naturally present at a high abundance, e.g. ^{1}H, and ^{19}F, others are not, e.g., ^{13}C and ^{15}N, Table 1. Insertion of a sample into a strong magnetic field leads to its constituent nuclei forming magnetic moments. This strong magnetic field, denoted B_0, causes the moments formed to align themselves with or against B_0, leaving a small net equilibrium magnetisation aligned with the field, which is the cause of the comparative insensitivity of NMR techniques. Once formed, this net magnetisation precesses at a slight angle to that of B_0 in a process termed Lamour prescession (Claridge, 2016; Keeler, 2010). The orientation of magnetic moments about B_0 results from the number of energy levels which can be inhabited. This number can be calculated using the equation $2I + 1$ which for a spin-half nucleus such as ^{1}H would give two energy levels, +1/2 and -1/2 otherwise termed α and β or spin up and spin down respectively (Claridge, 2016; Keeler, 2010).

The net magnetisation precessing at Lamour frequency can be manipulated by the application of a radiofrequency pulse, denoted B_1, perpendicular to B_0 at a frequency resonant with that of Lamour precession. Application of B_1 pulses for differing durations consequently causes the net magnetisation to be rotated by differing angles, where the most common are 90^{o} and 180^{o}, and can be applied along the x, y, or z axes (Keeler, 2010). Combinations of these radiofrequency pulses as defined programs form the experiments performed when measuring NMR spectra. Detection of the magnetisation recorded in these spectra occurs when it 'cuts' a coil surrounding the bore into which the measured sample was inserted in an orientation perpendicular to that of the B_0 field with the currents generated amplified during the free induction decay (Keeler, 2010). Subsequently, the free induction decay (a function of time) which can be observed, but is not easily decipherable, is subjected to Fourier transformation (to provide a function of frequency) in order to generate a decipherable spectrum, which can then be analysed. These spectra are then displayed with peaks presented at chemical shift values relative to a reference frequency, rather than the frequency themselves, to facilitate comparison between spectra recorded on different magnets with potentially differing magnetic field strengths. Conversion between the frequency measured and the chemical shift presented utilises the equation $\delta(\text{ppm}) = ((\upsilon - \upsilon_{\text{ref}})/\upsilon_{\text{ref}})*10^6$ (Keeler, 2010).

The combination of different sequences of radiofrequency pulses forms the basis of NMR experiments. Varying the pulses applied therefore enables different information to be presented in the spectra generated. The NMR experiments performed during this project will therefore be introduced alongside how the information received from the analysis of these spectra can be used for carbohydrate structure characterisation and modification monitoring, magnetisation transfer diagrams for these experiments are shown in Figure 9.

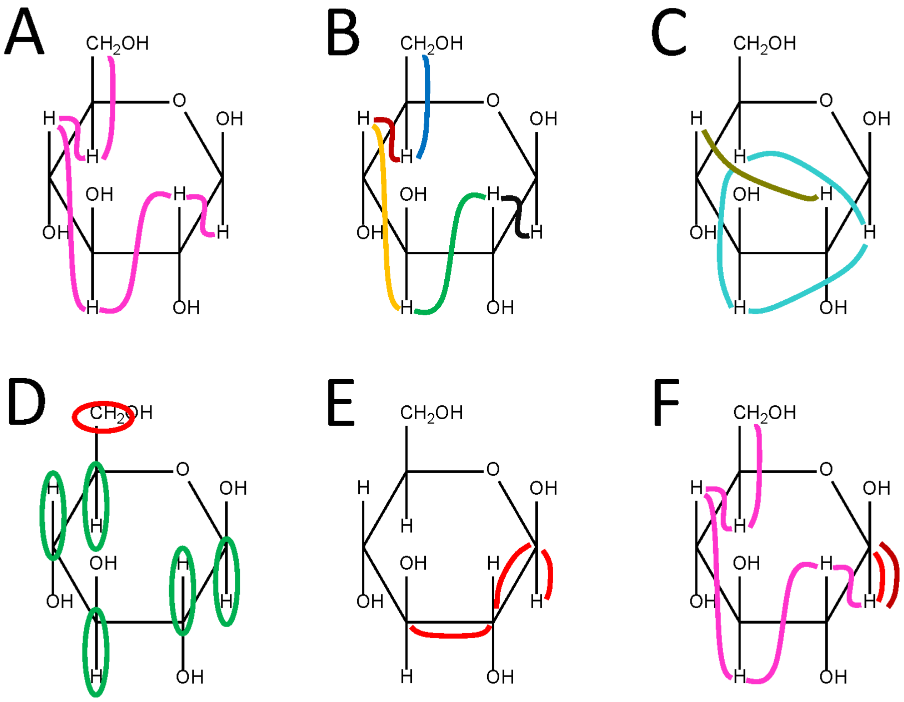


Figure 9: Magnetisation transfer diagrams for 2D ^{1}H,^{1}H-TOCSY (A), 2D ^{1}H,^{1}H-COSY (B), 2D ^{1}H,^{1}H-NOESY (C), 2D ^{1}H,^{13}C-HSQC/1D ^{13}C-DEPT (D), 2D ^{1}H,^{13}C-HMBC (E), and 2D ^{1}H,^{13}C-HSQC-TOCSY (F) experiments used during the research carried out for this project. The pink lines indicate the transfer of magnetisation through a series of protons within a spin system (A and F), while black, green, yellow, burgundy, and blue lines indicate this between protons bound to neighbouring carbon atoms (B), and cyan and khaki lines indicate transfer between protons located within 5Å of one another (C). In D) the green

ovals indicate secondary CH groups, and those is a red oval are primary CH_2 groups, while in E) the red lines indicate magnetisation transfer from proton to carbon atoms and then along part of a chain of carbons, and in F) the red and burgundy lines indicate the transfers between proton and carbon atoms.

2.1.1 – 1D ^1H Experiments:

1D ^1H NMR spectra contain signals for each of the protons present within the target compound. The 1D nature of these spectra, however, coupled with the limited range of separation on the ^1H axis means that peaks are often overlapped or otherwise difficult to distinguish from others within the same region of the spectrum. This is particularly relevant for carbohydrate samples, which have a very limited range of observable proton groups, i.e. CH, CH_2, or, less commonly, CH_3, while those of the hydroxyl and amine functional groups are rarely visible owing to their labile nature and consequent exchange with deuterium, which further limits the spread of peaks throughout a spectrum. Consequently in order to complete an in-depth structural characterisation these 1D ^1H experiments are best coupled with other, 1D ^{13}C and 2D, spectra. 1D ^1H experiments can also suffer loss of peaks which are overlapped by the water (HDO) signal; therefore several different water suppression techniques have been included as part of their pulse programs, with pre-saturation used most commonly. As can be seen from Figure 10 the spectra generated consist of a series of 1D peaks, with splitting evident on some. This splitting emerges due to coupling constants that result from geminal and/or vicinal hydrogen nuclei. When either a geminal or vicinal hydrogen is present this generates a two-spin system, resulting in four energy levels $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, and $\beta\beta$. Consequently, there are two possible $\alpha \rightarrow \beta$ transitions for spin one, i.e. $\alpha\alpha \rightarrow \beta\alpha$ and $\alpha\beta \rightarrow \beta\beta$ causing a doublet to emerge, by contrast for a one spin system when the proton represented by a peak has neither a geminal nor a vicinal hydrogen, there is only one possible $\alpha \rightarrow \beta$ transition, generating a singlet (Keeler, 2010). The pattern seen here, where the addition of a second spin doubles the number of lines into which a peak splits, follows for the addition of further spins, with each added spin causing the number of the peak's lines to double. The overlapping nature of these peaks can also be seen in Figure 10 to a limited extent within the anomeric region, here 4.3ppm to 5.3ppm, but more noticeably within the non-anomeric region, here 3.2ppm to 4.2ppm. The anomeric peaks will be of particular importance for the analysis of 2D spectra as their location, away from the other non-anomeric peaks, can provide a starting point from which connections can be traced. Anomeric peaks are shifted in this manner owing to the effects on their corresponding proton atom by both the electronegative position one and five oxygen atoms.

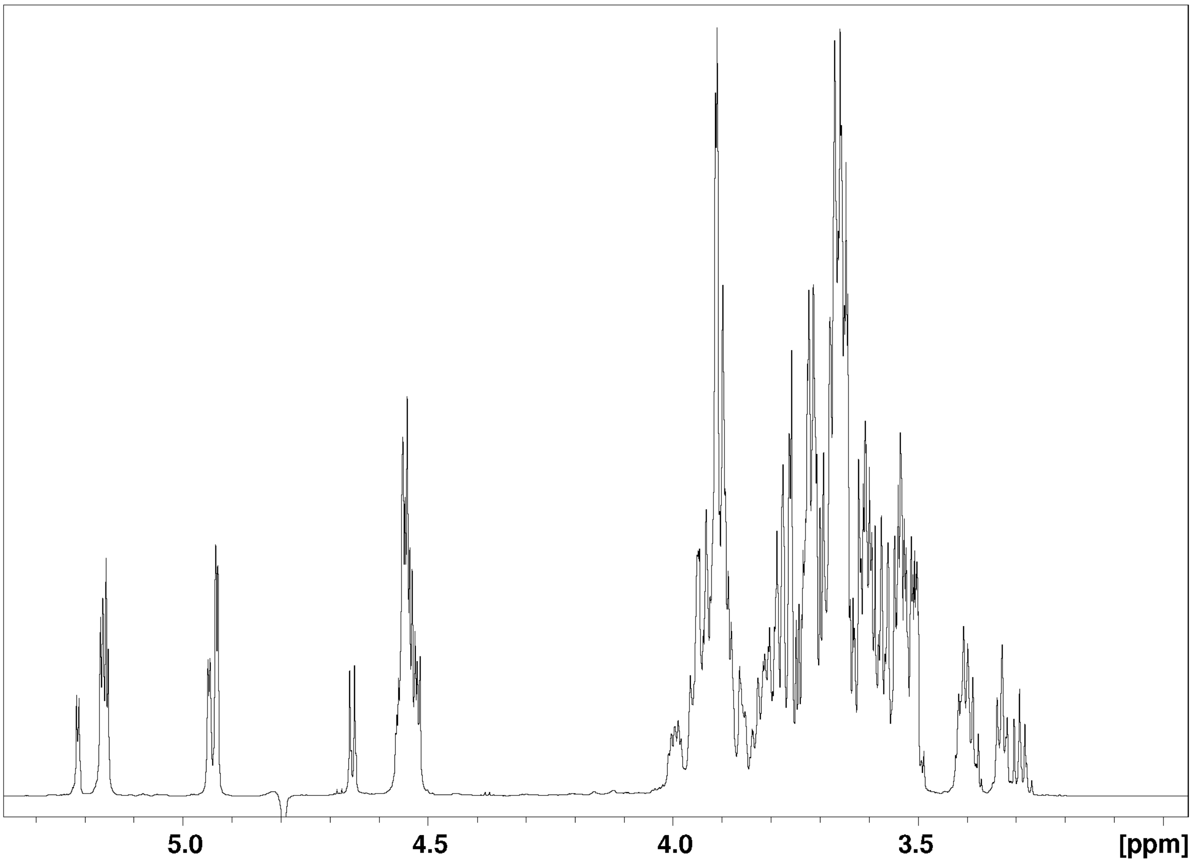


Figure 10: 1D ^{1}H spectrum of a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments.

2.1.2 – 1D ^{13}C Experiments:

As with 1D ^{1}H NMR spectroscopy, 1D ^{13}C experiments provide a means by which peaks can be generated for each of the carbon atoms present within the compound of interest. Whilst the same difficulties regarding crowding also occur in certain regions of 1D ^{13}C spectra the greater range of chemical shift values serves to reduce this effect considerably. The low proportion of ^{13}C isotopes compared to those of ^{12}C also simplifies these spectra as it would be unlikely for two ^{13}C atoms to be present next to each other, and as such ^{13}C-^{13}C interactions are unlikely to be present to cause signal splitting. Within the 1D ^{13}C experiments performed as part of this project a composite pulse decoupling (CPD) element was typically included to simplify the spectra produced by decoupling proton and hence removing any splitting which occurs as a result of the protons bound to carbon atoms. An example of the spectra which result from the pulse program typically used in this project is shown in Figure 11 with a series of 1D peaks evident, and splitting is less apparent.

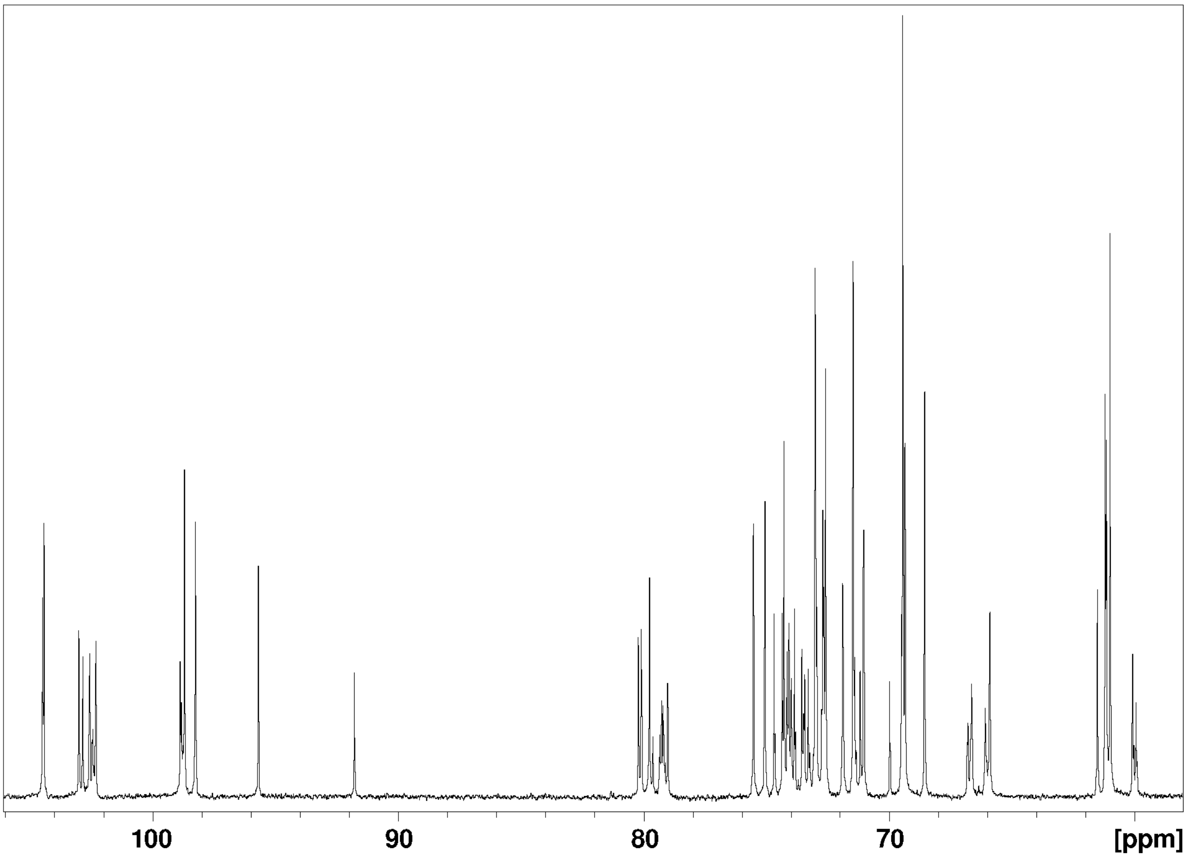


Figure 11: 1D ^{13}C spectrum for a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments.

2.1.3 – Distortionless Enhancement by Polarisation Transfer (DEPT) Experiments:

1D ^{13}C-DEPT spectra display a series of peaks across the same chemical shift range as the 1D ^{13}C spectra. However, unlike the 1D ^{13}C experiments transfer of magnetisation is required during these experiments, hence any quaternary carbons present within the sample will not be detected. Three variants of DEPT experiment have been developed; DEPT45, DEPT90, and DEPT135, of which only DEPT135 experiments were performed in this project. These three experiments each generate different spectra as a result of the different angle of rotation generated by the final radiofrequency pulse on proton. The experiments where this pulse results in a 45° angle of rotation, DEPT45, will display all non-quaternary carbon peaks in a positive manner. Adjusting this pulse such that the angle generated is 90°, i.e. DEPT90 experiments, results in spectra showing the peaks for tertiary carbons which are displayed in a positive manner, while if the angle produced is 135°, in DEPT135 experiments, both primary and tertiary carbons will be displayed in a positive manner whilst secondary carbons will be negative. The example spectrum shown in Figure 12 shows the tertiary carbons, between 71ppm and 110ppm, and the secondary carbons between 60ppm and 71ppm.

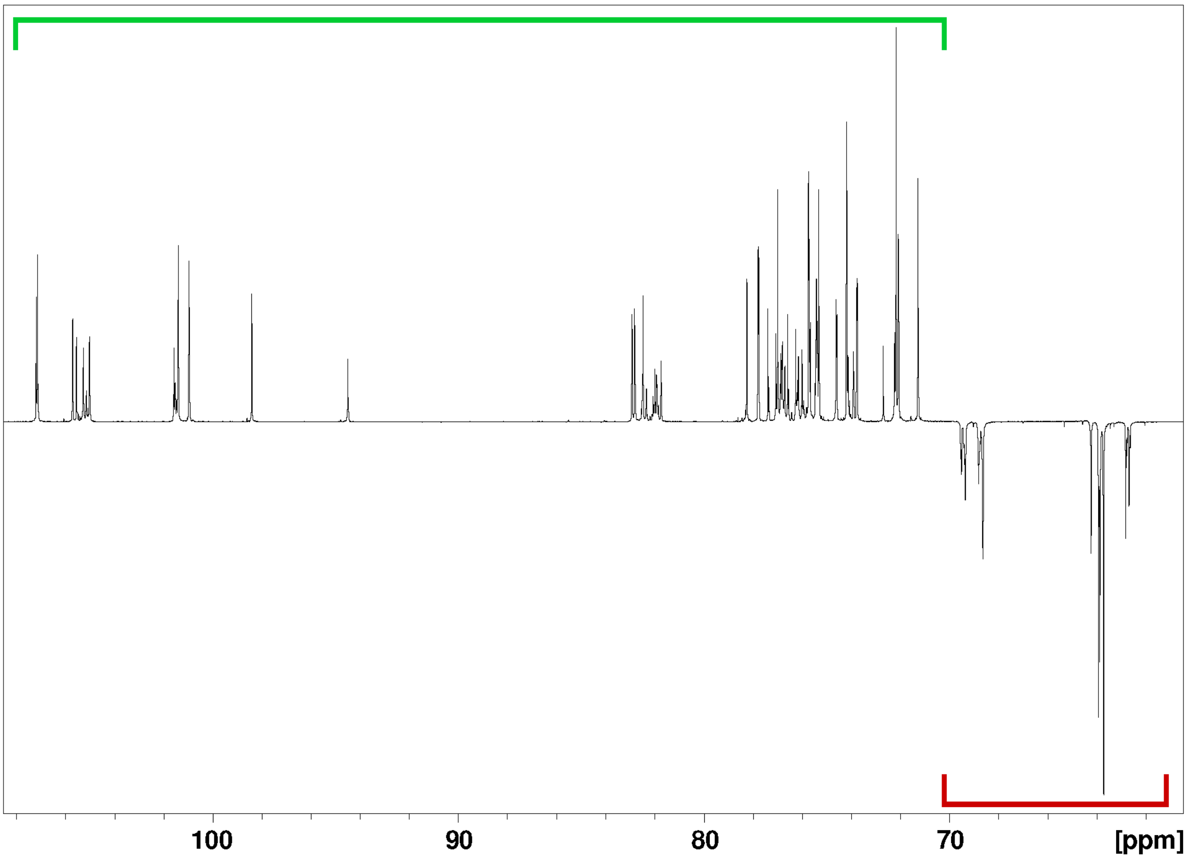


Figure 12: 1D ^{13}C-DEPT spectrum for a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments with peaks for CH groups in the region under the green bar and those for CH_2 groups in the region above the red bar.

2.1.4 – Correlation Spectrscopy (COSY) Experiments:

Whilst the 1D ^{1}H, 1D ^{13}C, and 1D ^{13}C-DEPT experiments have been introduced, the remaining experiments performed during this project were 2D, magnetisation transfer diagrams for these experiments are shown in Figure 9. One such 2D spectrum, a homonuclear 2D ^{1}H,^{1}H-COSY experiment, can be used to identify the chemical shifts for protons bound to carbon atoms which are coupled to each other. Consequently, 2D ^{1}H,^{1}H-COSY experiments can be used to identify the chemical shifts of neighbouring proton atoms and through this trace the order of atoms within a spin system. As a 2D experiment, overcrowding of spectrum regions is considerably reduced compared to what is observed in 1D ^{1}H experiments, however, in some regions of the spectra collected it can remain a problem. Use of 2D ^{1}H,^{1}H-COSY experiments which were double quantum filtered has served to improve the quality of the spectra produced as both cross-peaks and diagonal peaks can be phased rather than having to choose one or the other, with the presaturation element incorporated in experiments used in this project further improving the spectrum quality (Keeler, 2010), Figure 13**Error! Reference source not found.**.

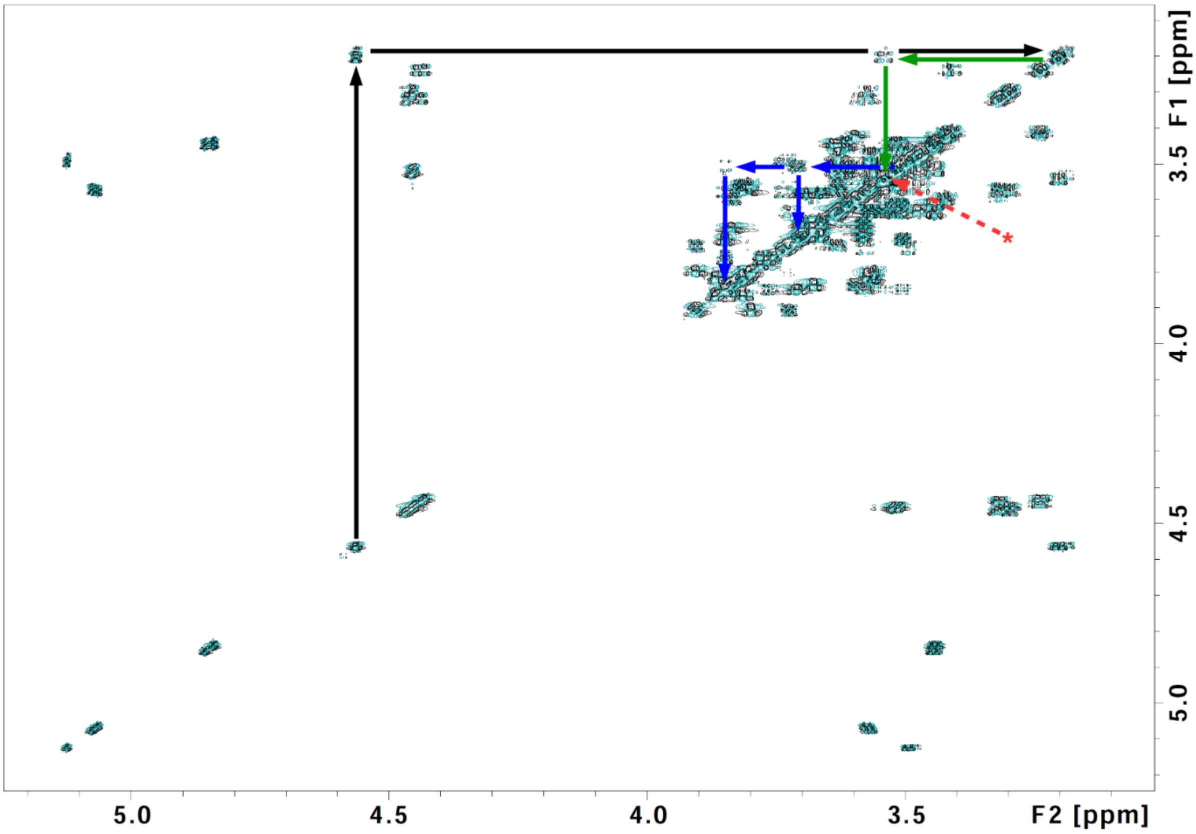


Figure 13: 2D ^1H,^1H-COSY spectrum for a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments with connections shown between the anomeric and C_2H positions (black), C_2H and C_3H (green), and C_5H and C_6H (blue), the red asterisk indicated the connections between C_3H and C_5H are too closely contained to be observable at this scale.

2.1.5 – Total Correlation Spectroscopy (TOCSY) Experiments:

Another type of 2D experiment, 2D ^1H,^1H-TOCSY spectra, functions via the transfer of magnetisation between protons bound to carbons within an unbroken chain of carbon atoms. This means that whilst 2D ^1H,^1H-TOCSY experiments cannot be used to track the order of carbon bound proton atoms within a chain, they can be used to highlight all of those present within the chain which need to be accounted for, this therefore lends itself to combinations with Correlation Spectroscopy (COSY) experiments, where transfer is only between a single set of neighbouring proton atoms, to provide a particularly powerful analytical tool. The key parameter within these experiments is the mixing time, which when set to be short only allows for the transfer of magnetisation across a part of the chain, or when set to be long could include the whole chain but may have detrimental effects. The spectra produced by 2D ^1H,^1H-TOCSY experiments consist diagonal peaks with cross peaks, to link those within the same spin system, Figure 14

.

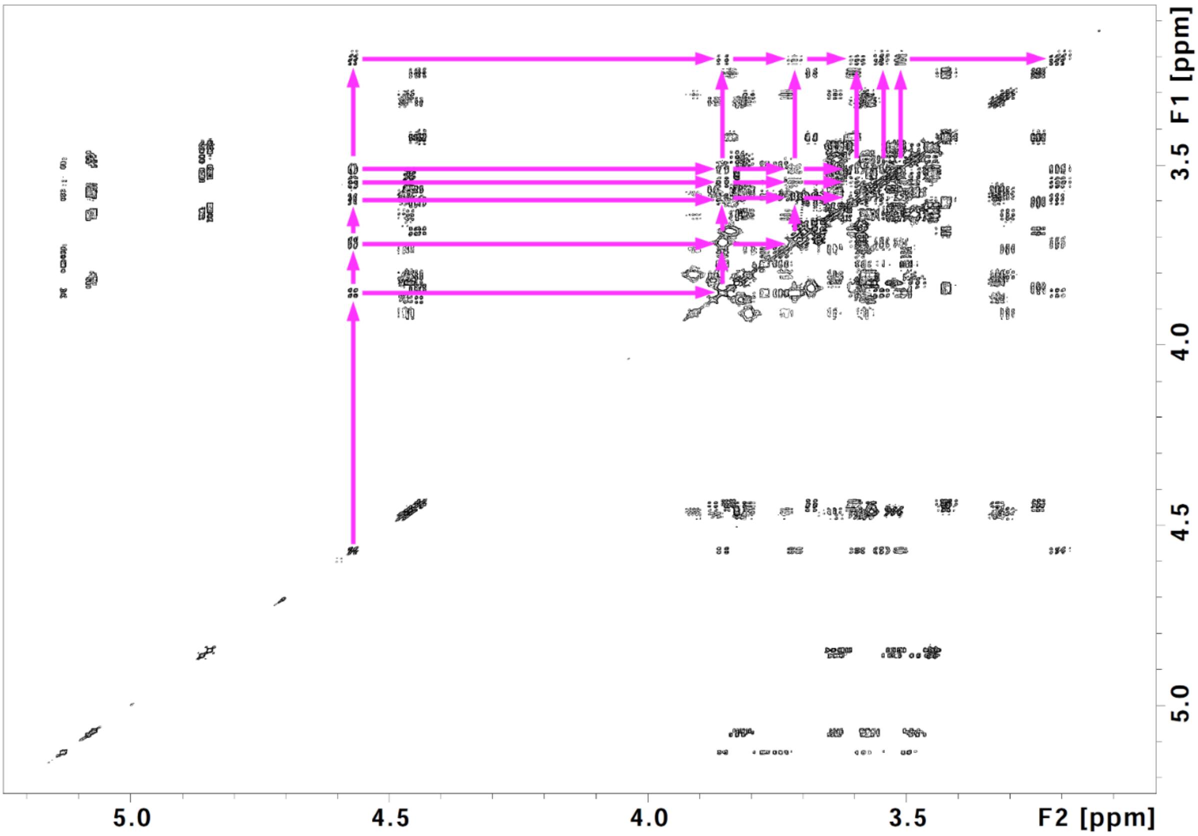


Figure 14: 2D ^1H,^1H-TOCSY spectrum for a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments with connections shown in magenta.

2.1.6 – Nuclear Overhauser effect (NOESY) and Rotating-frame nuclear Overhauser effect Spectroscopy (ROESY) Experiments:

Unlike the 2D experiments introduced thus far, where the through bond proximity of atoms has been determined, nuclear Overhauser effect methods, 2D ^1H,^1H-NOESY and 2D ^1H,^1H-ROESY, highlight atoms at a distance of less than about 5Å. As with 2D ^1H,^1H-TOCSYs, the most important parameter in both 2D ^1H,^1H-NOESY and 2D ^1H,^1H-ROESY experiments is the mixing time, which when short limits the distance between which magnetisation can travel for neighbouring proton atoms. While both experiments provide the same through space atom proximity information they differ in terms of the size of compound they are best suited to analyse where 2D ^1H,^1H-NOESYs are better suited for molecules which are either large or small while 2D ^1H,^1H-ROESY experiments work best for intermediately sized compounds (Raghothama, 2010). This is due to the nuclear Overhauser effect transitioning between being positive and negative, and hence being roughly zero, at the point corresponding to intermediately sized compounds, consequently limiting the spectra that can be recorded for them. In contrast to this, the nuclear Overhauser effect present for 2D ^1H,^1H-ROESY spectra does not transition through zero, therefore it can fill in the gap left by 2D ^1H,^1H-NOESY spectra. Both the spectra recorded for 2D ^1H,^1H-NOESY and 2D ^1H,^1H-ROESY experiments are very similar, in a manner which also matches the 2D ^{1}H,^{1}H-TOCSY spectra recorded with a series of diagonal and cross peaks present mostly, but not exclusively, for protons within the same spin system, an example 2D ^{1}H,^{1}H-NOESY spectrum is shown in Figure 15.

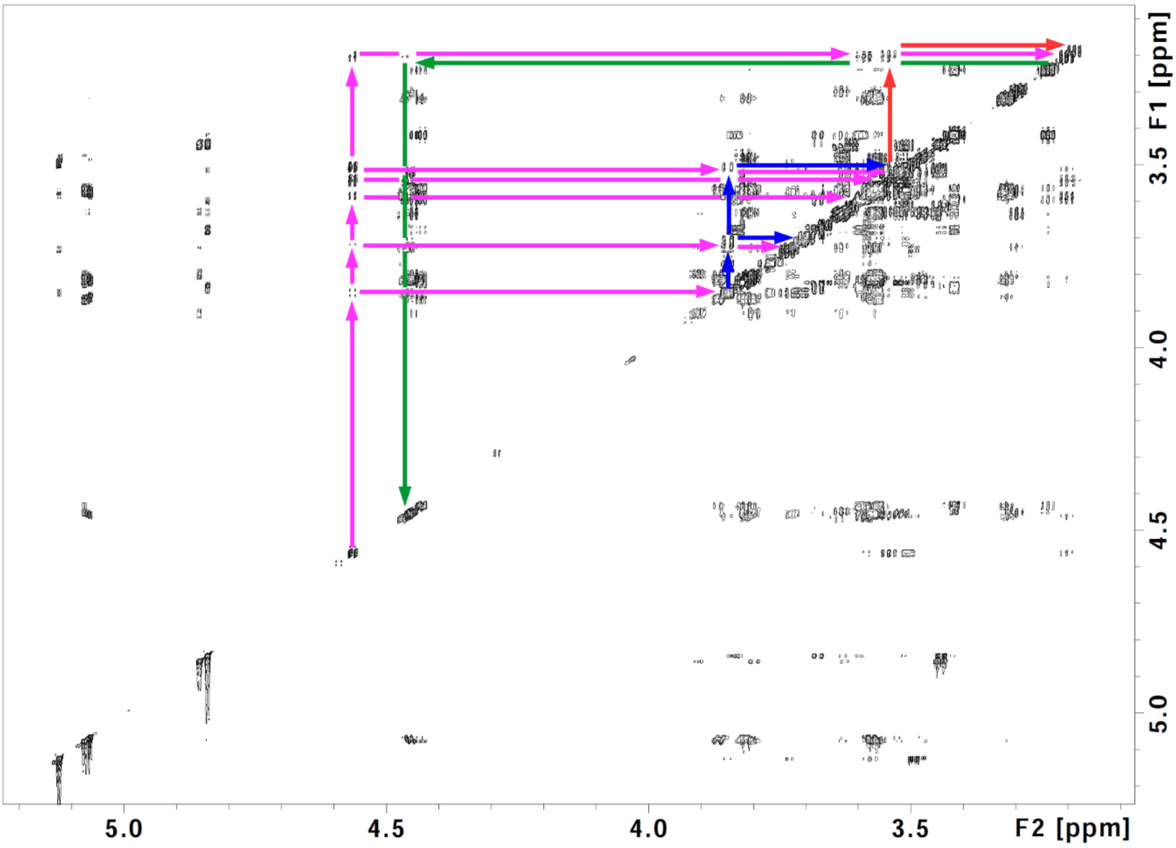


Figure 15: Example 2D ^{1}H,^{1}H-NOESY spectrum for a xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating unit fragments with connections shown in magenta for those observable from the anomeric proton, green for that observable from C$_2$H back to the anomeric proton of another residue, and red and blue for connections observed from non-anomeric protons.

2.1.7 – Heteronuclear Single-Quantum Correlation (HSQC) Spectroscopy Experiments:

The 2D ^{1}H,^{1}H-COSY, 2D ^{1}H,^{1}H-TOCSY, 2D ^{1}H,^{1}H-NOESY, and 2D ^{1}H,^{1}H-ROESY experiments introduced thus far have all been homonuclear, involving ^{1}H nuclei. In contrast, the 2D ^{1}H,^{13}C-HSQCs measured are heteronuclear and whilst spectra correlating ^{1}H and ^{15}N atoms are available, the experiments performed in this research all generated spectra correlating ^{1}H and ^{13}C atoms. These experiments rely on the insensitive nuclei enhanced by polarisation transfer (INEPT) block of radiofrequency pulses to transfer magnetisation from its starting point on proton atoms to the other atom type being measured, here ^{13}C. This is followed by a period of precession for the second atom before a reverse INEPT block transfers magnetisation back to the protons prior to detection. Unlike each of the 2D experiments previously introduced, the spectra recorded from 2D ^{1}H,^{13}C-HSQC experiments do not highlight linkages as cross peaks between diagonals. Instead of this, the peaks indicate the chemical shifts for each C–H pair of atoms on both ^{13}C and ^{1}H axes in this case. The principal benefit of 2D ^{1}H,^{13}C-HSQC experiments is to combine the peak separation seen for both 1D ^{1}H and 1D ^{13}C experiments such that instances of peaks overlapping a rare. These spectra therefore increase the ease with which individual atoms, or C–H atom pairs, can be assigned within a spectrum and consequently aid the structural characterisation as well as the monitoring of both modifications and interactions seen for the target compound, Figure 16.

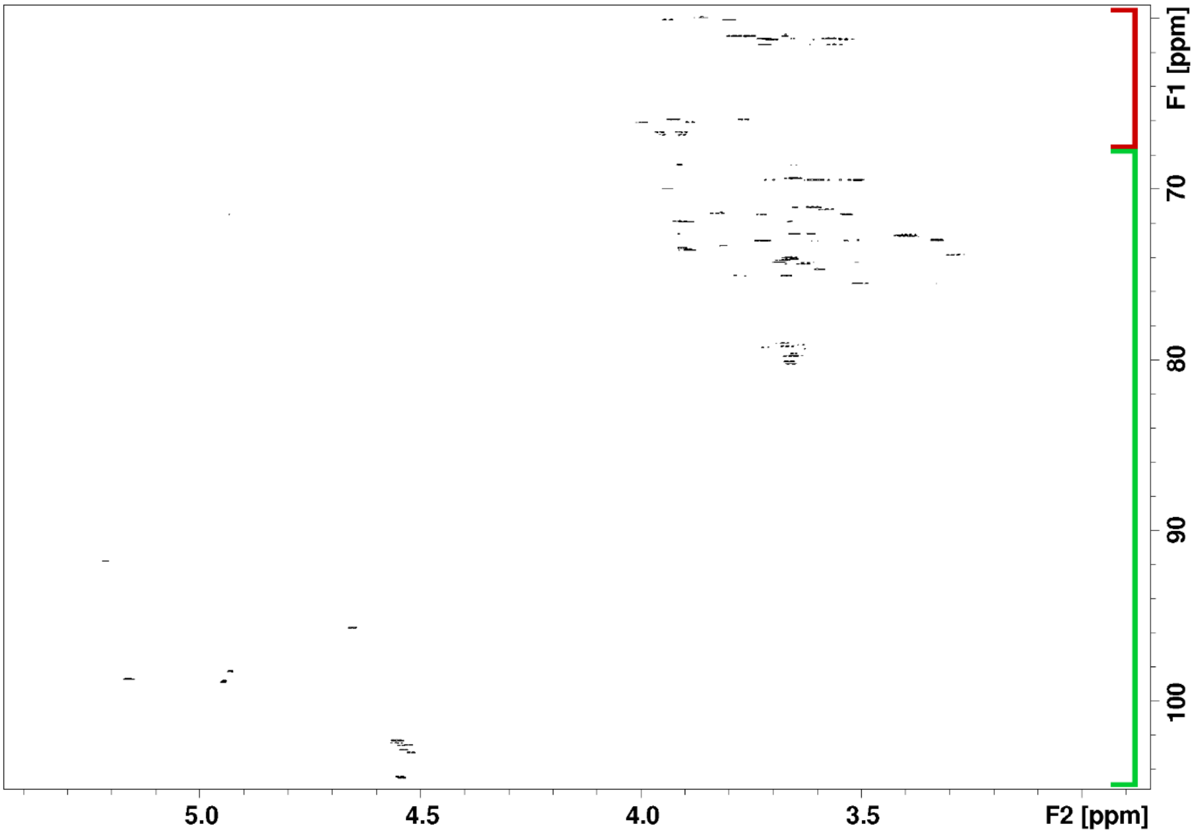


Figure 16: 2D ^{1}H,^{13}C-HSQC spectrum for a Xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating units fragments with CH groups shown in the region to the left of the green bar and those for CH_2 groups to the left of the red bar.

When measuring these spectra, as well as the 2D ^{1}H,^{13}C-HMBC and 2D ^{1}H,^{13}C-HSQC-TOCSY spectra which remain to be introduced, at high resolution, the timeframe for experiment completion becomes too long to be practical, therefore, non-uniform sampling methods are applied to reduce the time required. Non-uniform sampling measures a random selection of the indirect dimension increments, where the proportion of the total increments to be sampled can be set prior to acquisition, in a manner in where the missing increments can be backfilled during processing with multidimensional decomposition and/or compressed sensing equations. Coupled with the implementation of non-uniform sampling within the spectra recorded, switching from CPD to adiabatic decoupling allowed for greater acquisition times to be used by reducing the power of pulses being directed towards the probe, which in turn enables the measurement of higher resolution spectra.

2.1.8 – Heteronuclear Multiple Bond Correlation (HMBC) Spectroscopy Experiments:

2D $^{1}H,^{13}C$-HMBC experiments, like 2D $^{1}H,^{13}C$-HSQCs, are heteronuclear with correlations between the nuclei of ^{1}H and ^{13}C atoms used in the experiments performed during this research. However, whilst 2D $^{1}H,^{13}C$-HSQCs transfer magnetisation between bound proton and carbon nuclei, 2D $^{1}H,^{13}C$-HMBC experiments can then further transfer this magnetisation through one to three other atoms terminating at a ^{13}C, for the experiments performed in this case. Whilst the spectra derived from these experiments do not show peaks on a diagonal, they do contain the cross peaks between C–H atom pairs presented in a 2D $^{1}H,^{13}C$-HSQC spectrum, Figure 17. Consequently, combining 2D $^{1}H,^{13}C$-HSQC and 2D $^{1}H,^{13}C$-HMBC experiments proves particularly powerful analytically for the structural characterisation of target samples. As with the other 2D heteronuclear experiments, the separation of peaks across two axes minimises the potential for overlap, particularly when spectra are recorded at high resolution, however, some instances of this remain.

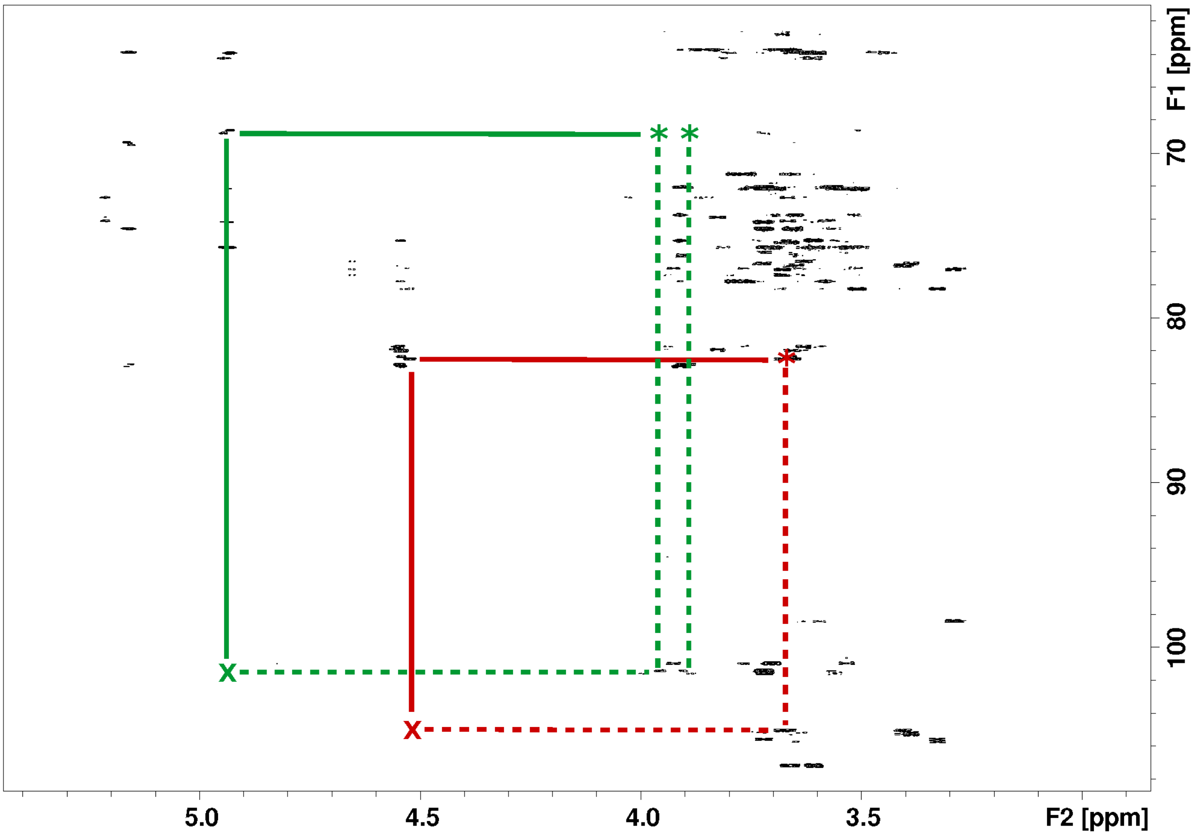


Figure 17: 2D $^{1}H,^{13}C$-HMBC spectrum for a Xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating units fragments with the linkage between two of the internal backbone glucose residues shown in red from position one of one residue (red X) and four of the other (red *) via a 2D $^{1}H,^{13}C$-HMBC cross peak. Also shown, in green, is the connection between position one of a xylose residue (green X) and position six of a backbone glucose (green *).

2.1.9 – Heteronuclear Single-Quantum Correlation Total Correlation Spectroscopy (HSQC-TOCSY) Experiments:

The final experiment to be introduced, 2D $^{1}H,^{13}C$-HSQC-TOCSYs, are a concatenation of the 2D $^{1}H,^{13}C$-HSQC and 2D $^{1}H,^{1}H$-TOCSY experiments previously introduced above. Consequently, the magnetisation transfers exhibited by this experiment are also a combination of the two experiments with an initial transfer of magnetisation from proton to carbon nuclei before returning to proton and transferring along each carbon bound proton atom within an unbroken chain by means of the methods previously introduced. The spectra which result from these experiments therefore have both the C–H atom pair peaks, seen in a standard 2D $^{1}H,^{13}C$-HSQC, alongside a series of cross peaks linking each of these C–H pair peak positions present for atoms within a spin system, Figure 18. The fact that there would consequently be multiple peaks for each hexose residue means that the spectra generated for these experiments can become very busy, with a greater potential than either the 2D $^{1}H,^{13}C$-HSQC or 2D $^{1}H,^{13}C$-HMBC experiments for overlapping peaks to complicate spectrum analysis.

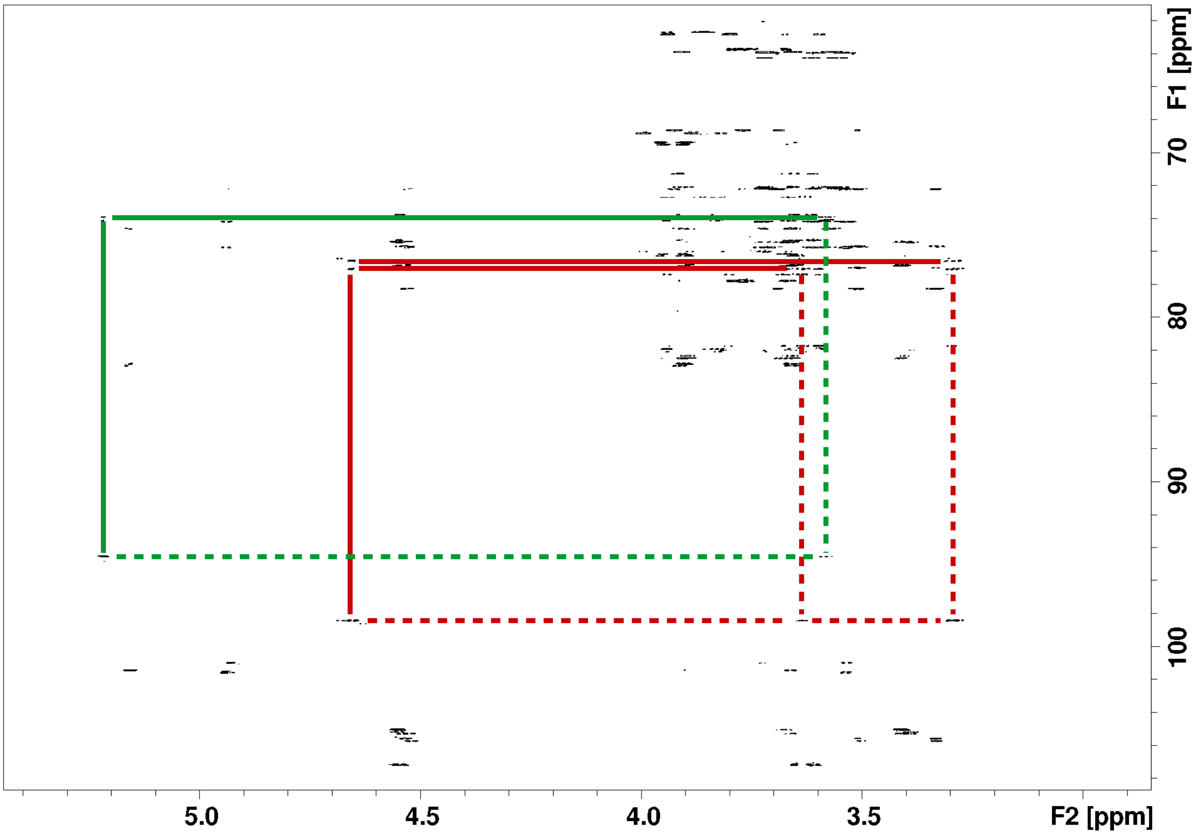


Figure 18: 2D $^{1}H,^{13}C$-HSQC-TOCSY spectrum for a Xyloglucan sample hydrolysed to between septamer and nonamer oligosaccharide repeating units fragments with connections shown between the reducing end α- (red) and β- (green) D-glucose residues' anomeric protons and other proton positions with their respective residues.

2.1.10 – NMR for Carbohydrate Structural Analysis:

NMR is a particularly powerful tool for the structural analysis of carbohydrates owing, amongst other factors, to its independence from the target compound's mass and ultraviolet (UV) activity, whilst also providing a wealth of data for the sample measured from the spectra recorded. The stereoisomers, epimers and anomers of both aldose and ketose residues of the same carbon chain length, introduced previously, give rise to a vast array of residues each with identical or near-identical masses making analyses via mass spectrometry (MS) alone unreliable. Similarly, the lack of UV activity in most of the monosaccharides which comprise complex carbohydrates prevents both detection and analysis via UV-based techniques. Using NMR for carbohydrate analysis provides a number of benefits that alternative structure characterisation methods do not, such as the ability to differentiate between hydroxyl groups being located either axially or equatorially to the plane of its monosaccharide's ring. This is important when determining residue conformation and configuration including anomer identification, which when applied to other CH groups, such as C_4H, can be used to determine which residue type, for example glucose or galactose, is present. Whilst MS alone could not be relied upon during carbohydrate structure characterisation, analysing a carbohydrate with both MS and NMR is particularly powerful. Masses determined via MS could indicate the likely number of residues present within an oligosaccharide structure that could help guide NMR spectrum analysis, which would be used to determine what the residues present are.

Equally, the different NMR methods introduced, with the resulting spectra combined for analysis, can be used to comprehensively characterise an oligosaccharide structure. A typical process by which this could occur would involve the collection of high resolution 1D ^{1}H, 1D ^{13}C and 2D ^{1}H,^{13}C-HSQC spectra to identify each atom. Comparison of 1D ^{13}C and 2D ^{1}H,^{13}C-HSQC spectra could also be used to determine if any quaternary carbons are present within the structure, which was not the case for the carbohydrates studied here. Analysis of either 2D ^{1}H,^{1}H-TOCSY or 2D ^{1}H,^{13}C-HSQC-TOCSY spectra then determines the atoms present within each observable spin system, or residue, that comprise the carbohydrate. Having determined which peak positions correspond to the same spin system, 2D ^{1}H,^{1}H-COSY spectra would be used to determine the order of these atoms within the spin system in a manner that would lead to assignments for each of the residues present but would not say anything about the linkages between them. These linkages would then be determined by 2D ^{1}H,^{13}C-HMBC experiments with magnetisation transfer across glycosidic bonds generating cross peaks. 2D ^{1}H,^{1}H-NOESY and 2D ^{1}H,^{1}H-ROESY spectra could then be used to support the case for any linkages determined provided the distance between the protons attached to glycosidically linked carbons did not exceed 5Å. The final element of the characterisation process would be the determination of the anomeric configuration for each of the atoms present which would be achieved either using 1D ^{1}H experiments, to measure the $^{3}J_{HH}$ couplings, provided overlapping peaks do not preclude it, or a 2D $^{1}H,^{13}C$-HSQC spectrum measured at sufficiently high resolution.

The nature of carbohydrates means that peaks are present in close proximity to that of water, therefore a suppression method would commonly need to be applied, however, this needs to be chosen correctly as something too severe, such as excitation sculpting, could lead to important peaks being ablated. Equally, any attempt to make the spectra recorded quantitative, via integration or maximum amplitude analysis, could be effected by the water suppression method used if those peaks present within the same chemical shift range as water are altered in terms of either their intensity or amplitude respectively. Once assigned, and potentially quantified, modifications made to the samples can be characterised using some of the experiments from the process described. Those experiments required to monitor these modifications will differ depending on which modification is being performed. 1D ^{1}H experiments as well as 2D $^{1}H,^{13}C$-HSQCs could be used to identify the differences resulting from modification, with further 2D $^{1}H,^{1}H$-COSY and 2D $^{1}H,^{13}C$-HMBC experiments of particular importance if peak shifting is exhibited, or additional peaks are generated.

Whilst the NMR experiments detailed above are a powerful tool for the analysis of carbohydrate samples, using them in combination with additional techniques that measure an independent property of the carbohydrate accentuates this analytical capability. The complementary technique used differed depending on the research being carried out, therefore in Chapter Four NMR was paired with mass spectrometry (MS), to determine whether the mass-to-charge ratios of the repeating units present matched the assignments made. However, for the sulfation experiments performed in Chapter Five NMR was paired with Fourier transform infrared (FTIR) spectroscopy such that sulfation was detected both indirectly, via CH chemical shift changes in NMR, and directly, via S=O and S-O vibrations in FTIR.

2.2 – Mass Spectrometry:

MS has had multiple applications during the structural analysis of carbohydrates and a number of protocol variations have been implemented in prior studies. These methods have been targeted to the monomer composition of oligosaccharides during glycan analysis alongside the structural characterisation of sulfated glycopolymers, however, MS performed in branched polymers has proved more difficult to successfully implement. In certain cases initial carbohydrate modification is required, via peralkylation or similar, to ensure that equal ionisation is achieved across the sample, or samples, tested despite the potential negative effects of partial sample loss, amongst others (Zaia, 2004; Kailema et al., 2014). Continuing advances in the analysis of carbohydrates by MS have generated improved ionisation methods alongside improved methods for quantification and novel software systems for carbohydrate MS data analysis. Further to these advances coupling MS methods to chromatography protocols has enabled improved carbohydrate structural characterisation, however, as these methods have not been used in this research they will not be introduced in any greater detail (Kailema et al., 2014). The use of advanced MS methods has enabled both oligosaccharide structural and positional isomer characterisation as well as that for anomeric configuration, particularly when ion mobility mass spectrometry is employed, although differentiation of stereoisomers remains out of reach (Ashline et al., 2007; Fenn and McLean, 2010; Hofmann et al., 2015).

Mass spectrometry methods differ dramatically based on the components contained within the mass spectrometer where there are two principal sources of this variation, the ion source and mass analyser (Glish and Vachet, 2003; Watson and Sparkman, 2007). The greatest source of mass spectrometry variants results from the ion source chosen with variants including chemical ionisation, electron ionisation, and fast atom bombardment, as well as the more commonly used electrospray ionisation (ESI), which has been further developed by adding a capillary electrophoresis component, and matrix assisted laser desorption/ionization (Glish and Vachet, 2003; Watson and Sparkman, 2007; Maxwell and Chen, 2008). Further MS method variation results from the mass analyser used such as the velocity dependent time-of-flight and ion motion dependent quadrupole systems, amongst others (Glish and Vachet, 2003). ESI-MS was the method implemented during this project owing to the softer ionisation protocol reducing the potentially extensive degradation caused to target carbohydrates, amongst other factors, with quadrupole time of flight mass analysis.

Mass spectra derived from these different experiments each present the mass-to-charge ratios of any target compounds and their breakdown products against ion abundance (Glish and Vachet, 2003). The mass spectra generated are complicated by the presence of multiple masses for each compound, resulting from different atom isotopes at a small number of locations within the target compound. Further to this, the mass displayed can differ from that expected for the ionised target compound owing the formation of adducts with differing counter cations, or anions depending on the mode of ionisation present. An example of this is the sodium adduct of the target compounds measured in this research, which was particularly prevalent, adding 23Da onto the expected masses of the compounds.

2.3 – Fourier Transform Infrared Spectroscopy:

As with both NMR and MS, Fourier transform infrared (FTIR) spectroscopy has been used for the characterisation of carbohydrate structures. One such method of carbohydrate analysis using FTIR has identified the mono- and disaccharides present in biological samples including honeys, syrups, and natural sweeteners whilst another has completed a similar analysis of the carbohydrates contained within fruit juices (Mellado-Mojica et al., 2016; Leopold et al., 2011). Similarly, FTIR has been used to identify carbohydrate polymers in terms of their level within samples resulting from extraction condition optimisation as well as during carbohydrate composition studies for different biological source materials (Baum et al., 2017; Pereira et al., 2009; Gómez-Ordóñez et al., 2011). The FTIR-based analyses of carbohydrate polymers have also been shown to be capable of identifying structural variations between different carbohydrate family members including sulfation level differences, notably for the glycosaminoglycan family, as well as between this family and other non-GAG family carbohydrates providing a method by which the purity of these samples can be analysed (Devlin et al., 2019a; Devlin et al., 2019b).

Further to these applications attenuated total reflectance FTIR (ATR-FTIR) spectroscopy has been applied to the analysis of cellulose crystallinity changes in a manner which would not be possible for most biophysical analysis techniques, owing to the polymer's crystalline nature preventing dissolution (Kljun et al., 2011). Alongside carbohydrate structural analyses, FTIR methods have been used for the analysis of carbohydrate interactions. One example of this is when studying those with proteins both in terms of confirming an interaction is present as well as studying the effect of interactions on α-helix and β-sheet levels within proteins during secondary structure rearrangement (Souillac et al., 2002). Interactions between carbohydrates have also been studied using FTIR spectroscopy, with suggestions being made regarding the interacting groups, due to differences between the spectra for the two carbohydrates alone and when mixed (Bianchera et al., 2014).

The FTIR spectroscopy method uses the splitting, and subsequent recombination, of an infrared light beam to generate an interference signal which can then be detected, transformed, and plotted (Gerwert and Kötting, 2010). This method examines vibrations, of a target compound that can be divided into two principal classes, stretching and bending, which equate to changes in the length and angle of atomic bonds respectively. These vibrations are measured principally at the level of individual, or a limited number of, bonds owing to the reduced mass of the atom involved. The nature of stretching vibrations as being dependent on the strength of bonds, results in higher frequencies for double and triple bonds compared to those of corresponding single bonds, i.e. C=O stretching vibration frequencies will be greater than those of C–O (Berthomieu and Hienerwadel, 2009). The FTIR method can be further developed by the inclusion of additional elements such as microimaging, macroimaging, attenuated total reflectance (ATR), transflection or transmission, given ATR-FTIR was used during the research carried out in this project the other additional elements will not be introduced in any greater detail. ATR-FTIR adapts the FTIR method to measure solid samples on a reflection element with mirrors incorporated into the spectrometer to direct infrared light to this element before being directed back towards the detector, resulting in a highly versatile method with sample preparation techniques often not required (Tiernan et al., 2020; Kazarian and Chan, 2013).

The FTIR spectra which result from these experiments compare the wavenumber of the stretching and bending vibrations against absorbance. Analysis of peak intensities present at wavenumbers corresponding to bending or stretching vibrations for the bonds of interest can be used to analyse the structure of a target compound. This method can therefore be used to directly analyse modifications which would otherwise not be possible via the alternate techniques thus far introduced. Sulfation provides a particularly relevant example of this, because NMR, for example, cannot detect the sulfate group directly, only infer its presence from changes in chemical shift position, however, it does provide greater information on the site of sulfation. Therefore, the combination of these techniques would prove particularly powerful analytically with FTIR able to provide a direct characterisation of the scale of sulfation while NMR can be used to identify the sites where this sulfation has occurred which, when quantified, could generate values for the proportion of each modification site.

2.4 – Chromatography:

Chromatographic techniques are routinely utilised for both the preparation and analysis of carbohydrates and proteins, with a wide range of separation methods viable for both types of molecule (Coskun, 2016). When applied to carbohydrates these techniques can include preparative size-dependent sample separation to generate homogeneous fractions for either subsequent structural analysis or inclusion in further characterisation assays. However, whilst carbohydrates are commonly separated using standard chromatography methods, high performance alternatives have also been applied. A wide range of these high performance chromatography methods are used for carbohydrate samples, with high performance anion-exchange chromatography (HPAEC), high performance size-exclusion chromatography (HPSEC), and hydrophobic interaction liquid chromatography (HILIC) being particularly prevalent (Cataldi et al., 2000; Simsek et al., 2013; Fu et al., 2013). Difficulties surrounding carbohydrate detection limits the potential detector variants which can be used, however a number of methods including pulsed amperometric detection (PAD), refractive index detection (RID), and evaporative light scattering detection (ELSD) have been shown to be viable (Mellado-Mojica et al., 2016; Dvořáčková et al., 2014).

By contrast, chromatography performed on protein samples is considerably simpler owing to a range of both preparative and analytical methods in common use with, in most cases, detection via the protein's innate ultraviolet absorbance. The most common chromatography method during recombinant protein purification is that of affinity chromatography where a range of tags have been developed to bind to their corresponding columns including, but not limited to, hexahistidine, maltose binding protein and glutathione-S-transferase tag sequences, with some affinity tags having additional benefits such as increasing protein stability (Costa et al., 2014; Young et al., 2012; Kimple et al., 2015). These affinity chromatography techniques can then be supplemented by ion-exchange, hydrophobic interaction, and/or size-exclusion chromatography methods used to separate proteins based on their surface charge, hydrophobicity and size respectively (Geng and Wang, 2008). In addition to the protein chromatography methods commonly used for the preparation of recombinant samples, coupling size-exclusion chromatography with multiangle laser light scattering (SEC-MALLS) generates an analytical protocol which can be used for the accurate determination of the molecular weight of target proteins (Thomsen, 2020; Some et al., 2019; Oliva et al., 2001).

2.5 – Differential Scanning Fluorimetry:

Differential scanning fluorimetry (DSF) experiments examine protein interactions with a binding partner monitoring the protein's stability, as assessed by melting temperature variations in a high throughput manner (Gao et al., 2020). Previous applications of DSF for protein interactions with proteins, carbohydrates and small molecules are widespread (Shao et al., 2020; Tobola et al., 2018; McClure et al., 2018; Støve et al., 2020). One of the common applications of differential scanning fluorimetry is the investigation of interactions between carbohydrates and proteins such as those of the coagulation cascade protein Antithrombin III with sulfated carbohydrates in a manner which allowed for the correlation of protein stability with resulting biological activity (Lima et al., 2013). Further DSF applications when studying carbohydrate protein interactions include comparing the effects of natural ligand competitors for binding galectin-3 (Sindrewicz et al., 2020) and determining the effect of specific sulfation pattern modifications during the interactions of glycosaminoglycan family members with fibroblast growth factors (Rudd et al., 2010). However, the principal advantage to differential scanning fluorimetry during the investigation of carbohydrate protein interactions is its ability to screen for interactions using a wide range of combinations as has been performed for glycosaminoglycans and their protein interaction partners (Uniewicz et al., 2010).

Differential scanning fluorimetry protocols measure protein unfolding with increasing temperature in the presence of a fluorescent dye using PCR equipment. The dye selected is used due to its fluorescent nature when present within a non-polar environment, such as that found within the hydrophobic core of proteins which becomes more available to penetration by the dyes during the initial phases of protein unfolding. Subsequent over exposure of the dye to the aqueous buffer conditions when protein unfolding is substantial reduces fluorescence back towards zero (Niesen et al., 2007; Walport et al., 2021). Alternatively, fluorescence can decrease post unfolding due to dissociation of the dye from the protein at high temperature, increased mobility of the dye, or increased mobility of the protein section bound to the dye as well as, as a consequence of the unfolded protein sequences aggregating. Melting temperatures are recorded at the mid-point of the protein unfolding transition where half of the individual instances of the measured protein are unfolded and the other half is intact (Gao et al., 2020; Byrne et al., 2018). Protocols using the addition of a fluorescent dye for detection function based on a number of key assumptions which should be met before analysis can be carried out. These include that; there is an absence, or sufficiently low presence, of hydrophobic patches on the protein surface that could affect baseline readings, the protein is stable at the start of the experiment, increasing temperature does not lead to protein structure rearrangement prior to unfolding, and there are no interactions between the protein and dye or chemical reactions between the protein, dye or other components in the sample mixture (Gao et al., 2020). Shifting of melting temperature to be either higher or lower can indicate the presence of interactions where the protein is stabilised or destabilised, respectively (Walport et al., 2021). The results of these experiments are presented as the recorded fluorescence at different melting temperatures, giving the pattern shown in Figure 19, where traces can be both normalised to the highest and lowest values surrounding the unfolding transition as well as abridged to remove readings beyond this period of transition.

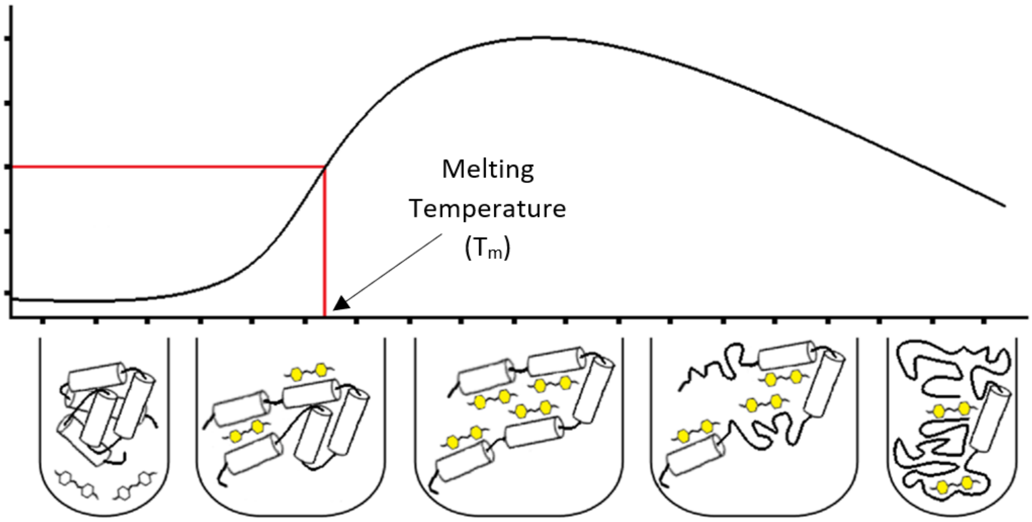


Figure 19: The interaction of SYPRO Orange Stain with a protein unfolding due to increased temperature during DSF measurement, and the consequent spectrum.

2.6 – Isothermal Titration Calorimetry:

Isothermal titration calorimetry (ITC) has developed wide ranging applications since it was first introduced in 1990 (Freire et al., 1990). Between 2011 and 2015, for example these interactions have be divided into a number of separate categories, including protein interactions with other proteins, small molecules, metals, nucleic acids, and other compounds, in total accounting for 67% of the total reactions run within this time frame. Other compounds studied via ITC include lipids, polysaccharides, synthetic compounds and nucleic acids in the absence of protein (Falconer, 2016). Whilst examples of protein-protein, protein nucleic acid, and protein-lipid interactions, alongside carbohydrate-small molecule, and enzyme kinetic studies have used ITC for the determination of thermodynamic data these will not be introduced any further (Pierce et al., 1999; Liang, 2008; Velazquez-Compoy et al., 2004; Situ et al., 2014; Zsila et al., 2018; Olsen, 2006). Multiple carbohydrate-protein interactions have been studied including the multivalent interactions of lectins and their binding ligands which each generated viable results in spite of the problems meeting assumptions required for a binding model owing to the potential for mixed binding modes (Wittmann, 2013; Dam and Brewer, 2002). Further to this the CH–π stacking interactions of both wild type and modified *Ralstonia solanacearum* lectin with α-methyl L-fucose have been examined using ITC leading to the conclusion that CH–π stacking is a driving force in protein carbohydrate interactions (Houser et al., 2020).

ITC measures the evolution of heat resulting from a protein-ligand interaction when compared to the temperature of a reference cell in arrangements similar to that shown in Figure 20. Experiments are performed as titration reactions with a number of injections, typically 20 to 40, of the ligand at higher concentration into a lower concentration protein sample or *vice versa* with each injection resulting in a heat change within the sample cell. This heat change is then accommodated by the application of varying current levels in order to maintain the sample and reference cells at a constant temperature difference. Quantitative analysis of the data recorded via integration, normalised to concentration, enables the temperature after each injection to be calculated and plotted. Such analyses of data produced by ITC can indicate the interaction stoichiometry, n, equilibrium dissociation constant, K_D, heat Capacity, ΔC, Gibbs free energy, ΔG, enthalpy, ΔH, and entropy, ΔS, values (Doyle, 1997; Leavitt and Freire, 2001; Walport et al., 2021). Molecular mechanism characterisation for any interactions observed can be achieved from the reporting of the parameters listed with the forces that drive the formation of interactions a focus of these characterisations (Doyle, 1997).

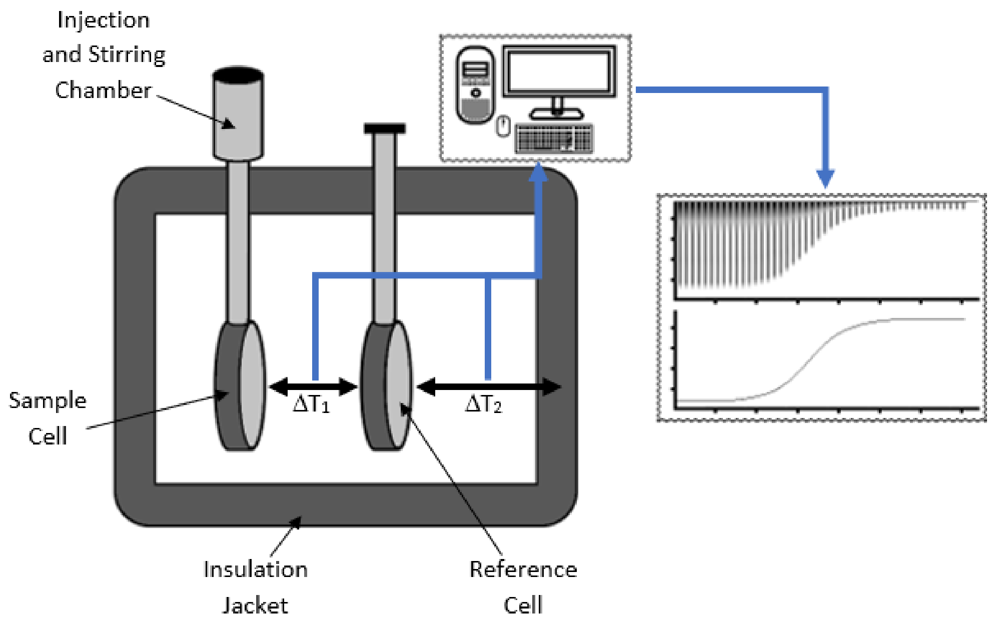


Figure 20: The configuration of a ITC calorimeter alongside the mechanism by which sample measurement generates resulting data.

Chapter 3 – Experimental Procedures

3.1 – Polysaccharide Preparation and Modification:

3.1.1 – Xyloglucan Hydrolysis:

Xyloglucan samples (3mg, 5mg/mL), dissolved in 100mM sodium acetate buffer pH5.5, were hydrolysed using 0.1U/mL xyloglucan-specific-endo-β-1,4-glucanase (xyloglucanase) at 37°C overnight. Samples were subsequently spin fractionated according to the protocol in section 3.1.2 and the <3kDa fraction dialysed against H_2O with repeat exchanges of the H_2O at regular intervals for 48 hours using 500Da molecular weight cut off cellulosic dialysis tubing . The dialysed less than 3kDa fraction was subsequently dried on a rotary evaporator, to reduce sample volume, before being flash frozen and freeze-dried.

3.1.2 – Spin Fractionation:

Carbohydrate samples were spin fractionated across 3kDa and 10kDa molecular weight cut off 15mL Amicon® ultra centrifugal filters at 3,000xg for 60 minutes. Samples were then washed by the addition of H_2O up to 15mL and centrifugation as above a further three times. Following the wash protocol, samples were flash frozen and freeze-dried.

3.1.3 – Size-Exclusion Chromatography:

Size-exclusion chromatography of the hydrolysed carbohydrate samples was performed using a Sepharose CL-6B column under gravity with 20% (v/v) ethanol in water as the mobile phase and 5mL fractions collected. These fractions were spotted onto a series of glass plates and dried at 37°C for one hour. Fractions showing sample present as a white spot were flash frozen and added to a lyophiliser overnight or until dried to dryness.

Subsequent size-exclusion chromatography of the hydrolysed and spin fractionated carbohydrate samples was performed using a Superdex 30 16/600 column, again in 20% (v/v) ethanol in water, on an ÄKTA purifier system with 1mL fractions collected. As previously, the fractions were spotted onto glass plates and dried with those showing a sample present being flash frozen and freeze-dried.

3.1.4 – Oligosaccharide Acetylation for Mass Spectrometry:

Samples to be analysed via mass spectrometry were acetylated by dissolving the carbohydrate in 50μL pyridine and 50μL anhydrous acetic anhydride and incubated at 50°C overnight in a fume hood. Acetylated samples were subsequently opened and returned to incubation at 50°C, again in a fume hood, to allow the pyridine and remaining anhydrous acetic anhydride to boil off. The samples were subsequently dialysed against water for 48 hours with regular water changes and dried to dryness. The dried, post-dialysis, samples were dissolved in deuterated methanol and characterised by nuclear magnetic resonance spectroscopy and mass spectrometry.

3.1.5 – Carrageenan Sample Preparation:

Carrageenan samples (3mg, 5mg/mL), were dialysed against H_2O prior to use in sulfation and desulfation reactions using 3.5kDa molecular weight cut off cellulosic tubing to remove the degraded disaccharide fragments present in the starting material samples. Repeated H_2O exchanges were performed at regular intervals over 48 hours. Dialysed samples were then freeze-dried.

3.1.6 – Sulfation of Carbohydrates in 1-Butyl-3-Methylimidazolium Chloride:

Carbohydrate samples (0.1mmoles) were dissolved in 1-butyl-3-methylimidazolium chloride (5.4g, density = 1.081g/mL) and supplemented, post optimisation, with both 0.1mmoles of 4-dimethylaminopyridine and 5mmoles of pyridine sulfur trioxide complex. These samples were placed in a heat block at 85^{o}C with stirring, and samples were collected at 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, and 24 hours. A control time point was also collected at 24 hours which lacked the pyridine sulfur trioxide complex. Prior optimisation of these conditions included varying concentrations for both 4-dimethylaminopyridine, between 0.01mmoles and 1mmole, and pyridine sulfur trioxide complex, between 0.05mmoles and 5mmoles.

Subsequent extraction was performed by precipitating in 50mL cold ethanol, saturated with sodium acetate, with later filtration across a Buchner funnel and dissolution in H_2O. The extracted samples were then spin fractionated using 3kDa molecular weight cut off 15mL Amicon® Ultra Centrifugal Filters at 3,000xg for 60 minutes. A wash protocol was performed with an initial centrifugation period at 3,000xg for 60 minutes to reduce the sample volume followed by overnight incubation in saturated sodium chloride with a 60 minute centrifugation period at 3,000xg. Finally, there were three washes in 15mL H_2O again at 3,000xg for 60 minutes. Once washed the greater than 3kDa sample was flash frozen and added to a lyophiliser overnight or until dried to dryness.

3.1.7 – Sulfation of Heparin in Dimethylformamide:

Heparin samples modified via dimethylformamide based sulfation were first dialysed against RO water, using 3kDa molecular weight cut off dialysis tubing, to remove impurities with repeat H_2O exchanges at regular intervals for 48 hours before being freeze dried overnight on a lyophiliser. The samples were then converted to their tetrabutylammonium salt by passing them across a column loaded with Dowex Marathon C beads in their tetrabutylammonium form using milli-Q water as a mobile phase and subsequently dialysed, as previously, and freeze-dried overnight on a lyophiliser.

Once dry, 0.1mmoles of the tetrabutylammonium heparin salt was dissolved in 5mL of dimethylformamide and supplemented with 0.5mmoles pyridine sulfur trioxide complex before being added to a heat block with stirring for up to 24 hours. Samples were collected after 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, and 24 hours, with a control, also collected at 24 hours, from which the pyridine sulfur trioxide complex was absent. Three reaction temperatures were used, 35°C, 55°C, and 85°C.

Sulfated heparin samples were extracted by precipitation in 50mL of cold ethanol, saturated with sodium acetate, with subsequent filtration across a Buchner funnel and dissolution in H_2O. The extracted samples were then spin fractionated using 3kDa molecular weight cut off 15mL Amicon® Ultra Centrifugal Filters at 3,000xg for 60 minutes. A wash protocol was performed with an initial centrifugation period at 3,000xg for 60 minutes to reduce the sample volume followed by three washes in 15mL H_2O again at 3,000xg for 60 minutes. Once washed the greater than 3kDa sample was flash frozen and added to a lyophiliser overnight or until dried to dryness.

3.2 – Protein Production and Purification

3.2.1 – Competent Cell Preparation

Competent cells were grown as for the protein expression, below, with some differences. Briefly, 50µL of competent BL21 STAR™ (DE3) *Escherichia coli* cells were streaked onto a LB agar plate without antibiotics and incubated at 37°C overnight. A single colony of the resulting cells was then removed and added to a 5mL starter culture before being incubated at 37°C and 240rpm overnight. This starter culture was then added to a 250mL LB broth expression medium and incubated at 37°C and 180rpm before the cells were harvested via centrifugation at 4°C and 4,200xg for ten minutes.

The pelleted cells were gently resuspended in pre-cooled TFB1 buffer (100mM rubidium chloride, 30mM potassium acetate, 10mM calcium chloride, 50mM manganese chloride, 15% (v/v) glycerol, pH5.8), on ice, and subsequently incubated on ice for five minutes. The resuspended cells were again centrifuged at 4°C and 4,200xg for ten minutes before the resulting pellet was resuspended, in TFB2 buffer (10mM 3-(N-morpholino)propane sulfonic acid, 10mM rubidium chloride, 75mM calcium chloride, 15% (v/v) glycerol, pH 6.5), and incubated on ice for between fifteen minutes and one hour. Following this incubation, the resulting competent cells were separated into 50mL aliquots, flash frozen and stored at -80°C until required for use.

3.2.2 – Plasmid Constructs

The plasmid constructs used for protein expression and purification were provided by Prof. Dave Fernig, for fibroblast growth factors 1 and 2, and Prof. Lu-Yun Lian, for cyclophilins B and D. They were subsequently sequenced by Eurofins Genomics using standard T7 Primers prior to being used for protein overexpression, with the principal elements highlighted in Table 2.

Table 2: Protein, affinity tag, and affinity tag cleavage sequences for the plasmid constructs received

Target Protein	Vector	Antibiotic Resistance	Affinity Tag	Cleavage Sequence	Tag Location
Fibroblast Growth Factor 1	pETM-11	Kanamycin	His_6	TEV Protease	N-terminus
Fibroblast Growth Factor 2	pET-14b	Ampicillin	N/A	N/A	N/A
Cyclophilin B	pETM-11	Kanamycin	His_6	TEV Protease	N-terminus
Cyclophilin D	pETM-11	Kanamycin	His_6	TEV Protease	N-terminus

The target protein sequences contained within the constructs provided were applied to the ExPASy ProtParam function (Gasteiger et al. in Walker (ed.), 2005) to calculate parameters required during overexpression. Selected results from this analysis are shown in Table 3 for the affinity-tag cleaved target protein.

Table 3: ExPASy ProtParam data calculated for the target proteins used during Differential Scanning Fluorimetry interactions screening experiments (post tag cleavage)

Target Protein	Amino Acids	Molecular Weight (g/mol)	Computed pI	Extinction Coefficient (Cysteines)	Extinction Coefficient (Reduced)
Fibroblast Growth Factor 1	143	16,047	7.88	17,545	17,420
Fibroblast Growth Factor 2	155	17,254	9.58	16,180	15,930
Cyclophilin B	184	20,346	9.25	16,960	16,960
Cyclophilin D	166	17,831	9.07	10,220	9,970

3.2.3 – Plasmid Miniprep

A single colony was harvested from an LB agar plate containing the transformed cells of interest and used to inoculate a 10mL LB broth culture supplemented with either 58mM kanamycin sulfate or 269mM ampicillin sodium salt as appropriate and incubated at 37°C with shaking at 240rpm overnight. The resulting plasmids were purified according to the manufacturer's instructions for a low copy plasmid in the Bioline Isolate II Plasmid Mini Kit used.

3.2.4 – Plasmid Concentration Determination

Plasmid concentration was determined by measurement on a Thermo Scientific Nanodrop 2000c Spectrophotometer in Nucleic Acid mode.

3.2.5 – Bacterial Cell Transformation

50μL competent BL21 STAR™ (DE3) cell aliquots were supplemented with 1μL of the plasmid construct for expression. The resulting cell and plasmid mixtures were incubated, on ice, for thirty minutes before being heated at 42°C for thirty seconds. Following a further two minute incubation period, again on ice, the transformed cells were further supplemented with 450μL SOC broth (0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10mM sodium chloride, 2.5mM potassium chloride, 10mM magnesium chloride, 10mM magnesium sulfate, and 20mM glucose) and incubated at 37°C with shaking at 240rpm for sixty minutes.

100μL of the transformed cells were added onto LB agar plates containing 58mM kanamycin sulfate, or 269mM ampicillin as appropriate, and the remaining 400μL pelleted via centrifugation at 1,650xg for five minutes. 300μL of the supernatant was removed, and the pellet resuspended in the remaining 100μL before being added to the LB agar plates, which were subsequently incubated at 37°C overnight in a static incubator.

3.2.6 – Protein Expression

A single colony was harvested from the target protein's LB agar plate, containing the transformed cells of interest, and used to inoculate a 20mL of LB broth starter culture, supplemented with 58mM kanamycin sulfate, or 269mM ampicillin sodium salt as appropriate. The resulting samples were then incubated at 37°C with shaking at 240rpm overnight.

After incubation, the 20mL inoculated LB broth was added to 1L of the same LB expression medium supplemented with either 58mM kanamycin sulfate or 269mM ampicillin sodium salt and incubated at 37°C with shaking at 180rpm until the absorbance at OD_{600} reached around 0.45. Once this value had been reached the expression media were cooled to 18°C whilst maintaining shaking at 180rpm and once the OD_{600} reached 0.6 a final 500μM concentration of IPTG was added and the expression media left to incubate at 18°C with shaking at 180rpm overnight.

3.2.7 – Bacterial Cell Harvest

E.coli cells used for protein expression were harvested from their LB expression media via centrifugation at 9,100xg for five minutes at 4°C. The supernatant produced was discarded and the pellet resuspended in either a **Low Salt** (20mM phosphate, pH6.8, 150mM sodium chloride, 1mM dithiothreitol) or **Low Imidazole** (20mM phosphate, pH6.8, 500mM sodium chloride, 10mM imidazole 1mM dithiothreitol) buffer, for Heparin HiTrap and His-Trap purifications, respectively, and stored at -20°C until required for purification.

3.2.8 – Bacterial Cell Lysis

Resuspended cell pellets were thawed on ice with the addition of a final concentration of 1.25mM calcium chloride, 25mg/mL deoxyribonuclease I, 5mg/mL lysozyme, and 3mM magnesium chloride as well as 100µL protease inhibitor cocktail, EDTA free (composed of 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), aprotinin, bestatin, E-64, leupeptin, and pepstatin A). The pellets were then lysed by passing the samples across a Constant Systems cell disruptor at 28,000 Psi. Lysed cells were centrifuged at 38,800xg for thirty minutes with resulting the supernatant filtered through a 0.44mm syringe filter and stored on ice until required for affinity chromatography. The pellet formed by centrifugation was resuspended in 8M urea and stored on ice until required for SDS-PAGE.

3.2.9 – Heparin HiTrap Affinity Chromatography

A 5mL HiFliQ heparin FPLC affinity column was equilibrated using sequential washes in filtered and degassed milli-Q water, **Low Salt Buffer**, **High Salt Buffer** (20mM phosphate, pH6.8, 2M sodium chloride, 1mM dithiothreitol), and a repeat **Low Salt Buffer** on an ÄKTA Start system using a 5mL/min flow rate. The supernatants produced as described above were then loaded onto the ÄKTA Start at the same 5mL/min flow rate used for column equilibration with a flow through fraction collected for use in SDS-PAGE.

The subsequent elution protocol comprised a fifteen column volume wash in **Low Salt Buffer** followed by an increasing gradient of **High Salt Buffer** between 0% and 100% over ten column volumes and included the collection of 4mL fractions during the sodium chloride elution gradient. Analysis of the chromatogram generated, collected at an absorbance of 280nm, enabled the determination of protein containing fractions, which were collected and stored on ice for subsequent purification and analysis via SDS-PAGE.

3.2.10 – HisTrap Affinity Chromatography

A 5mL HisTrap™ FF Crude FPLC affinity column was equilibrated using sequential washes in filtered and degassed milli-Q water, **Low Imidazole Buffer**, **High Imidazole Buffer** (20mM phosphate, pH6.8, 500mM sodium chloride, 500mM imidazole, 1mM dithiothreitol), and a repeat **Low Imidazole Buffer** on an ÄKTA Start system using a 5mL/min flow rate. The supernatants produced as described above were then loaded onto the ÄKTA Start system at the same 5mL/min flow rate used for column equilibration with a flow through fraction collected for use in SDS-PAGE.

Similar to heparin affinity chromatography, the elution protocol comprised a fifteen column volume wash in **Low Imidazole Buffer** followed by an increasing gradient of **High Imidazole Buffer** between 0% and 100% over ten column volumes and included the collection of 4mL fractions during the imidazole elution gradient. Analysis of the resulting chromatogram, collected at 280nm, indicated the protein containing fractions, which were collected and stored for subsequent analysis via SDS-PAGE.

3.2.11 – Tobacco Etch Virus Protease Affinity Tag Cleavage

Proteins purified using a hexahistidine tag had this affinity sequence cleaved, using Tobacco Etch Virus (TEV) Protease, prior to further purification methods. Briefly, the target protein containing fractions were combined and added to a section of 3.5kDa dialysis tubing to dialyse against a 50mM tris, pH8.0, 150mM sodium chloride, 1mM dithiothreitol, 0.5mM ethylenediaminetetraacetic acid buffer overnight at 4°C, having first been supplemented with a 250μL 1mg/mL aliquot of TEV protease in 50mM tris, pH8.0, 200mM sodium chloride, 1mM β-mercaptoethanol buffer.

The cleaved hexahistidine tags were subsequently separated from the previously-tagged target proteins using HisTrap affinity chromatography where the protein containing flow through samples were stored until required for later use.

3.2.12 – PD10 Column Buffer Exchange

Protein containing samples were buffer exchanged between affinity and size exclusion chromatography purification steps using a PD10 column. Briefly, the column was equilibrated using five column volumes of the buffer into which the protein was to be exchanged, after which 2mL of the protein was added, with the resulting eluate discarded. 500μL of the buffer into which the protein was to be exchanged was then added and again the eluate was discarded. A subsequent 3mL of this buffer was then added to the column in order to elute the protein with the eluate stored until required for later use.

3.2.13 – Size-Exclusion Chromatography

Following the affinity chromatography methods, target protein containing eluted fractions were further purified via size-exclusion chromatography across a Superdex 75 HiLoad 26/600 column. This was equilibrated using sequential washes in filtered and degassed milli-Q water, and a 50mM tris, pH6.8, 100mM sodium chloride, 1mM dithiothretol buffer using a 1.60mL/min flow rate on an ÄKTA Pure system. The affinity chromatography fractions shown via UV trace and SDS-PAGE analysis to contain the protein of interest were combined, concentrated and buffer exchanged before being loaded onto the SEC column.

The elution protocol performed in 50mM tris, pH6.8, 100mM sodium chloride, 1mM dithiothreitol buffer used one column volume for protein elution during which time forty-nine 4mL fractions were collected, subsequently a second column volume of buffer was run through the column as a wash. Protein containing samples, identified via analysis of the UV trace at 280nm, were stored on ice for subsequent analysis via SDS-PAGE.

In cases where protein yield post-affinity chromatography was insufficient to purify using the Superdex 75 HiLoad 26/600 column, purification was performed using a Superdex 75 10/300 column instead. The elution protocol remained similar with two column volumes in 50mM tris, pH6.8, 100mM sodium chloride, 1mM dithiothretol buffer used, however, the flow rate was reduced to 0.75mL/min and twenty-nine 1mL samples were collected in place of the larger fractions collected normally.

3.2.14 – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Samples collected during protein purification were prepared for SDS-PAGE by mixing 4:1 with 5x loading buffer (10% (w/v) sodium dodecyl sulfate, 10mM dithiothreitol, 20% (v/v) glycerol, 200mM tris pH6.8, 0.05% (w/v) bromophenol Blue, and 4M urea) and heated to 95^{o}C for five minutes. 7μL of each sample, alongside one 10μL molecular weight marker sample per gel, were loaded onto the SDS-PAGE gels, comprised of a 15% resolving component (375mM tris, pH8.8, 15% (w/v) acrylamide, 0.1% (w/v) sodium dodecyl sulfate, 0.1% (w/v) ammonium persulfate, and 10mL N,N,N',N'-tetramethylethane-1,2-diamine (TEMED)) and a 4% stacking component (125mM tris, pH6.8, 4% (w/v) acrylamide, 0.1% (w/v) sodium dodecyl sulfate, 0.1% (w/v) ammonium persulfate, and 10mL N,N,N',N'-tetramethylethane-1,2-diamine (TEMED)). Gels were run at voltages generating a current below 30mA per gel in running buffer (25mM tris, 200mM glycine, 0.1% (w/v) sodium dodecyl sulfate), typically runs were started at 100V until the samples had collected at the bottom of the stacking gel at which point the voltage was raised to 200V, when running two gels, with the total duration being around 60 minutes.

3.2.15 – Coomassie Staining

SDS-PAGE gels were stained using coomassie blue. Each gel was initially heated in a 50% (v/v) ethanol, 10% (v/v) acetic acid solution using microwave irradiation for thirty seconds before being left on a rocker for a further 5 minutes. The gels were subsequently switched to a 5% (v/v) ethanol, 7.5% (v/v) acetic acid solution supplemented with 400µL 95% (v/v) ethanol, 0.25% (w/v) brilliant blue G stain and returned to the rocker to stain overnight.

3.2.16 – Protein Concentration via Centrifugation

Combined protein samples, from both affinity and size-exclusion chromatography purifications, were concentrated using 3kDa molecular weight cut off 15mL Amicon® ultra centrifugal filter units. These concentrator tubes were centrifuged at 3,000xg until such a point as the volume had been adequately reduced. Both the >3kDa and <3kDa fractions were measured on a Thermo Scientific Nanodrop 2000c Spectrophotometer in Protein A280 mode to determine any little loss of protein.

3.2.17 – Protein Concentration Determination

Protein concentration was determined by measurement of the absorbance at OD_{280} on a Thermo Scientific Nanodrop 2000c Spectrophotometer in Protein A280 mode and Beer-Lambert Law calculations using the extinction coefficients derived from the ExPASy ProtParam function (Gasteiger et al. in Walker (ed.), 2005).

3.3 – Analytical Techniques:

3.3.1 – Nuclear Magnetic Resonance:

Carbohydrate samples prepared, in 600µL D_2O unless otherwise stated, for NMR spectroscopy were supplemented with 100µM sodium trimethylpropanesulfonate (DSS) and inserted into 5mm glass NMR tubes. Protein samples dissolved in 50mM tris, pH6.8, 100mM sodium chloride, 1mM dithiothreitol buffer were supplemented with 30µL D_2O before being inserted into the 5mm glass NMR tubes. Spectra were collected on either a 600MHz Bruker Avance II+ spectrometer or a 800MHz Bruker Avance Neo spectrometer, both fitted with TCI CryoProbes using various experiments depending on the purpose, Table 4. Experiments were performed at 300K unless otherwise stated. Spectra were processed and analysed using Bruker TopSpin 3.4 and Bruker TopSpin 4.0 software.

Table 4: Acquisition parameters for the spectra recorded for the different stages of research including the type of spectrum, and specific pulse program, recorded as well as the number of points and spectral width alongside the resulting resolution for each of the different elements of the project investigated:

Sample	Experiment	Pulse Program	Points (#)		Scans (#)	Spectral Width (ppm)		FID Resolution (Hz)		Spectro-meter Frequency (MHz)
			F2	F1		F2	F1	F2	F1	
Polysaccharides	1D ^{1}H	zgpr[a]	---	65536	32	---	10	---	0.18	600
	1D ^{1}H	zgpr[b]	---	9588	96	---	12	---	2.00	800
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2[b]	400	1400	64	90	10	90.6	11.4	800
XGase Hydrolysis	1D ^{1}H	zgpr	---	65536	32	---	10	---	0.18	600
Post Sepharose CL-6B Size Exclusion Chromatography	1D ^{1}H	zgpr	---	12288	64	---	12	---	1.17	600
	2D ^{1}H,^{1}H-COSY	cosydfphpr	512	2048	16	10	10	23.5	5.86	600
	2D ^{1}H,^{1}H-COSY	cosydfphpr	1024	4096	16	10	10	11.7	2.93	600
	2D ^{1}H,^{1}H-TOCSY	cltocsypr[c d]	264	768	32	4	4	18.2	6.25	600
	2D ^{1}H,^{1}H-TOCSY	cltocsypr[c d]	1024	4096	32	4	4	4.69	1.17	600
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2	400	2048	8	90	10	67.9	5.86	600
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2	1024	4096	8	90	10	26.5	2.93	600
	2D ^{1}H,^{13}C-HSQC	hsqcedetgpsp.3	400	2048	8	90	10	67.9	5.86	600
	2D ^{1}H,^{13}C-HSQC	hsqcedetgpsp.3	1024	4096	8	90	10	26.5	2.93	600
	2D ^{1}H,^{13}C-HMBC	hmbcgplpndqf	262	868	160	80	4	46.1	5.53	600
Post-Spin Fractionation	1D ^{1}H	zgpr	---	65536	64	---	10	---	0.26	800
	2D ^{1}H,^{1}H-COSY	cosydfphpr	800	4096	8	10	10	20	4.00	800
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2	400	1024	16	70	5	70.4	7.75	800
	2D ^{1}H,^{13}C-HMBC	hmbcgpndqf	400	2048	16	220	10	110.7	8.00	800
High Resolution Oligosaccharide Structural Characterisation	1D ^{1}H	zgpr	---	65536	64	---	10	---	0.26	800
	2D ^{1}H,^{1}H-COSY	cosydfphpr	2048	4096	8	8	8	6.25	3.05	800
	2D ^{1}H,^{1}H-TOCSY	dipsi2gpphpr[e]	2048	4096	8	8	8	6.25	3.05	800
	2D ^{1}H,^{1}H-NOESY	noesygpphpr[f]	1024	4096	8	4	4	6.25	1.60	800
	2D ^{1}H,^{1}H-ROESY	roesyphpr.2	1024	4096	8	8	8	12.5	3.05	800
	2D ^{1}H,^{1}H-ROESY	roesyphpr.2[g]	1024	4096	8	4	4	6.25	1.60	800
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2[h i j]	8192	4096	8	50	8	2.46	3.05	800
	2D ^{1}H,^{13}C-HMBC	hmbcetgpl3nd[i j]	6144	4096	8	50	8	3.28	3.05	800
	2D ^{1}H,^{13}C-HSQC-TOCSY	hsqcdietgpsisp[i k]	8192	5120	8	50	9	2.46	2.68	800
	1D ^{13}C-DEPT	deptsp135	---	65536	2048	---	80	---	0.49	800
	1D ^{13}C	zgpg30	---	32768	2048	---	80	---	0.98	800

	1D ^{13}C	zgpg30	---	65536	2048	---	80	---	0.49	800
Post-Superdex 30 Size Exclusion Chromatography	1D ^{1}H	zgpr	---	65536	32	---	11	---	0.2	600
	1D ^{1}H	zgpr	---	65536	64	---	11	---	0.2	600
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2[i]	3072	3072	8	52	6	5.11	2.34	600
Acetylation	1D ^{1}H	zgpr[l]	---	16384	256	---	10	---	1.0	600
Different Supplier Samples	1D ^{1}H	zgpr	---	65536	64	---	11	---	0.2	600
	2D ^{1}H,^{1}H-COSY	cosydfphpr	2048	4096	8	8	8	4.68	2.34	600
	2D ^{1}H,^{1}H-TOCSY	cltocsypr[c d]	2048	4096	8	10	10	5.86	2.93	600
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2[i j]	3072	3072	8	52	6	5.11	2.34	600
	2D ^{1}H,^{13}C-HMBC	hmbcetgpl2nd.2[ij]	6144	4096	24	70	8	3.44	2.34	600
	2D ^{1}H,^{13}C-HSQC-TOCSY	hsqcdietgpsisp[m]	6144	4096	8	70	9	3.44	2.52	600
	1D ^{13}C	zgpg30	---	32768	6144	---	60	---	0.55	600
Sulfation and Desulfation	1D ^{1}H	zgpr	---	66560	64	---	10	---	0.18	600
	2D ^{1}H,^{13}C-HSQC	hsqcetgpsisp2	512	2048	96	100	10	58.9	5.87	600
Protein Expression	1D ^{1}H	zgesgp	---	8192	1024	---	14	---	2.06	600

[a] These spectra were run at a range of temperatures between 283K and 333K

[b] These spectra were run at 333K

[c] The pulse program for this experiment is taken from Griesinger et al., 1988.

[d] The mixing time (d9) used for this experiment was 120μs.

[e] The mixing time (d9) used for this experiment was 240μs.

[f] The mixing times (d8) used for these experiments were 1s and 2s.

[g] The mixing times (p15) used for these experiments were 0.4s, 0.8s and 1.6s.

[h] The relaxation delays (d1) used for these experiments were 2s and 5s.

[i] Non Uniform Sampling was used for these experiments at 25%.

[j] These spectra used adiabatic decoupling.

[k] The mixing times (d9) used for these experiments were 30μs, 60μs, 120μs, and 240μs.

[l] The spectra were run in deuterated methanol, CD_3OD.

[m] The mixing time (d9) used for this experiment was 60μs.

3.3.2 – Mass Spectrometry:

Acetylated samples, prepared in methanol, were measured by the Department of Chemistry's Analytical Services Team using electrospray ionisation mass spectrometry with positive ion polarity on an Agilent 6540 UHD Accurate Mass Q-TOF LC/MS mass spectrometer.

3.3.3 – Fourier Transform Infrared Spectroscopy:

Polysaccharide samples were measured, as dry powder, by Dr. Cassio Lima using attenuated total reflectance Fourier transform infrared spectrscopy on a Bruker Invenio FT-IR Spectrometer.

3.3.4 – Differential Scanning Fluorimetry:

Differential scanning fluorimetry assays were performed, in triplicate, on a StepOne Plus Real-Time PCR machine using SYPRO Orange dye and increasing temperature between 25^{o}C and 95^{o}C in increments of 0.3^{o}C. 5μM protein, either fibroblast growth factor 1 or 2, or cyclophilin B or D, were measured, both in the presence of 0.5mg/mL carbohydrate and when these were absent, in 50mM tris, pH 6.8, 100mM NaCl, 1mM DTT buffer. Data were analysed using GraphPad Prism 9 with normalisation to the maximum and minimum absorbance measurements in each dataset followed by fitting to a Boltzmann sigmoidal equation. Melting temperatures were then calculated at the midpoint of transition generated by protein unfolding. Data points prior to the minimum and subsequent to the maximum were disregarded, thus leaving solely the sigmoidal transition during protein denaturation.

3.3.5 – Isothermal Titration Calorimetry:

Isothermal titration calorimetry assays were performed on a Malvern MicroCal PEAQ-ITC Automated calorimeter at 25^{o}C using a 19 injection method with a subsequent system wash. 400μL of either 40μM FGF-1 or 60μM CypB were used as the cell sample in 50mM tris, pH 6.8, 100mM NaCl, 1mM TCEP buffer. 120μL of the carbohydrate at the required concentration in the same buffer was used as the syringe sample. Data analysis was performed using MicroCal PEAQ-ITC Analysis Software.

Chapter 4 – Carbohydrate Structural Analysis Nuclear Magnetic Resonance Method Development

4.1 – Introduction

The analysis of carbohydrate samples by NMR has significant challenges to be overcome in order to be successful, owing to the extensively overlapped signals for different instances of the same residue type within a polymer. This is balanced by the distinct pattern of signals for different but structurally similar residue types and the clear separation of anomeric peaks from those for other CH pairs within the spin system. This leads to difficulty in characterising the structures of similar sequences when they extend beyond oligosaccharide lengths. NMR applied to carbohydrate structure characterisation typically starts with 1D, mostly natural abundance, ^{1}H and ^{13}C experiments which, when supplemented by 2D ^{1}H,^{1}H-COSY, 2D ^{1}H,^{1}H-TOCSY, and 2D ^{1}H,^{1}H-NOESY as well as 2D ^{1}H,^{13}C-HSQC, 2D ^{1}H,^{13}C-HSQC-TOCSY and 2D ^{1}H,^{13}C-HMBC experiments, proves particularly powerful analytically, provided signals do not extensively overlap (Tuomivaara et al., 2015; Komatsu and Kikuchi, 2013; Ding et al., 2016). Collection of these spectra allows for the assignment of any residues present alongside any quantitative analyses required to determine the composition of individual carbohydrate structures or manifold carbohydrate mixtures (Ohmenhaeuser et al., 2013; Sassaki et al., 2014; Palhares et al., 2019). They also allow for modification site identification and reaction quantification when atom-level resolution is achieved.

The diverse properties of carbohydrates has led to wide ranging applications within both commercial and biopharmaceutical products, amongst others, and they remain a key focus of research. Application of high resolution NMR to study the carbohydrate structures present in these products, particularly where a relationship between structure and consequent properties is known, is therefore highly important. This is particularly true of biopharmaceutical carbohydrates, such as the anticoagulant heparin, where the target polysaccharide requires comprehensive analysis to ensure the structure present will exhibit the desired properties without adverse side effects (Onishi et al., 2016; Yates and Rudd, 2016). However, equally important for biopharmaceutical and commercial products is the analysis of contaminants present, which can lead to inconsistent properties within the resulting sample and, in biopharmaceuticals, consequent biological activities. Failure to identify biologically active contaminants within biopharmaceuticals can prove to have major consequences for product safety, as illustrated by the contamination of heparin samples with oversulfated chondroitin sulfate leading to a crisis between 2007 and 2008 (Liu et al., 2009; Beni et al., 2011).

Xyloglucan, a test-case polysaccharide of industrial relevance used here for NMR method development, which acts as a thickening agent for commercial products, differs structurally in a source species dependent manner with nineteen sidechain variants identified (Tuomivaara et al., 2015). The *Tamarindus indica* sourced samples analysed are composed of four repeating unit variants differing solely with regards to the number and location of D-galactose residues (Hsieh and Harris, 2009), Figure 21. Using spectra recorded with resolution at, or approaching, the natural line width, the proportion of each repeating unit present, a key factor in carbohydrate composition and modification studies, becomes identifiable. Whilst complete characterisation of native carbohydrate structures is important when determining the relationship with consequent properties, it also becomes vital for the characterisation of both the sites and extent of modification, should they be required. The aim of the research presented in this chapter is to describe the development of an NMR based method for carbohydrate structural characterisation, with spectra collected at atom-level resolution, which could be applied to examining carbohydrate modification samples at a later point. Further to this, this chapter will aim to collect and analyse samples separated in a size-dependent manner, to enable verification of any assignments made, as well as a series of samples from different suppliers, to identify their similarities and differences.

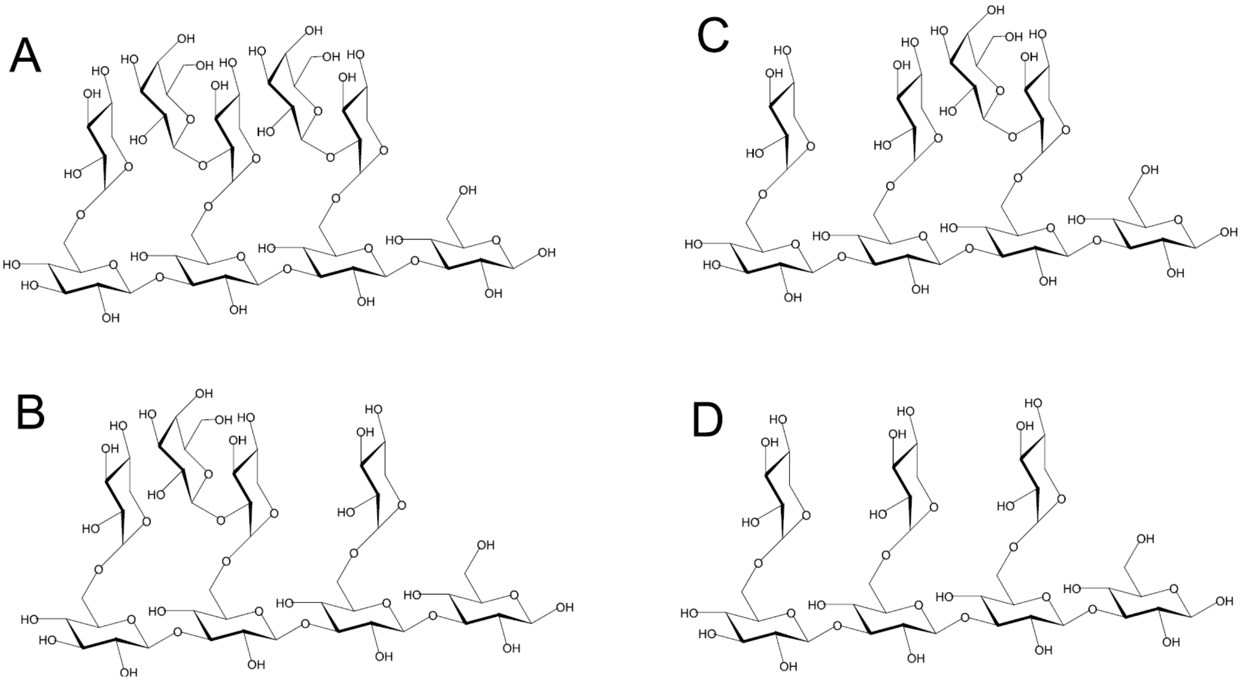


Figure 21: Haworth projections showing the XLLG (A), XLXG (B), XXLG (C) and XXXG (D) xyloglucan repeating units.

4.2 - Results

4.2.1 – High Resolution NMR Method Development for Polysaccharide Sample Analysis

Any newly derived NMR method for carbohydrate structural analysis should place considerable importance on the resolution, where possible, of distinct signals for individual C-H groups present as an overarching goal. In the case of polysaccharide samples this has some additional limitations derived from their molecular weight and viscosity, for example, as well as the field strength of the spectrometer used. This resolution at, or approaching, the natural line width was, however, not achieved during initial experiments for the gel forming xyloglucan test case polysaccharide therefore the effect that the sample's highly viscous nature has on tumbling rate, relaxation time, and consequently line-shape and resolution was targeted for improvement. Natural line width is defined, for the purposes of this research, as the point at which further acquisition or processing parameter optimisation fails to improve the resulting spectrum. This was first attempted by varying the sample concentration, between 0.5mg/mL and 20mg/mL, to determine whether it would alter sample viscosity such that resolution would increase without an excessive decrease in signal intensity. The viscosity of the highest concentration samples necessitated a pre-treatment at 100^{o}C to enable insertion into NMR tubes. Sample concentration variation showed no effect on the line width of peaks contained within the spectra, Figure 22, only resulting in an expected variation in signal intensity. Therefore a 5mg/mL concentration was chosen for samples used during subsequent optimisation steps such that signal intensity is sufficient for detection using a range of one and 2D experiments without the requirement for pre-treatment during preparation.

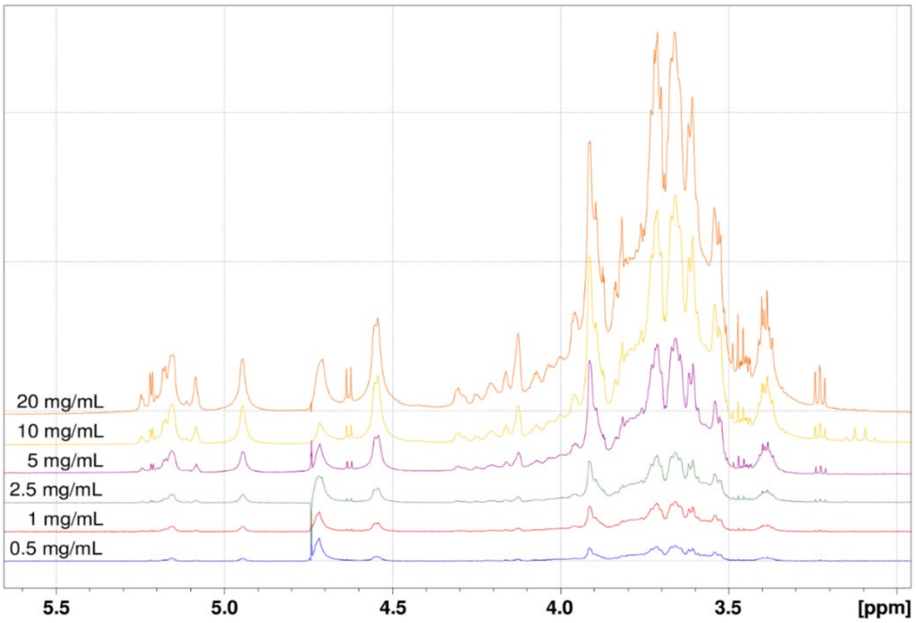


Figure 22: The carbohydrate region of 1D ^{1}H NMR spectra collected on polysaccharide xyloglucan samples at the concentrations indicated.

Having abandoned sample concentration as a means to increase resolution by reducing viscosity, spectrum collection temperature was next targeted for optimisation. Increasing sample temperature to 100°C during pre-treatment steps reduced viscosity, therefore a temperature gradient, of 5°C steps between 10°C and 60°C, was investigated to partially mimic this effect. 1D ^{1}H spectra, Figure 23, measured for each temperature point exhibited peak intensity increases with increasing temperature that occurred concurrently with improved resolution, most noticeably for the peak at 4.55ppm in Figure 23, denoted by the black arrow, representing the D-glucose and D-galactose residues' anomeric CH pair. This peak increases from a low intensity singlet at 10°C to being partially split into a doublet at 60°C having roughly tripled in intensity. Whilst varying the spectrum collection temperature resulted in marginally increased resolution at 60°C, that observed remained far below the resolution at, or approaching, the natural line width which would constitute the optimised measurement conditions, therefore further optimisation using additional methods to reduce sample viscosity, by reducing the molecular weight, were required.

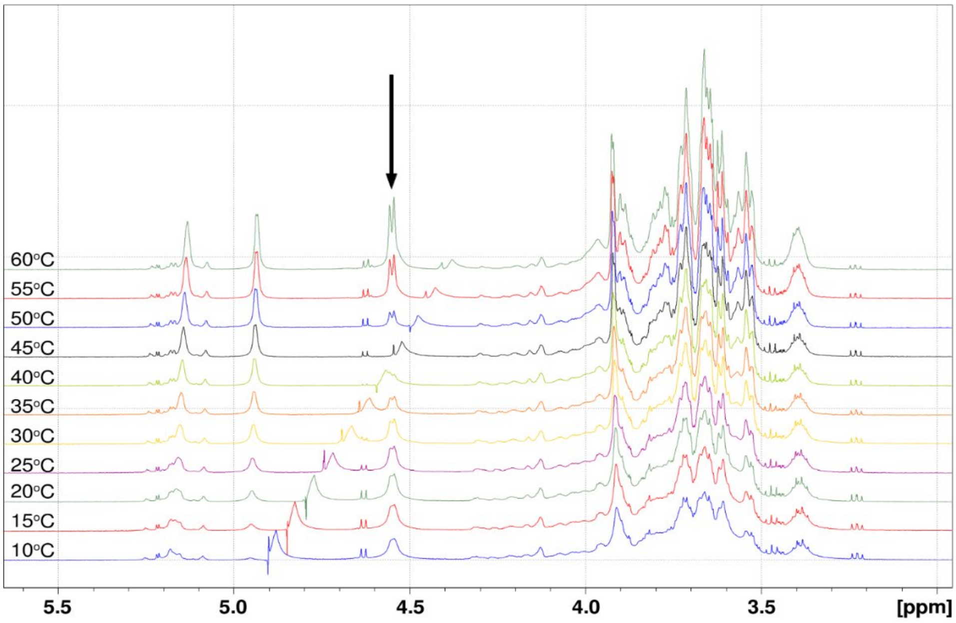


Figure 23: The carbohydrate region of 1D ^{1}H spectra collected on a 5mg/mL xyloglucan sample at the temperatures indicated.

Whilst falling below the resolution at, or approaching, the natural line width required, the conditions determined thus far, using a 5mg/mL sample for measurement at 60°C, did allow for the collection of 2D ^{1}H,^{13}C-HSQC spectra from which a series of assignments could be made, Figure 24A. These assignments identified the backbone D-glucose, terminal and further linked sidechain D-xylose, and terminal sidechain D-galactose residues expected for a *Tamarindus indica* sample from previously published assignments (Tuomivaara et al., 2015; Ding et al., 2016). However, a series of L-arabinose peaks visible within the sample could not be accounted for by the repeating unit structures expected from the proton and carbon assignments defined during previous characterisations (Tuomivaara et al., 2015; Ding et al., 2016). Integration of the anomeric ^{1}H peaks, Figure 24B, normalised so the combined terminal and further linked D-xylose integrals represented three residues, Figure 24C, was used to attempt an estimation of the proportion of each residue type present.

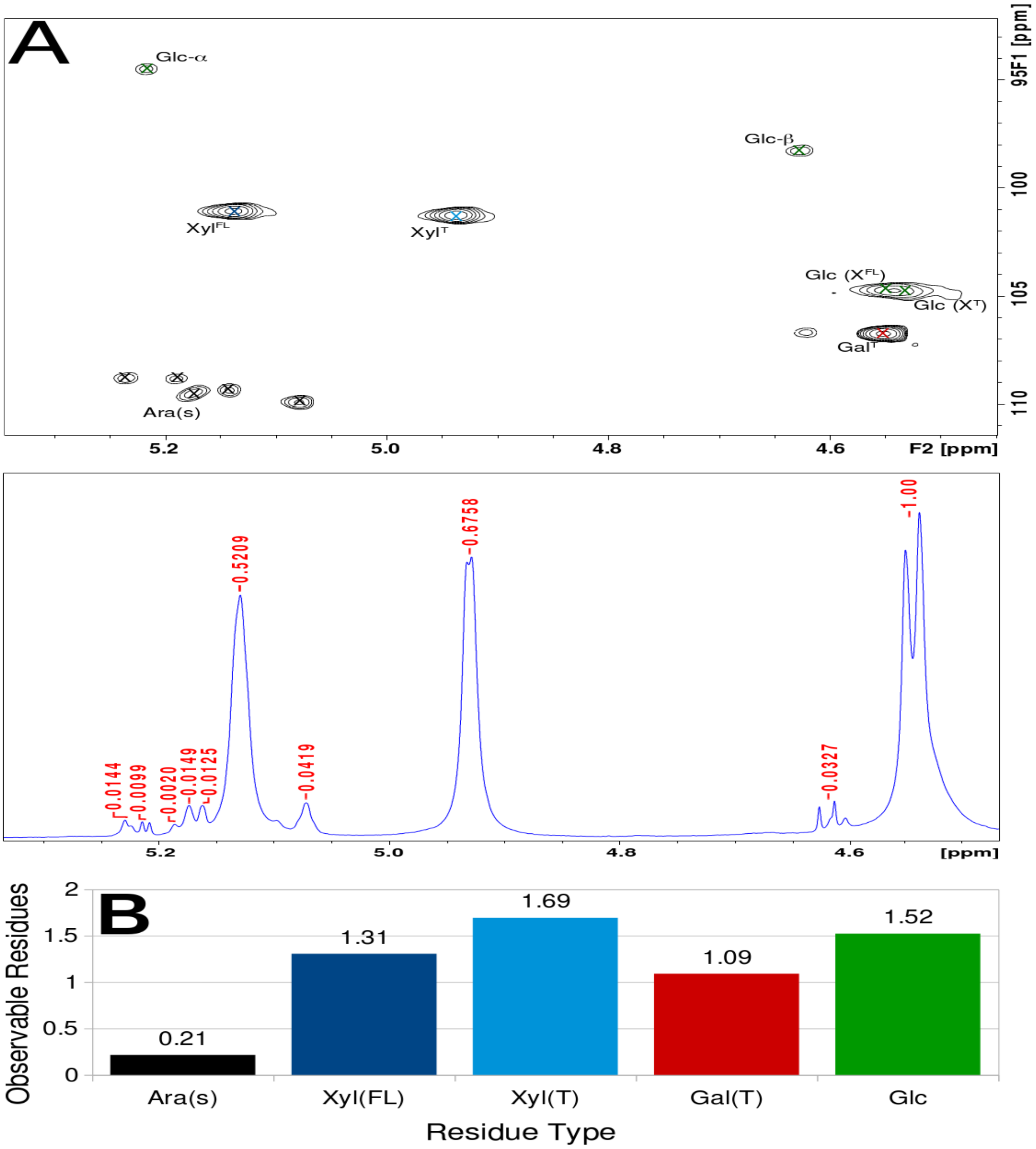


Figure 24: The anomeric region of the A) 2D ^{1}H,^{13}C-HSQC, upper, and 1D ^{1}H, lower, spectra measured for polysaccharide xyloglucan showing the assignments and integrals normalised to their being a combined 3 residues for terminal and further linked D-xylose as suggested by previously published research, respectively. Here Glc, Xyl, and Gal represent D-glucose, D-xylose, and D-galactose respectively while FL indicates the residue is further linked and T indicates that it is terminal within a side chain. Glucose residue followed by X^{FL} or X^{T} indicate that the xylose they are bound to is further linked or terminal within a side chain respectively. B) shows the calculated number of each residue type within the average repeating unit (Tuomivaara et al., 2015; Hsieh and Harris, 2009).

As both D-galactose and L-arabinose have previously been identified in xyloglucan side chains, analysis of the integrals collected, Figure 24A, supposed that terminal L-arabinose and D-galactose are the only residue types bound to further linked D-xylose (Tuomivaara et al., 2015). Therefore, of the 1.3 further linked D-xylose residues identified, 0.2 are linked to a terminal L-arabinose residue, and the remaining 1.1 residues to a terminal D-galactose, Figure 24B. As the peaks for D-galactose and D-glucose were considerably overlapped, removing the D-galactose component (1.1 residues) gave a value of 1.5 identifiable D-glucose residues. There is therefore an underestimation of the number of backbone D-glucose residues, as the 1.5 residues identified cannot support the 3 D-xylose residues per repeating unit from the previously identified *Tamarindus indica* xyloglucan structures (Tuomivaara et al., 2015). This underestimation of the number of backbone D-glucoses present is believed to result from the particularly fast relaxation of the less flexible backbone residues. This is coupled with the possibility that the extremely broad nature of the backbone D-glucose signals makes them undetectable both visually in the spectrum and, consequently, during integration. The need for increased accuracy during integral measurement, as well as being able to assign signals to specific atoms within distinct repeating units during modification, further highlights the importance of increasing spectrum resolution far beyond that achieved thus far.

Spectrum resolution increases for a 5mg/mL test case xyloglucan sample have been observed with increasing measurement temperature, however they do not resolve signals of different repeating units and do not give accurate integrals required for a useful carbohydrate structure characterisation method. Given that viable methods for resolution improvement on a viscous polysaccharide sample have been exhausted, further optimisation via polymer hydrolysis to oligosaccharide fragments was conducted and coupled with size dependent separation to remove any uncleaved polysaccharide.

4.2.2 - High Resolution NMR Method Development for Oligosaccharide Samples

Enzymatic hydrolysis of the xyloglucan samples could be used to produce repeating units small enough to enable the collection of high-resolution NMR spectra. Use of an enzyme specific to xyloglucan would also enable the determination of whether the L-arabinose observed was present in this polymer or a contaminant. Therefore the test-case sample was hydrolysed using xyloglucan-specific *endo*-β-D-1,4-glucanase, hereafter xyloglucanase, across a range of concentrations between 0.1U/mL and 1U/mL. Time courses for these samples, including that for 0.1U/mL enzyme shown in Figure 25, indicated that xyloglucan hydrolysis had run to completion within 30 minutes with dramatically increased resolution, exhibited as decreased peak line width, and no further changes after this point. The increase in spectrum resolution observed occurred in parallel with a complete change to the viscosity of the sample, which became close to that of water. The peaks in Figure 25, including two appearing at 5.21 and 4.65 ppm for the α and β reducing end D-glucose anomers, have the expected doublet splitting pattern, with pairs of doublets observable for the terminal and further linked D-xylose residues, present at 4.93 and 5.15 ppm, respectively. Peaks for the terminal D-galactose and backbone D-glucose residues, present between 4.48 and 4.58ppm, were also far better resolved, with multiple doublets emerging, although they are considerably overlapped. Alongside sample rheology changes, increased resolution coincided with increased signal intensity, not shown in Figure 25, both of which make high resolution 2D experiments more feasible. Reduction of the sample viscosity post-hydrolysis to being as water removed the need for either high temperature pre-treatment or data collection at 60°C, further simplifying the conditions required for the collection of spectra recorded at the natural line width on target carbohydrates.

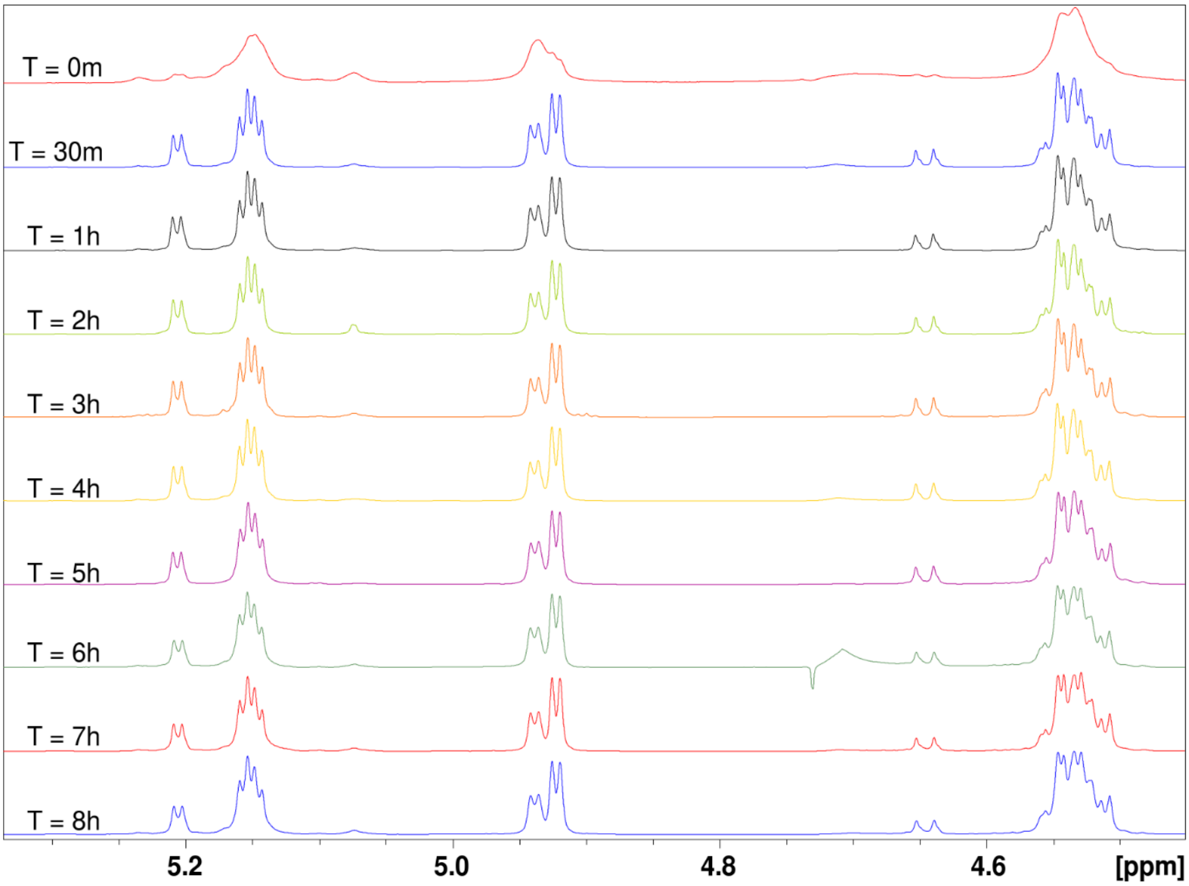


Figure 25: A time course of 5mg/mL xyloglucan treated with xyloglucanase prior to enzyme addition, indicated as T = 0m, as well as over the course of 8 hours with samples collected at the times indicated. Spectra are shown at the same heights to emphasise changes in resolution.

Assignment of the spectra for xyloglucan oligosaccharide samples, coupled with the suggestion that hydrolysis had run to completion, enabled the conclusion that the previously reported *Tamarindus indica* repeating units had been produced (Tuomivaara et al., 2015).

Electrospray ionisation mass spectrometry performed on the acetylated xyloglucan oligosaccharides sample, Figure 26, indicated that the repeating units, or one of their adducts, had been produced. However, mass spectrometry also identified a series of lower molecular weight structures of which one corresponded to a fragment potentially derived from one of the repeating units which was undetectable in the 1D ^{1}H and 2D ^{1}H,^{13}C-HSQC NMR spectra collected thus far.

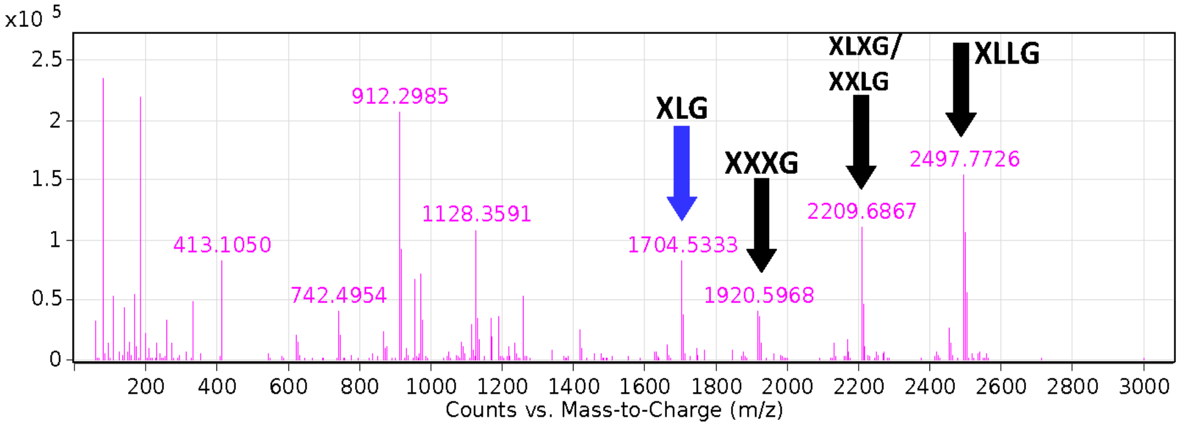


Figure 26: Mass Spectrum for the hydrolysed and acetylated xyloglucan test case sample with the molecular weights for the nonasaccharide repeating unit A (XLLG), octasaccharide units B and C (XLXG and XXLG) and heptasaccharide unit D (XXXG) indicated by the black arrows, while the hexasaccharide repeating unit fragment XLG is indicated by the blue arrow.

Increases to spectrum resolution generated during xyloglucanase hydrolysis, compared with that seen for spectra measured on the polysaccharide, suggested that the L-arabinose residues were part of a contaminant compound. The L-arabinose peaks did not increase in intensity with hydrolysis, as would have been expected in line with the other residues if L-arabinose were a constituent part of the xyloglucan polymer. While this determination is only indicative based on sample hydrolysis, with L-arabinose peaks remaining identifiable despite being present at a low level, separation in a size-dependent manner would support the mass spectrometry results and serve to conclusively demonstrate whether they form part of a contaminant polymer. This separation would also serve to make subsequent assignment of xyloglucan spectra less problematic provided the L-arabinose residues are not a constituent of the xyloglucan.

Separation of the xyloglucan oligosaccharides on a Sepharose CL-6B size exclusion chromatography column generated multiple fractions which were combined post-1D ^{1}H spectrum analysis into four combined fractions, Figure 27. 1D ^{1}H NMR analysis of the combined separation fractions indicated the L-arabinose residues were confined to the first sample. This, coupled with the broad line width and low intensity is indicative of a contaminant polysaccharide not present as part of the target xyloglucan chain, hence it was not characterised further. Similarly, analysis of the 1D ^{1}H spectra collected on the fourth combined fraction was indicative of an oligosaccharide mixture expected from xyloglucanase hydrolysis. The absence of any peaks for L-arabinose indicates that the separation of the two carbohydrates, post-hydrolysis, was complete. Size-exclusion chromatography of this nature was incapable of separating the repeating units contained within the hydrolysed xyloglucan sample therefore a less time intensive method for the separation of xyloglucan from the L-arabinose based polymer was sought.

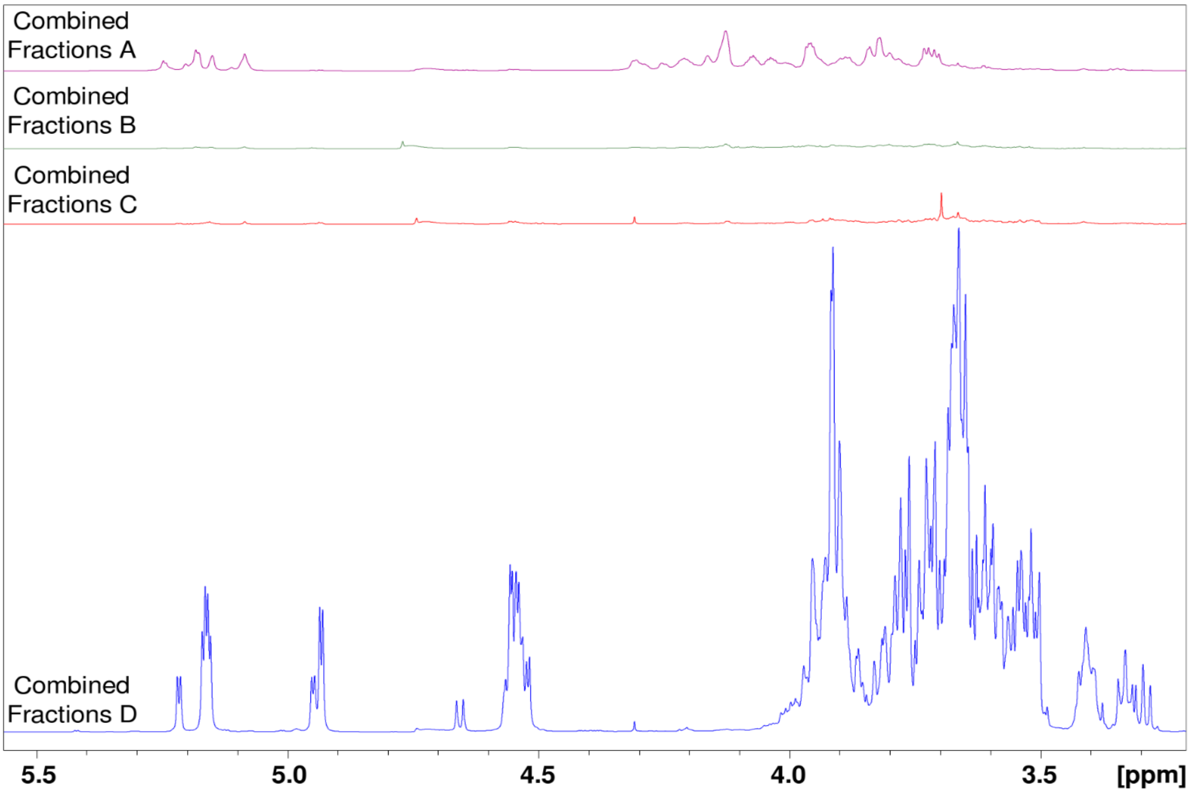


Figure 27: 1D ^{1}H NMR spectra collected on the combined fractions samples eluted from a Sepharose CL-6B size exclusion chromatography column, where combined fractions eluted from the column close to its dead volume are designated 'Combined Fractions A' and subsequent fractions combined due to having identical 1D ^{1}H NMR traces, are designated Combined Fractions B to D.

This, less time intensive, method was separation of the hydrolysed sample between 3kDa and 10kDa molecular weight cut off membranes using centrifugal filters for spin fractionation which would capture any contaminant polysaccharides but allow the repeating units to pass through. This proved equally effective as the above size-exclusion chromatography for the separation of the contaminant polymer from the xyloglucan oligosaccharides, as can been by comparison of Figure 27 and Figure 28, with a reduction in the time required from two days to four hours, thus making it a more feasible step in a carbohydrate analysis methodology. Using these two cut off sizes generated three fractions >10kDa, 3kDa to 10kDa, and <3kDa, Figure 28, with the L-arabinose containing polysaccharide confined to the >10kDa fraction whilst the xyloglucan was confined to the <3kDa fraction. As with size-exclusion, it was not possible to separate the repeating units which, whilst useful for an NMR structure characterisation method, is not essential at the present point, however, 2D spectra based assignment of the target compound may become less efficient without it.

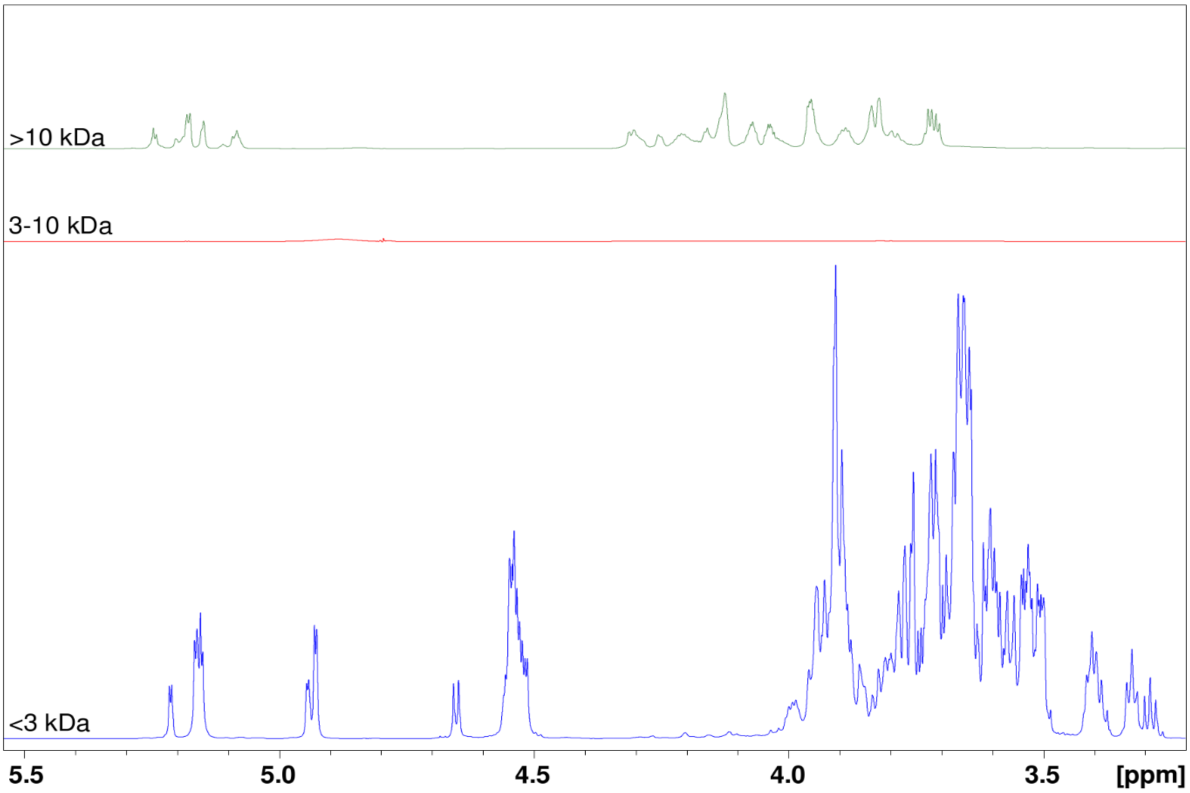


Figure 28: 1D ^{1}H NMR spectra measured on the spin fractionation samples collected when using 3kDa and 10kDa MWCO concentrator tubes to separate the hydrolysed sample into >10kDa, 3 to 10kDa, and <3kDa fractions.

The increase in signal intensity and resolution observed for 1D ^{1}H NMR spectra was also evident in the 2D $^{1}H,^{13}C$-HSQC spectra collected on the hydrolysed sample, Figure 29B, compared to the intact polysaccharide, Figure 29A. The singlet broad peaks seen in Figure 29A, collected at 60°C, have been replaced by peaks with clearly identifiable splitting patterns, Figure 29B. This figure focusses solely on the peaks for the D-galactose (Gal^{T}) residues at 4.55/107.5ppm and backbone D-glucose ($Glc(X^{T})$ and $Glc(X^{FL})$) residues at 4.54/105.5ppm. Improved resolution for the D-galactose peak, now present as a doublet, still only suggested one variant present whereas the two D-glucose residue types previously observed increased to at least four variants (subsequently assigned to $Glc(X^{T})_{(p2\&3)}$, $Glc(X^{FL})_{(p2\&3)}$, $Glc_{(p1)}(X^{T}{}_{(p2)})$, and $Glc_{(p1)}(X^{T}{}_{(p2})$, see later) based on the largely overlapping doublet peaks generated.

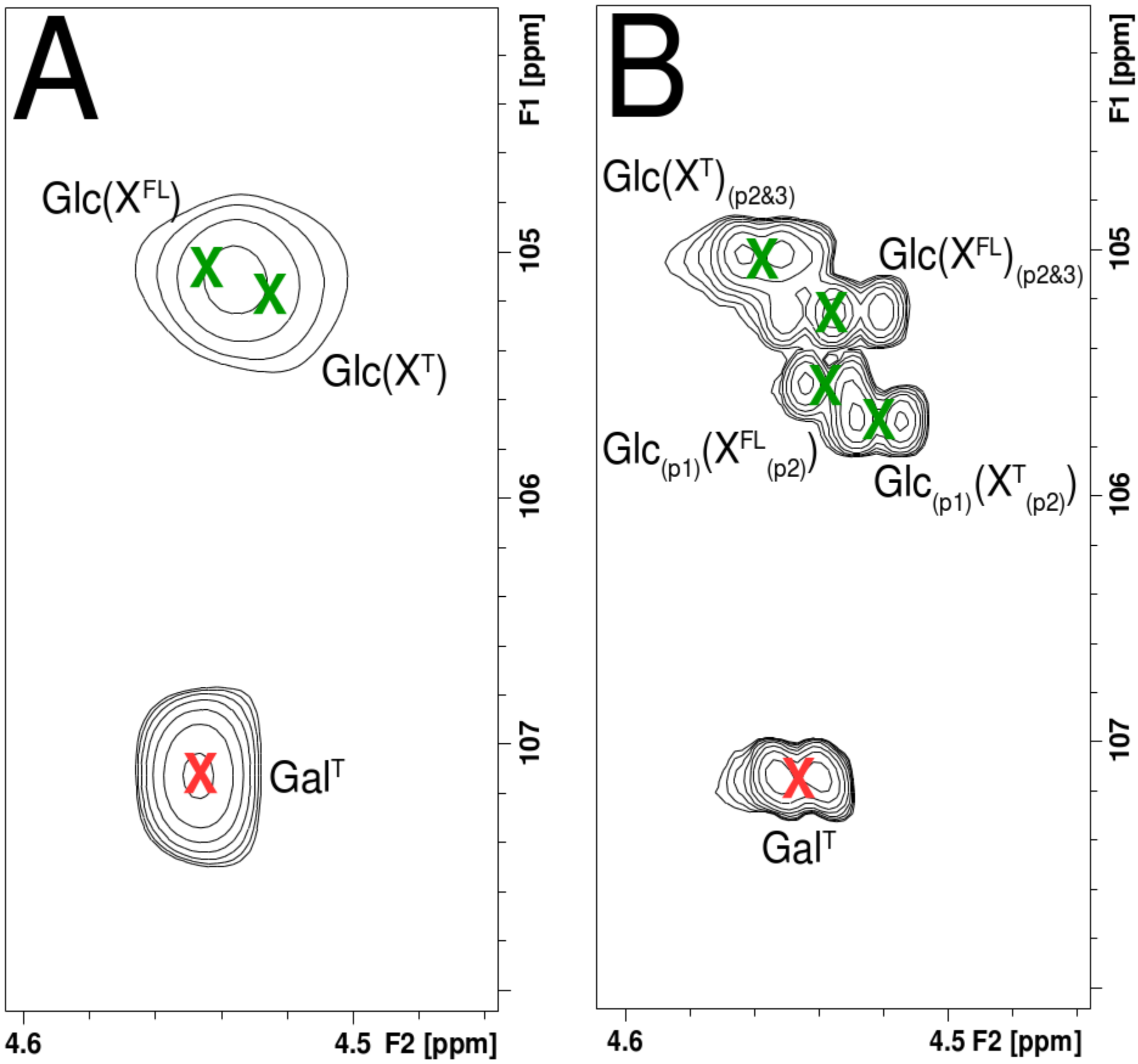


Figure 29: The 2D ^{1}H,^{13}C-HSQC spectral region for terminal D-galactose and non-reducing end and central (within the repeating units) D-glucose residues for A) the polymeric xyloglucan sample and B) the post-hydrolysis and separation oligosaccharide repeating units sample.

The improvements made by the hydrolysis and separation of the xyloglucan test-case samples assist with achieving the requisite resolution, which is a key factor for NMR-based structure characterisation method development for carbohydrates, their modifications and study of protein interactions. Whilst this was not possible for polysaccharide spectra, significant progress towards it has been made via hydrolysis to oligosaccharides and separation of the target compound from any contaminants. As the desired atom-level resolution remains to be achieved, further optimisation will be performed focussing on the NMR experimental parameters for the 2D spectra. These will need to be tailored to generate the best possible resolution whilst avoiding extended acquisition times in experiments where decoupling is being utilised, which could adversely affect the spectrometer and probe, particularly given the size of the dataset to be collected. Whilst not completely achieving the desired results, the improvements made have answered one of the principal questions posed thus far, i.e. the L-arabinose residues have been identified as part of a contaminant within the sample, rather than being an integral part of the xyloglucan itself.

4.2.3 – Parameter Optimisation for High-Resolution Spectrum Improvement

Having reduced the molecular weight and, therefore, the linewidth of the signals, further improvement was explored to enhance the resolution of the 2D spectra through the optimisation of parameters. This focussed on those parameters used for both spectrum acquisition and subsequent processing by reducing sweep width to the minimum possible, switching to adiabatic decoupling, increasing the number of points measured, using non-uniform sampling and optimising the processing parameters used.

The principal change made to the acquisition parameters was the use of adiabatic decoupling which, by reducing the pulse power directed to the probe, allows for the longer acquisition times. In the indirect dimension the acquisition time can be simply increased by increasing the number of increments, however this makes the experiment unacceptably long. This was mitigated by the implementation of non-uniform sampling, with 25% of indirect dimension points collected, reducing the measurement time considerably. To take full advantage of the high-resolution measurements, the spectra were processed with twice the number of acquired points in each dimension and using the sine bell squared window function with phase shift parameters 3.5 and 2.7 for the direct and indirect dimensions respectively. Further to this processing of non-uniform sampling spectra was achieved using compressed sensing (CS) rather than the standard multidimensional decomposition (mdd).

The spectra collected following computational parameter optimisation had dramatically increased resolution, Figure 30, such that two D-galactose residue variants can be identified, Figure 30B, in place of the one previously suggested, Figure 30A. The appearance of a second D-galactose doublet in the anomeric region could be hypothesised from the potential presence of D-galactose residues on two of the three side chains within each *T.indica* xyloglucan repeating unit, hence its appearance is not unexpected. The assignments made for these two peaks correspond to terminal D-galactose residues in a side chain neighbouring the D-glucose residue present at the reducing end, Figure 30E 9, or non-reducing end, 10, respectively.

Similarly, the number of peaks for the anomeric protons of D-glucose residues not present at the reducing end increases from at least four doublets to potentially twelve, depending on the assignments made. The peaks previously identified as being the non-reducing end backbone D-glucoses with a D-galactose either present, Figure 30A iv, or absent, iii, from the neighbouring side chain correspond to the same two peaks, 8 and 7, seen in the increased resolution spectrum, Figure 30D. The additional peaks seen therefore all represent D-glucose residues present at neither the reducing nor the non-reducing ends, Figure 30C. The prior assignments made, for D-glucose residues whose side chains do, Figure 30A ii, and do not, i, contain a D-galactose residue, no longer hold true with updated assignments suggesting that these represent backbone D-glucoses adjacent to the non-reducing and reducing end residues respectively.

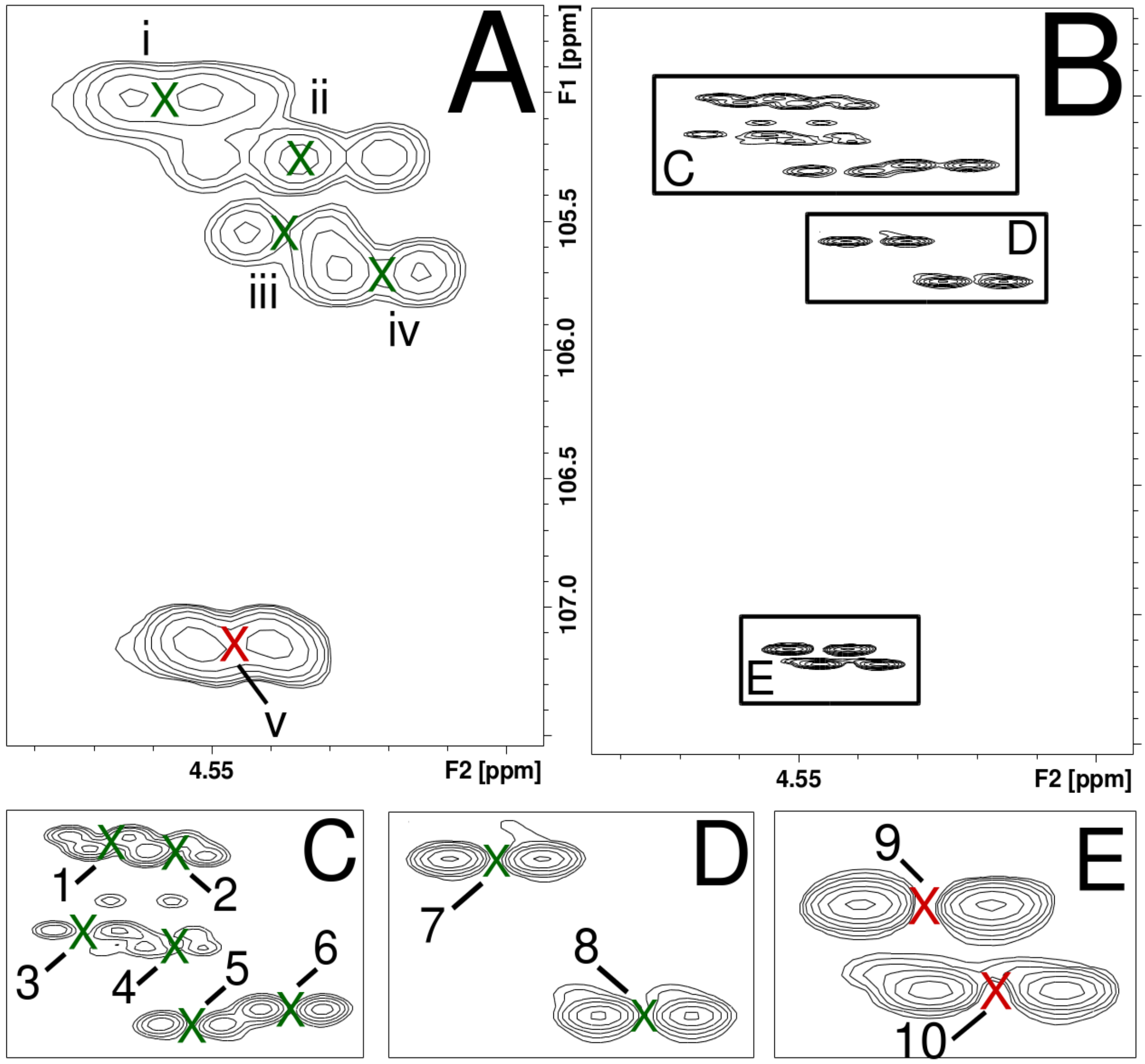


Figure 30: 2D ^{1}H,^{13}C-HSQC spectral region for the terminal D-galactose and non-reducing end and internal D-glucose residues for A) the hydrolysed and separated oligosaccharides and B) the improved spectrum generated by optimisation of both acquisition and processing parameters. C), D), and E) correspond to the inserts shown in B) for central D-glucoses, non-reducing end D-glucoses and terminal D-galactoses respectively.

The split between the presence or absence of D-galactose is then accounted for within these sets of peaks with Figure 30C 6 and 5 representing the side chains branching off D-glucose residues next to the non-reducing end which either contain, or lack, a D-galactose residue, respectively. Similarly, D-glucose residues next to the reducing end whose side chain either contains or lacks a D-galactose residue are represented by Figure 30C peaks 1 and 2 respectively. As will be detailed later in this results chapter, the atom and residue positions corresponding to these peaks were identified via the use of principally HSQC, HSQC-TOCSY and, importantly for residue position within the repeating unit, HMBC, NOESY and ROESY spectra. Alongside the peaks from the repeating units seen in Figure 30C there are those, peaks 3 and 4, generated by breakdown product D-glucose residues. The updated assignments between Figure 30A and Figure 30C-E indicate the atom-level resolution targeted for carbohydrate structural characterisation has been achieved and the method can thus be applied.

A further observation made from these computationally enhanced spectra is that the 2D ^1H,^{13}C-HSQCs are approaching the natural line width seen by 1D ^1H NMR. Figure 31 highlights four peaks from the 2D ^1H,^{13}C-HSQC anomeric region with red and green lines indicating corresponding ^1H peaks. Those residues present as a β-anomer, $^3J_{H1H2}$ = 8Hz to 10Hz, Figure 31A and B, are close to the natural linewidth with side peaks, represented by the green lines in Figure 31B, distinctly visible if considerably overlapped with the major peaks, red lines. These side peaks can only be identified in some of the panels representing α-anomers, $^3J_{H1H2}$ = 3Hz to 4Hz, Figure 31C and D. One example is in panel D, above the peak corresponding to the red lines, and another is the side peak seen in the panel C 1D ^1H spectrum, at 5.209ppm, which is indistinguishable from the main peak in the 2D ^1H,^{13}C-HSQC measured and present in that panel.

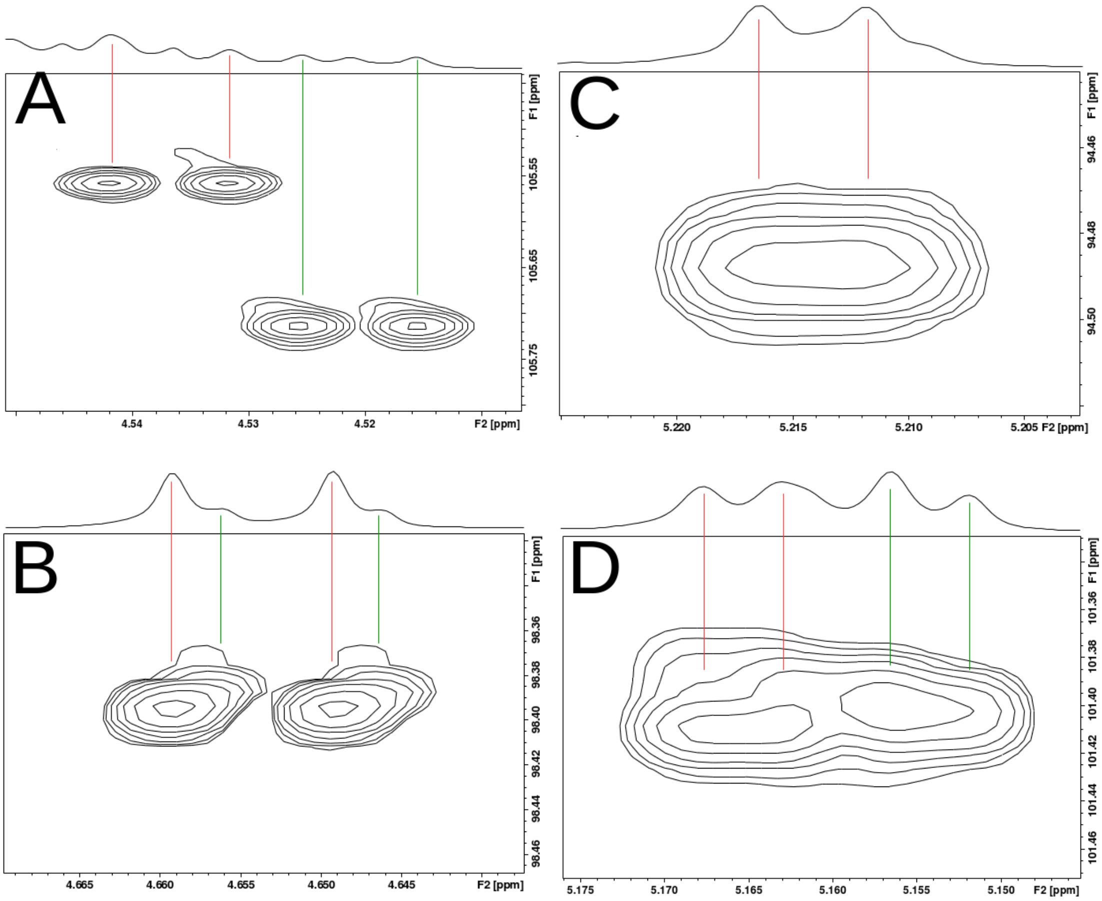


Figure 31: 2D ^{1}H,^{13}C-HSQC spectral regions for four of the anomeric region's peaks, representing A) non-reducing end D-glucose residues, B) reducing end β-D-glucose residues, C) reducing end α-D-glucose residues, and D) further linked D-xylose residues.

Parameter improvements made for spectrum collection, supplementing those for sample preparation, have resulted in resolution at, or approaching, the natural line width which was the desired outcome for carbohydrate structural analysis. Having achieved this resolution, the test-case Xyloglucan sample used during its development required assignment for each of the four repeating units as well as the two breakdown products observed. Generating these assignments provided the requisite starting point for the application of repeating unit quantitative analyses once repeating unit separation had been achieved.

4.2.4 – Application of the Method Developed for Structure Characterisation of a Xyloglucan Sample

As described in the Technical Introduction, the 2D experiments performed each have different abilities, which, when combined allow for carbohydrate structure characterisation. These spectra were each recorded on the test-case xyloglucan sample used during method development, resulting in a comprehensive series of assignments, Appendix 1. The high resolution 2D ^{1}H,^{13}C-HSQC spectra measured, Figure 32, are composed of three distinct regions: that for the CH$_2$ group, A, non-anomeric secondary hydroxyl group, B, and anomeric secondary hydroxyl group, C, protons. From each of the three regions it is possible to see groups of peaks representing residues at the same position but in different repeating units where the pattern of residues surrounding that being represented differ, such as for the non-reducing end and internal D-glucoses shown in Figure 30.

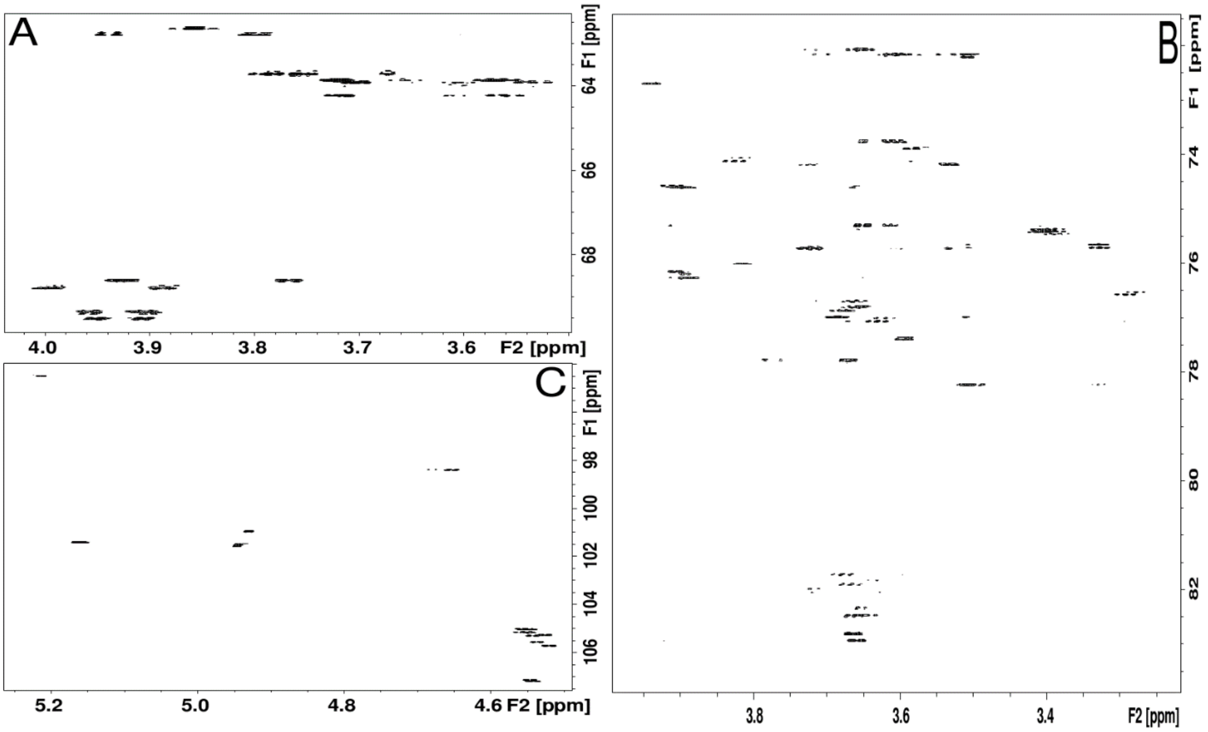


Figure 32: The CH$_2$ group, A, non-anomeric secondary hydroxyl group, B, and anomeric secondary hydroxyl group, C, regions of a 2D ^{1}H,^{13}C-HSQC spectrum recorded on the xyloglucan test-case oligosaccharide sample.

Using the assignment strategy outlined in the technical introduction, six of the residue types assigned in previous publications (Tuomivaara et al., 2015; Ding et al., 2016) were present within the anomeric region for the test-case xyloglucan carbohydrate, these were the α- and β-reducing end D-glucoses, the non-reducing end and internal D-glucoses, the D-xylose residues terminal within a side chain as well as those linked to further residues, and the D-galactose residues terminal within a side chain. Pairing 2D ^{1}H,^{13}C-HSQC and 2D ^{1}H,^{13}C-HSQC-TOCSY spectra, as shown in Figure 33, identified peaks for each of the carbon bound protons present for each residue. Figure 33, showing the two non-reducing end D-glucose variants, highlights the process by which the non-anomeric protons within in the same residue as each assigned anomeric proton were identified. Also evident is the utility of the high resolution spectra collected with separate peaks for the two variants identifiable for the first, second, and forth atom positions, denoted by the green and blue lines. While the 2D ^{1}H,^{13}C-HSQC-TOCSY can also split the third, fifth, and two sixth position atoms into higher and lower peaks, those from the 2D ^{1}H,^{13}C-HSQC are so comprehensively overlapped that the exact assignment position cannot be determined.

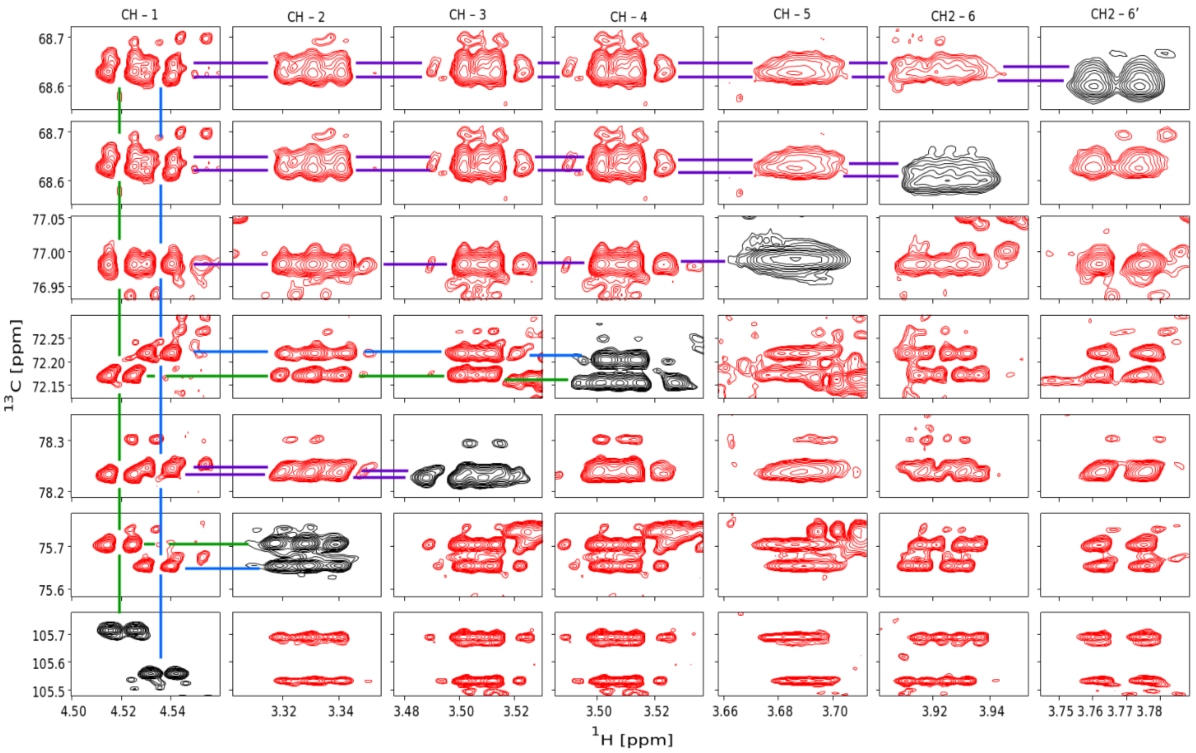


Figure 33: Tile plot showing the 2D $^{1}H,^{13}C$-HSQC peaks for the non-reducing end D-glucose reside, black, alongside 2D $^{1}H,^{13}C$-HSQC-TOCSY cross peaks, red. Green and blue lines indicate where the two variants, i.e. with a D-galactose either present or absent in the internal (second position) residue's bound side chain, can be differentiated in the 2D $^{1}H,^{13}C$-HSQC peak and the purple lines, where they cannot be. Whilst HSQC-TOCSY spectra were also measured with mixing times of 30μs, 60μs, and 120μs, that shown here was run at 240μs. No peak shifting, which could have resulted from heating, due to prolonged mixing time, was observed in the240μs spectrum compared to that with a 30μs mixing time.

Having used 2D $^{1}H,^{1}H$-TOCSY and 2D $^{1}H,^{13}C$-HSQC-TOCSY spectra paired to 2D $^{1}H,^{13}C$-HSQCs to identify which peaks represented CH groups within the same residues, 2D $^{1}H,^{1}H$-COSY spectra were used to determine the order of these peaks. Corroboration of these peak orders, determined by the analysis of 2D $^{1}H,^{1}H$-COSY spectra, was provided by analysis of the 30μs 2D $^{1}H,^{13}C$-HSQC-TOCSY spectrum which showed CH group peaks at most two atom positions away from each starting peak. Using the 30μs 2D $^{1}H,^{13}C$-HSQC-TOCSY spectrum proved particularly useful for identifying peaks in the crowded sections of the 2D $^{1}H,^{1}H$-COSY spectrum, which corresponded to protons from CH pairs bound to non-anomeric secondary hydroxyl groups, owing to the greater distribution of peaks across the two dimensions. This proved successful for the vast majority of observable peaks, however, the CH_2-6 protons from the reducing end α-D-glucose anomer had only one identifiable peak with a pseudo-quartet multiplet structure, Figure 34A. The 1D ^{13}C-DEPT135 experiment performed, Figure 34B, where the α-D-glucose peak is represented by the black arrow confirmed this peak represents both carbon bound protons from the CH_2 group with very close chemical shifts. The unusual multiplet structure of the peak is caused by a strong coupling due to the close chemical shift values. Peak identification for corresponding atom positions in each of the observable residues is important for both structure characterisation and subsequent modification monitoring, however, this generates an incomplete series of assignments as the position of each residue within the repeating unit also needs to be determined.

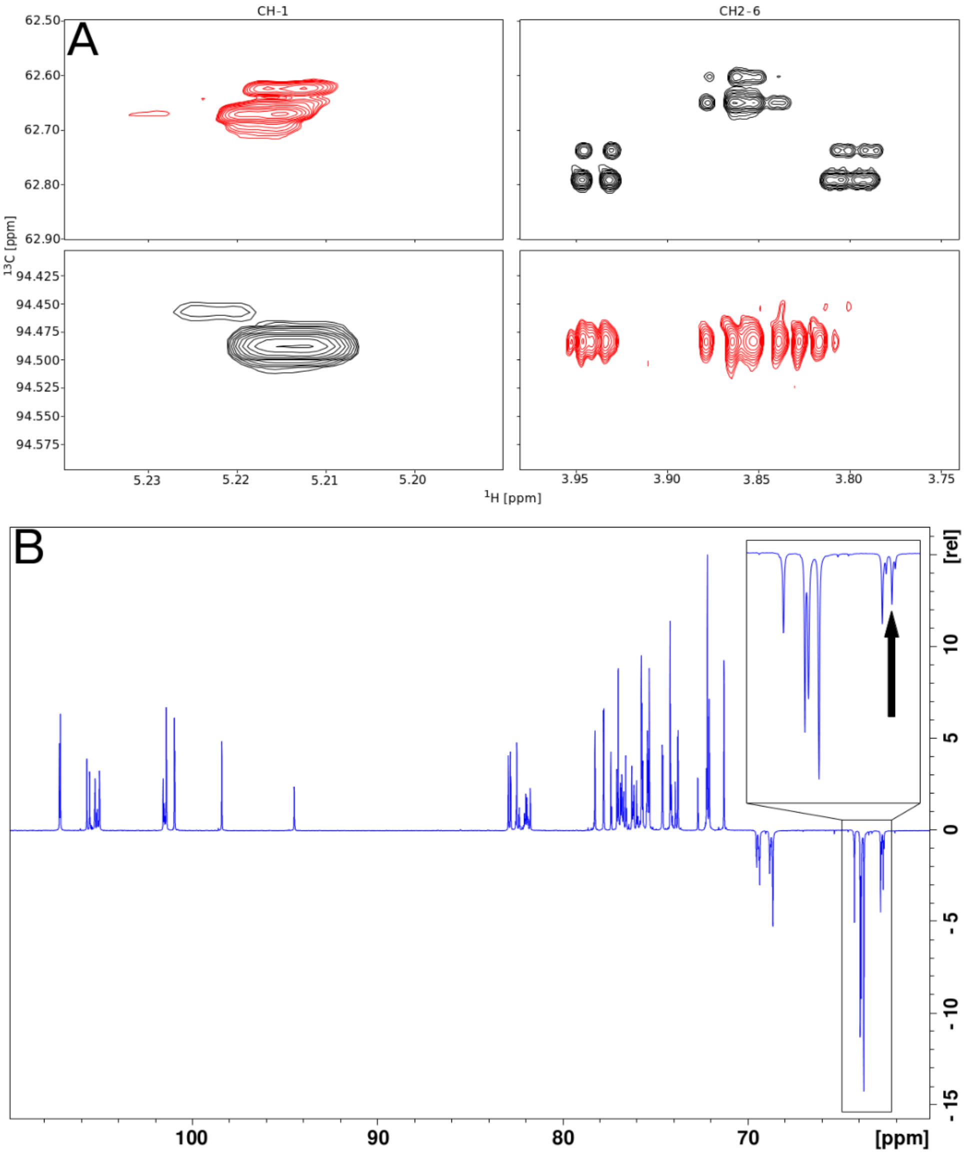


Figure 34: A) showing the 2D ^{1}H,^{13}C-HSQC, black, and 2D ^{1}H,^{13}C-HSQC-TOCSY, red, peaks for the α-D-glucose position 1 and 6, at 3.85/62.65ppm, carbon atoms. B) showing a 1D ^{13}C-DEPT135 spectrum for the hydrolysed xyloglucan sample with the position 6 α-D-glucose peak indicated by the black arrow.

Once assignments had been made for the atoms within individual residues, further spectra were used to determine the linkages between these residues. Two options exist when characterising such linkages, either using 2D ^{1}H,^{1}H-NOESY and/or 2D ^{1}H,^{1}H-ROESY experiments to identify the atoms present within a certain distance of each carbon bound proton or using 2D ^{1}H,^{13}C-HMBC spectra to identify the atoms present within two to three covalent bonds. During the characterisation of the test-case xyloglucan sample, 2D ^{1}H,^{13}C-HMBC spectra were used to identify the linkages present while 2D ^{1}H,^{1}H-NOESY and 2D ^{1}H,^{1}H-ROESY spectra were used for confirmation purposes only. From the 2D ^{1}H,^{13}C-HMBC spectra, it was possible to identify connections between D-glucose residues within the backbone chain, Figure 35A, as well as linkages between the anomeric D-xylose protons and the CH_2-6 D-glucose protons, B and C, and between the anomeric D-galactose proton and the C-2 D-xylose proton, D. Multiple instances are evident for the majority of these linkages, indicative of the different variants for both residue types across the four repeating units. One example of this is shown in Figure 35A where the anomeric peaks from the non-reducing end and internal D-glucose residues differ depending on both their position within the repeating unit and the presence of side chain D-galactose residues.

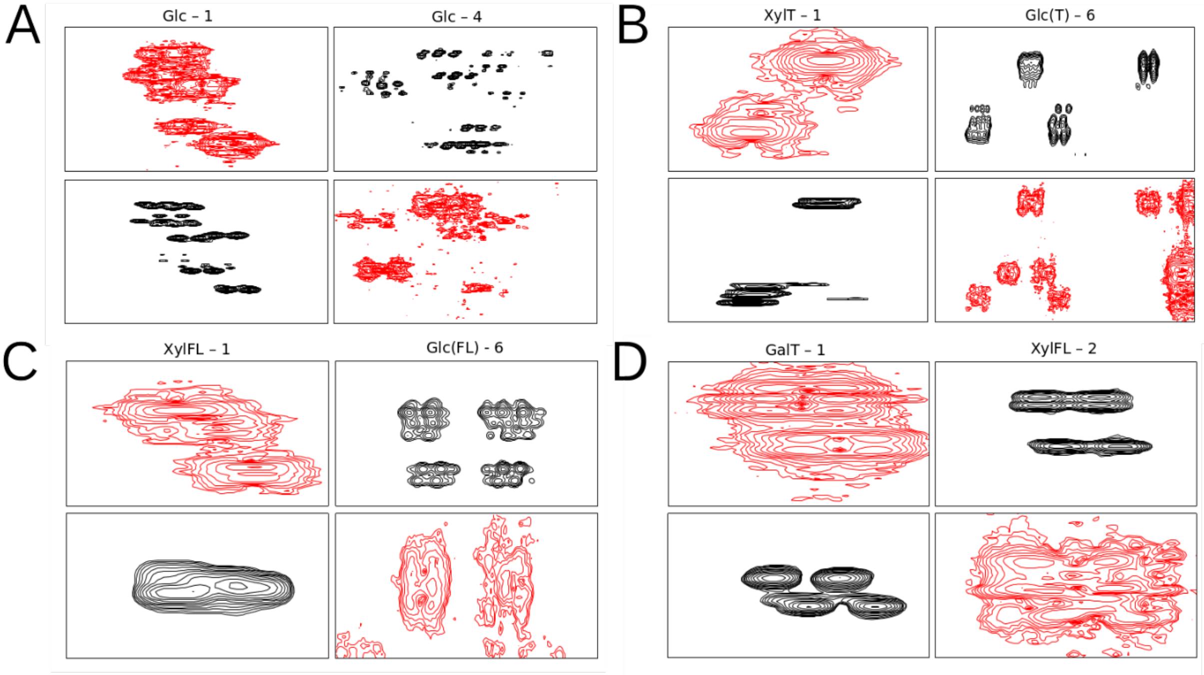


Figure 35: Tile plots showing the linkages between backbone D-glucose residues, A, backbone D-glucose residues and terminal D-xylose residues, B, backbone D-glucose residues and further linked D-xylose residues, C, and further linked D-xylose residues and terminal D-galactose residues, D. 2D ^{1}H,^{13}C-HSQC peaks are shown in black and 2D ^{1}H,^{13}C-HMBC peaks in red. The HMBC spectrum regions displayed are processed to be in the magnitude mode for the ^{1}H dimension.

By combining the partial assignments made from the 2D ^{1}H,^{13}C-HSQC, 2D ^{1}H,^{1}H-TOCSY, 2D ^{1}H,^{13}C-HSQC-TOCSY, and 2D ^{1}H,^{1}H-COSY spectra with the linkages determined by 2D ^{1}H,^{13}C-HMBC, 2D ^{1}H,^{1}H-NOESY, and 2D ^{1}H,^{1}H-ROESY experiments a comprehensive series of individual assignments for each of the repeating units' residues was recorded. The test-case xyloglucan was found to contain the four expected repeating units shown in Figure 21 alongside two structures which appear to be repeating unit fragments, XLG and XXG, produced during the method development and sample assignment process. In certain cases, most notably for the non-reducing end and internal D-glucose residues, the pattern of peaks observed in the 2D ^{1}H,^{13}C-HSQC-TOCSY spectra appeared to support the assigned positions of residues within their repeating units, as determined by 2D ^{1}H,^{13}C-HMBC. This is because the two peaks assigned to the non-reducing end have a 2D ^{1}H,^{13}C-HSQC-TOCSY peak pattern which is distinct from that for the peaks assigned to the internal (second position) D-glucose residue within the repeating unit, and both of these are distinct from the internal (third position) D-glucose residue peaks.

The peak assignments made for the test-case xyloglucan 2D $^{1}H,^{13}C$-HSQC spectrum anomeric region are shown in Figure 36, where a code sequence is used for simplicity. The initial character represents the structure, i.e. A is XLLG, B is XLXG, C is XXLG, D is XXXG, E is XLG, and F is XXG, and is followed by the backbone position within the structure i.e. 1-4 going from the non-reducing to reducing ends, which has either an α or β added to represent the anomer. Following this, there is either a G, X, or L to represent D-glucose, D-xylose and D-galactose residue respectively with the final number representing the atom position with each residue i.e. 1 in Figure 36 which shows only the anomeric region. The comprehensive assignment tables for the xyloglucan test-case sample are given in Appendix 1.

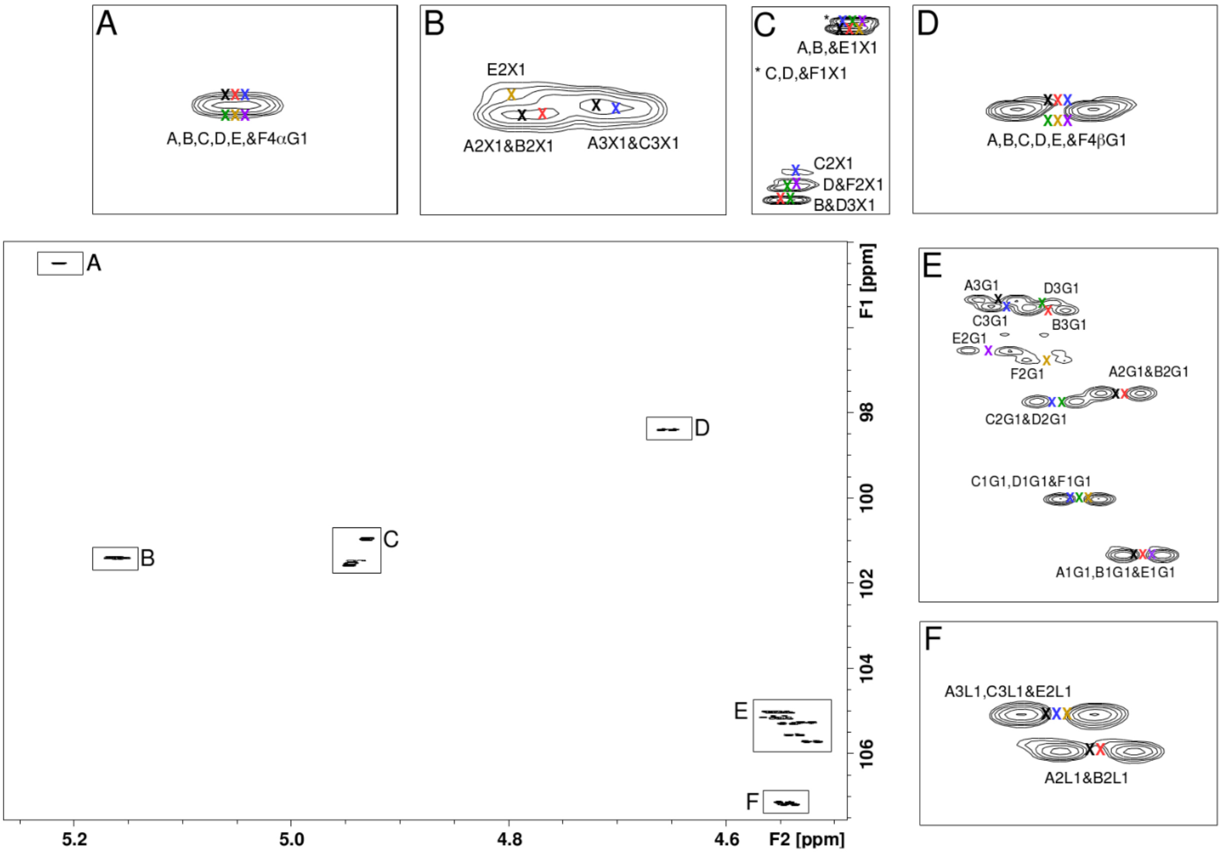


Figure 36: The test-case xyloglucan 2D $^{1}H,^{13}C$-HSQC spectrum anomeric region with boxes showing the assignments of individual peaks or groups of peaks for reducing end α-D-glucose, A, further linked D-xylose residues, B, terminal D-xylose residues, C, reducing end β-D-glucose, D, not-reducing end backbone D-glucose resides, E, and terminal D-galactose residues, F.

Whilst the hexasaccharide XLG repeating unit fragment was expected from the mass spectrometry data, Figure 26, the presence of a second, pentasaccharide, XXG, fragment in Figure 36 was unexpected given both nuclear magnetic resonance and mass spectrometry data were collected from the same sample, albeit with subsequent acetylation for mass spectrometry. While the mass spectrometry data showed peaks in the region around m/z 1420, where one would be expected for the XXG repeating unit fragment, they are of very low intensity and, given the relative sensitivities of the two techniques, should have been larger if they corresponded to the XXG peaks in the nuclear magnetic resonance spectrum. An attempt to reconcile this difference was therefore required, with size-exclusion chromatography used on the repeating unit sample, in a manner which could serve to separate both the repeating units and any fragments present from one another, that also validated the assignments made.

4.2.5 - Size Exclusion Chromatography for Repeating Unit Separation and Assignment Verification

Improvements made to spectral resolution and the resulting high resolution spectra collected for the xyloglucan test-case sample generated a comprehensive series of assignments, shown in Appendix 1. Achieving this highlighted the effect that differing linked residues, as well as those located more distant in a chain, have on the chemical shifts of peaks for residues at the same position of different repeating units, which in many cases remain either wholly or partially overlapped. Therefore, fractionation via size-exclusion chromatography on a Superdex 30 16/600 column was performed and the resulting fractions analysed via both NMR and mass spectrometry. Size exclusion chromatography was intended to separate the repeating units, and repeating unit fragments, into individual fractions which could be used to validate, or if necessary correct, the assignments made for the combined repeating units test-case sample.

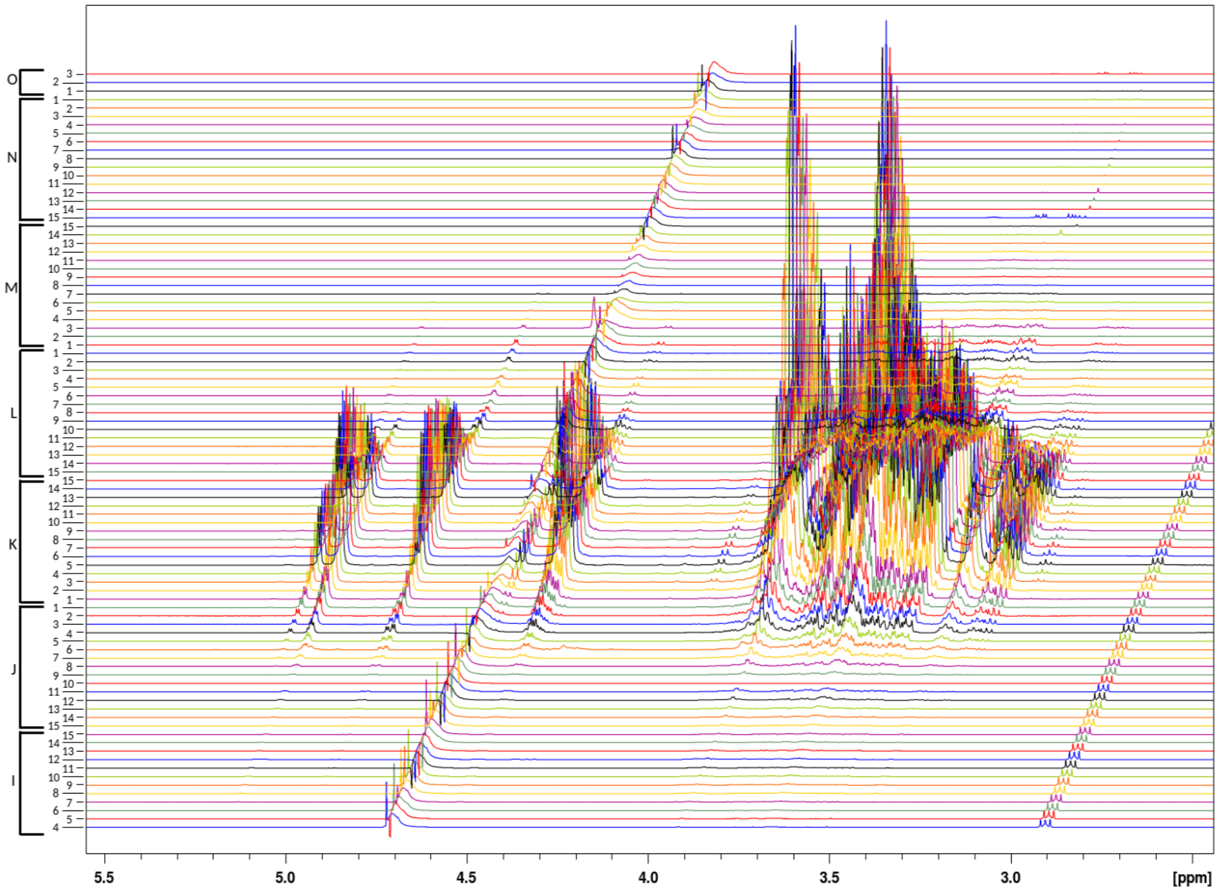


Figure 37: The carbohydrate region of ninety, 1D ^{1}H NMR screening spectra, separated using a Superdex 30 16/600 Column, containing either the xyloglucan repeating units or nothing that could be detected by NMR.

Separation across the Superdex 30 16/600 column generated one hundred and two samples each giving an indication that carbohydrate was present when spotted onto a glass plate and dried before being further characterised by 1D ^{1}H NMR spectroscopy. The spectra collected revealed that the first twelve samples consisted of either nothing detectable via NMR or the L-arabinose-based contaminant which must have undergone partial degradation to allow it to pass through the 3kDa MWCO filtration membranes used during spin fractionation. The remaining ninety fractions, Figure 37, either contained no detectable sample or the xyloglucan repeating units, twenty-eight of which had xyloglucan present in sufficient concentration for high resolution 2D ^{1}H,^{13}C-HSQC spectrum measurement when subjected to integral intensity analysis.

Collection of these high resolution 2D ^{1}H,^{13}C-HSQC spectra showed the rise and fall of peaks present in the different repeating unit fractions as expected given their relative elution volumes from the SEC column and the assignments made, suggesting that the assignments for the unfractionated sample are correct. Subsequent analysis of the peaks present in the fractions derived, using the same assignment process described above, confirmed this with assignments matching those made to the spectra for the combined repeating unit sample. However, assignment of the 2D ^{1}H,^{13}C-HSQC spectra indicated that the different repeating units had not separated completely during chromatography, with peak patterns present that would be expected for mixtures of both XLLG with XLXG/XXLG as well as XLXG/XXLG with XXXG. These spectra also indicated that, whilst the two repeating unit fragments become predominant over the XXXG repeating unit in the later fractions, they do not separate from the intact repeating units completely, Figure 38, Appendix 2.1 and Appendix 2.2.

Whilst the predominant peaks present in the first samples, Figure 38 panels J4 to K5, come from the XLLG repeating unit, minor peaks also appear, from panel J4, for protons present in XXLG. By panel K4 peaks begin to emerge which are indicative of the repeating unit XLXG and become equal in intensity with the other two repeating units by panel K6. Also present in panel K4, becoming predominant by panel L14, and continuing until panel L8 are peaks derived from the repeating unit fragments XLG and XXG. Whilst it can be seen that XXXG is the major intact repeating unit present in panels L15 and L14, it is difficult to determine the fraction in which it appears using NMR alone, similarly the fraction in which XLLG disappears is difficult to determine. This is owing to the peak overlap between those for D-glucose residues with branches including a D-galactose and those with branches comprising solely D-xylose at both positions 2 and 3 which causes XLLG to become indistinguishable from XLXG/XXLG if both are present, as does XXXG.

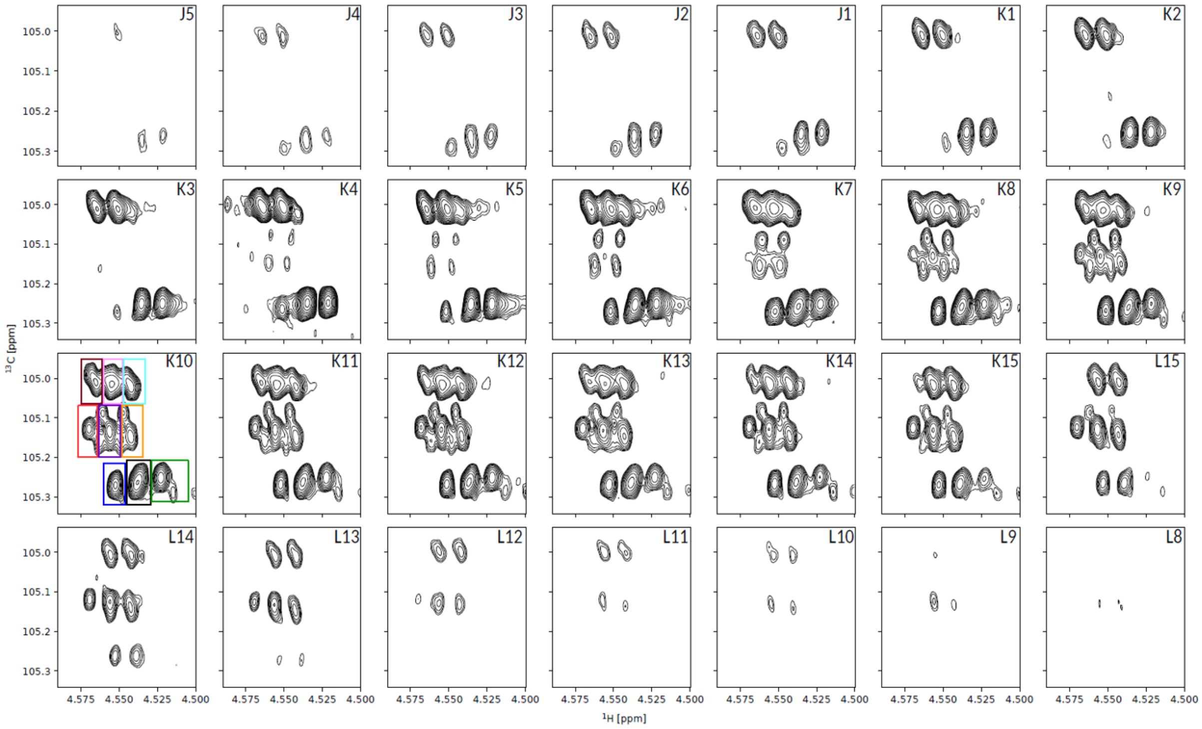


Figure 38: Tile plot showing the internal (second and third) position D-glucoses' anomeric CH cross peaks present from the 2D ^{1}H,^{13}C-HSQC spectra collected on the SEC separated fractions. The coloured boxes present in tile K10 indicate the regions used for the measurement of integral intensities.

The presence of the four repeating units and two repeating unit fragments in each of the fractions used to generate the spectra for Figure 38 can be determined by mass spectrometry, Figure 39, provided that they have first been acetylated. The complementary nature of these two techniques allows for each fraction's composition to be characterised, with mass spectrometry corroborating the suppositions that can be made from analysis of the NMR data. The mass spectrometry data, Figure 39, supports the conclusion that there is both considerable overlap between the XLLG and XLXG/XXLG repeating units as well as between the XLXG/XXLG and XXXG repeating units.

Whilst varying effectiveness of both ionisation and detection mean mass spectrometry was not wholly quantitative, general trends in the abundance values recorded for the size-exclusion chromatography fractions were used to indicate when each repeating unit or fragment was eluted relative to each other. This showed the pattern expected given each repeating unit's number of residues with the largest, XLLG, eluting first followed by the remainder in a size-dependent manner, i.e. XLXG and XXLG, eluting together, then XXXG, XLG, and finally XXG, Figure 39. The abundance values recorded were also suggestive of the relative amounts of each repeating unit or fraction compared to each other, however, these indications cannot be treated as conclusive based on this technique. They show that XLLG was present in greater proportion than the other repeating units, with XXXG present in a lesser proportion, Figure 39. The two observable fragments can also be seen as having been present in roughly equal amounts to the XLXG/XXLG and XXXG repeating units.

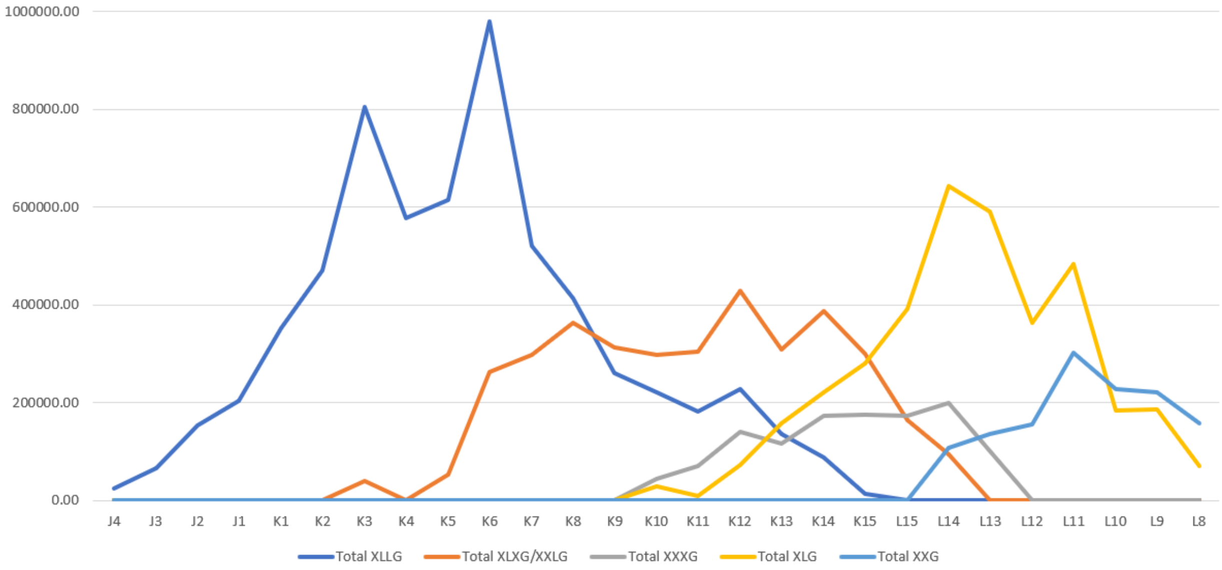


Figure 39: Abundances, recorded in the mass spectrometry Quantitative Analysis Reports, for the three repeating unit, and two fragment, molecular weights where 2497.8 – XLLG – is blue, 2209.7 – XLXG/XXLG – is orange, 1920.5 – XXXG – is grey, 1704.5 – XLG – is yellow, and 1421.4 – XXG – is cyan.

Combining the use of high resolution 2D NMR spectroscopy with mass spectrometry serves to generate a comprehensive and accurate series of assignments. Corroboration of assignments, or otherwise, derived from the analysis of fractions separated in a size-dependent manner further enhances both the capability and reliability of the structural characterisation method developed. However, whilst both of the methods have generated extensive data for the target sample which has been presented thus far, further information can be extracted. Therefore, the collection and analysis of integral intensity data from the NMR spectra measured should act to support the mass spectrometry results while also highlighting the increase or decrease in the intensity of individual peaks. This will be useful for the characterisation of natural polysaccharides; however, it should prove particularly beneficial during later modification monitoring. One of the targets for size-exclusion chromatography, performed on the repeating unit sample, was the separation of the different repeating units and repeating unit fragments. This was only partially achieved from the methods used here. Further, more sensitive, size-exclusion chromatography techniques could have been applied to this sample, however, the lack of required equipment and time prevented this from being completed. These techniques would, however, be an important aspect of any subsequent work to be carried on xyloglucan repeating unit structure characterisation.

4.2.6 – Integration of the Xyloglucan Test-Case Sample for Repeating Unit Analysis:

Integral intensity measurements for peaks in the anomeric region of the spectra collected for the twenty-eight SEC separated samples were performed using scripts developed courtesy of Dr. Rudi Grosman. The integral intensities generated by these scripts, present in Figure 40 as well as Appendices 2.3 and 2.4, mostly represent individual, clearly defined, doublets. In certain cases where two or more assignments are made on wholly or partially overlapping peaks separation proved impossible. Therefore, in these instances, the considerably overlapping regions were conflated into one chemical shift region for measurement. The integral intensities data collected indicated that repeating units containing D-galactose residues at positions two and/or three are present in the earlier fractions, with the XXXG repeating unit difficult to isolate given the presence of the XLXG and XXLG repeating units as well as the two repeating unit fragments, XLG and XXG, Figure 40 as well as Appendix 2.

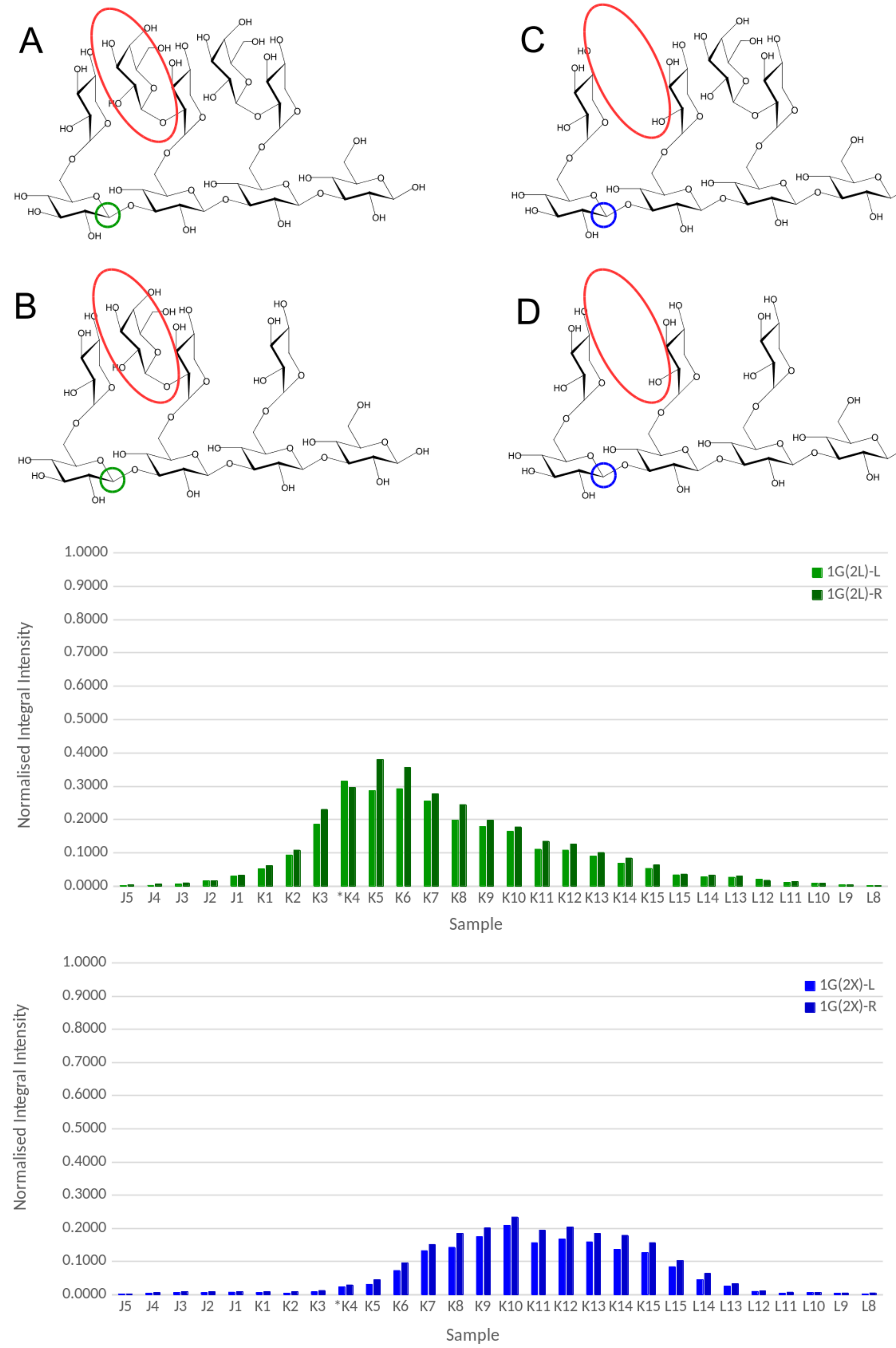


Figure 40: Integral intensity data for the non-reducing end D-glucose residue where the neighbouring side chain either contains, green, or lacks, blue, a D-galactose residue as indicated by the structural diagrams above. The normalised integral intensity corresponds to the intensity of those measured as a proportion of the greatest intensity integral.

The integral intensities collected for non-reducing end D-glucose residues, Figure 40, indicate the repeating units containing a D-galactose in the second backbone D-glucose’s side chain, i.e. XLLG and XLXG, are present in earlier fractions than those where the D-galactose is absent. Those repeating units without a D-galactose in this position, i.e. XXLG and XXXG, become the dominant compounds within the mixture by fraction K10, however, it is not possible to determine which of the two repeating units (XXLG or XXXG) is present in the greatest quantity. It can also be observed, Figure 40, that the proportion of repeating units containing a D-galactose residue in the second D-glucose’s side chain is greater than for the repeating units, where it is absent.

This distribution of intensities for the integrals measured on the non-reducing end D-glucose residues correlated closely with those for the internal (second position) D-glucose with the XLLG and XLXG repeating units present earlier, and at a considerably higher intensity, than those for XXLG and XXXG, Appendix 2 Figure 13. Similarly, the internal (third position) D-glucose integral intensity measurements correlate with those seen for the residues at positions one and two in terms of both the distribution pattern of D-galactose containing and absent repeating units as well as the relative abundances of these two repeating unit types, Appendix 2 Figure 16. From these correlating distribution patterns it is possible to suggest which fractions contain which repeating units, however, this would require further validation from both the D-xylose and D-galactose integrals and, as the signals measured are not specific to an individual repeating unit, any supposition could not be conclusively made based on NMR integration alone.

As would be expected given the close correlation between the integrals of backbone D-glucose residues, at both the non-reducing end and internal positions within the repeating units, the fractions containing D-xyloses present in side chains either containing or absent of a D-galactose residue match with those seen for the D-glucose residues, Appendix 2 Figure 12, Figure 15, and Figure 19. The intensities measured for terminal D-xylose residues give values similar to those seen for the corresponding D-glucoses, however, the intensities for the further linked D-xylose residues do not, even when taking into consideration that a portion of these measurements will come from the XLG and XXG fragments. The abundance of these residues is slightly greater than that which would be expected from the combined reading from the two D-glucoses. However, the severely overlapped nature of the peak used to derive these integral intensities, seen in Appendix 2 Figure 4, is the most likely potential cause of this discrepancy.

The integrals measured for the D-galactose residues also correspond to both those measured for D-glucose and D-xylose in terms of the distribution and intensity, Appendix 2 Figure 14 and Figure 18. It can be seen that the abundance of the second position side chain D-galactose residue is less than that for the third position side chain across a number of fractions, however, the third position side chain D-galactose peak contains a contribution from the XLG fragment with the intensity difference observed partially corresponding to that expected for this structure. There is, however, an increased intensity around fractions K4 to K8 which cannot be assigned to the XLG repeating unit fragment. This may suggest that there is a greater proportion of repeating units with internal (third position) D-glucose residues' sides chain containing a D-galactose residue compared to those of the internal (second position) D-glucose, however, this is far from certain.

The integral intensities measured for the reducing end D-glucose α and β anomers cover the expected range of fractions given the measurements derived from the other residues within the repeating units. There were no chemical shift differences seen in either of these anomers derived from the surrounding residues, however, their distribution indicates 58% β-D-glucose present across the range of sample containing fractions, slightly lower than the 64% previously reported for the monosaccharide in aqueous solution (Ha et al., 1991). These integrals also account for the reducing end α and β anomers from the XLG and XXG repeating unit fragments whose non-reducing end D-glucose residue integrals are similarly incorporated with that for the intact repeating units. Whilst the internal D-glucose intensity for these fragments is separate, Appendix 2 Figure 17, the values measured correlate with what was expected given the measurements for the other residue types within the compound.

Across the range of doublets measured, the left and right peaks have been treated as, and are displayed as, separate measurements in both Figure 40 and Appendix 2.4. In the majority of cases the intensities for two peaks in a doublet match each other quite closely, however, there are a few examples, notably non-reducing end side chain terminal D-xylose and positions two and three further linked D-xylose in Appendix 2 Figure 11 and figure 12 respectively, where this is not the case. The previously noted severely overlapping nature of the further linked D-xylose peak may account for this discrepancy for those residues, however, the same cannot be said for the non-reducing end side chain terminal D-xylose which has two clearly defined, if slightly overlapping, peaks present, Appendix 2 Figure 2. Also worthy of note is the asterisk for the K4 fractions in Figure 40 as well as Appendix 2.3 and Appendix 2.4 used to indicate that the values for this fraction have been halved. A scaling error in the baseline of this fraction, being twice the height of the others, introduced during processing necessitated this amendment.

Integral analysis of the peaks contained within the anomeric region of the spectra recorded provide vast amounts of data which can be used to both support the assignments made and provide an indication for the proportion of different repeating units present within the polysaccharide. Whilst the nature of the assignments make differentiation between the XLLG and XLXG or XXLG and

XXXG repeating units, and consequently a comprehensive determination, difficult, certain suppositions can be made. The integrals recorded show that more than half of the internal side chain D-xylose residues are further linked to a D-galactose giving a proportion greater than would have been expected if the four repeating units were present in equal proportions. Whilst it cannot be determined whether this relates to elevated levels of the XLLG, XLXG, and XXLG repeating units or reduced levels of XXXG, it clearly indicates that the repeating units are not equally distributed. Coupling the integral intensity data with the mass spectrometry results in Figure 39 both of the above causes likely contribute, with a reduced amount of the XXXG repeating unit evident alongside an elevated level of XLLG.

The optimisation of an high resolution NMR based method for structure characterisation, when coupled with MS, enabled the assignment of peaks with far greater specificity than existing techniques. Using the method developed allowed for the assignment of individual atoms on each residue of each repeating unit as opposed to the more general assignments that were previously published (Tuomivaara et al., 2015; Ding et al., 2016; Komatsu and Kikuchi, 2013). Having developed a method for the analysis of a xyloglucan test-case carbohydrate via high resolution spectrum collection and assignment as well as integral intensity and mass spectrometry data analysis there are several uses to which it could be applied. One such use is the characterisation of differences present in samples of the same polysaccharide provided by different suppliers. Whilst such an analysis would not require all of the steps discussed above, notably SEC separation, when examining a carbohydrate previously used for high resolution analysis many of the processes remain relevant.

4.2.7 – Structural Characterisation of Different Suppliers' Xyloglucan Samples

The development of a high resolution structural characterisation method for the analysis of carbohydrate samples enables myriad applications. One such application is the analysis of carbohydrates provided by different suppliers such that variations in the resulting samples' properties can be concluded as either resulting from differences in carbohydrate structure or something independent of this. Other applications for the method developed include the comprehensive analysis of both starting material and modified samples, to which it will be applied in later chapters, such that differences to peak position and intensity can be observed and from this conclusions drawn. Therefore, eight samples, from a range of different suppliers, have been measured using the method developed, and tested on a xyloglucan test-case sample, with the resulting spectra analysed to identify points of both similarity and difference. Tile plots for the previously assigned major peaks are shown in Appendix 3.1 while those for newly identified additional peaks are present in Appendix 3.2.

Each of the eight different suppliers' samples, containing *T.indica* xyloglucan used in research for commercial applications, measured showed the same series of major peaks as were present in the xyloglucan test case sample indicative of structures comprising the same four repeating units. The peaks contained in the anomeric secondary hydroxyl and CH_2 regions are shown in Figure 41 and Appendix 3. There is, however, a difference in the peaks assigned to the repeating unit fragments which are present in differing concentrations as can be seen in Figure 41 where suppliers 1 and 3 have the greatest intensity peaks while suppliers 6 to 8 contain only trace amounts of these fragments. This region also contains two doublets between 105.4 and 105.5ppm which are not present in the test case xyloglucan sample. As the test case sample was supplied from the same source as Figure 41 supplier six this suggests potential batch to batch variations in the carbohydrate provided. Two of the other test case xyloglucan major peak regions, those for the non-reducing end side chain D-xylose anomeric proton and the bound D-glucose CH_2 protons, present in Appendix 3 Figures 4 and 10 respectively, show differences between the suppliers. The non-reducing end side chain D-xylose anomeric proton regions show a series of low intensity doublets in samples 1 to 5 with a relatively high intensity additional peak seen directly to the left of the major peak for suppliers 2 to 4. These same five samples appear to contain an additional pair of doublets in the bound D-glucose CH_2 linked region suggesting that different, closely related, and considerably lower intensity repeating unit variants are also present in these samples.

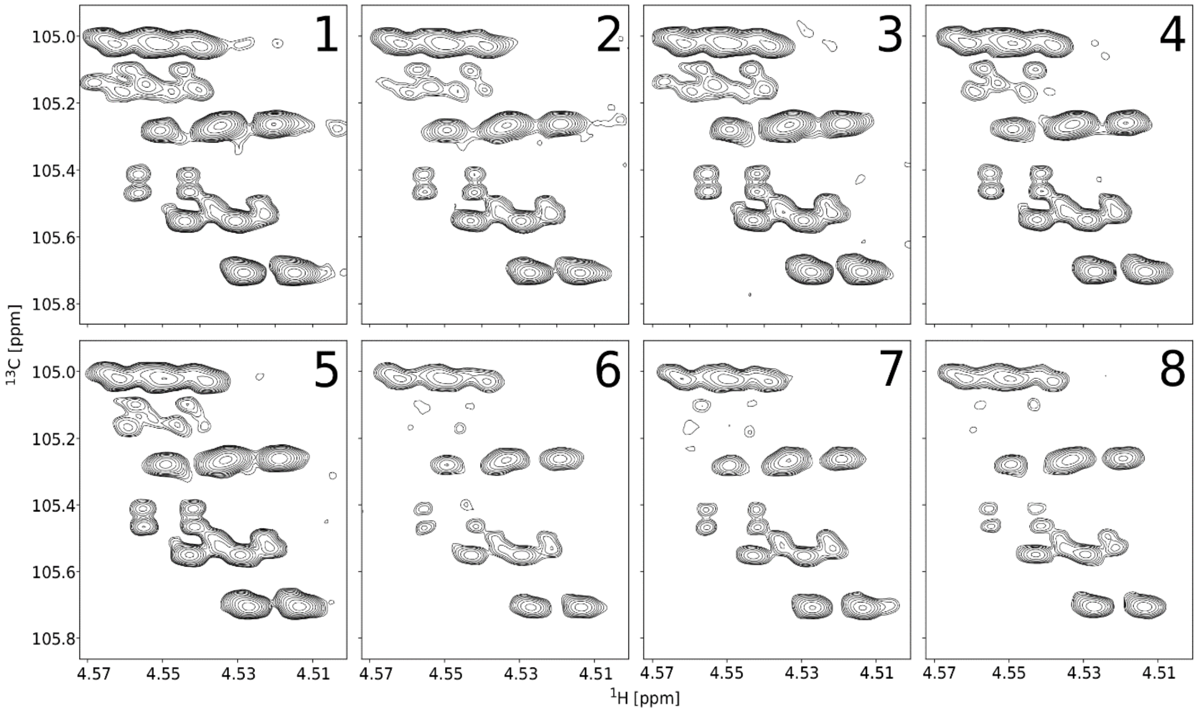


Figure 41: Tile plot showing the non-reducing end and internal D-glucoses' anomeric ^{1}H peaks for eight different supplier samples.

Alongside the additional peaks contained within the same regions as the major peaks from the test-case xyloglucan, several exist within the typical carbohydrate region which do not correspond directly to previously assigned signals, Figure 42 and Appendix 3.2. Within the anomeric region there are four additional regions containing peaks not present in the test case xyloglucan, shown in Figure 42, as well as Appendix 3 Figures 11, 12, 13, and 14. Three of these four peaks, each being doublets, are present in suppliers 2 to 4 and appear to be present at considerably lower intensity than the assigned repeating unit peaks, although integral intensity measurement and analysis would be required to confirm this. The fourth additional section within the anomeric region contains doublets in the spectra for suppliers 2 and 3 again appearing to be of reduced intensity.

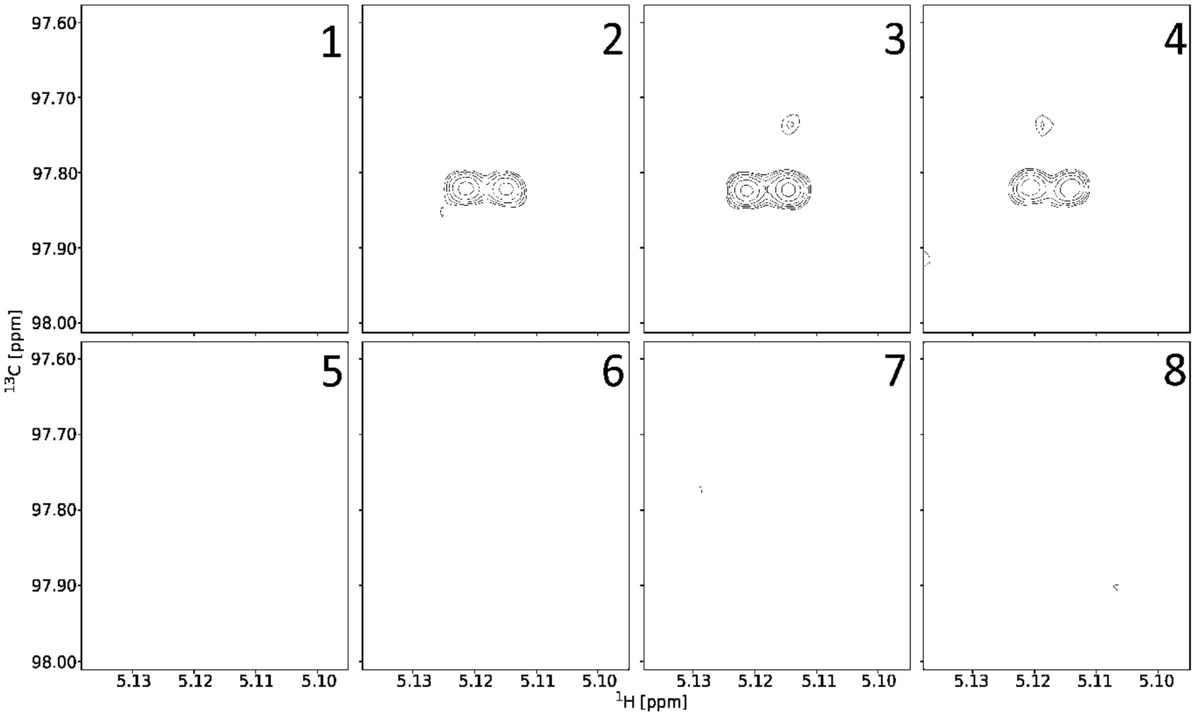


Figure 42: Tile plot showing one of the regions containing additional peaks within the anomeric region of the spectra collected for eight different supplier samples.

Further to the additional peaks seen within the anomeric regions there are three additional sections containing peaks within the CH_2 region of the spectra collected. These peaks are spread across a greater range of the different supplier samples, with the first present in suppliers 2 to 4, the second in suppliers 3, 4, 5, and 8 and the third only present in supplier 8. These different peaks can be seen in Appendix 3 Figures 15, 16, and 17 respectively. While the second and third additional sections contain peaks which are present in the expected pattern i.e. a pair of doublets present at the same carbon chemical shift, the first is not, containing a series of singlet peaks in a diagonal line, suggesting that these peaks are not derived from a carbohydrate.

The spectra collected using the high resolution structural characterisation method developed show that the major peaks present within the eight different suppliers' xyloglucan samples comprise the same repeating unit structures. The previously assigned peaks for the two repeating unit fragments seen in the test case xyloglucan were also identified to greater or lesser extents within all eight of the samples tested. Further to the peaks seen in the test case xyloglucan a series of additional peaks can be observed for some the samples tested in both the anomeric and CH_2 group regions. Some of these peaks occur in close proximity to previously assigned signals from the repeating units, suggesting a variation on the same residue. From the contour levels present for these additional peaks they appear to be present at a considerably lower intensity, however, this would require integral intensity analysis for confirmation.

The mass spectra recorded on the eight different suppliers' samples indicated considerable differences to the repeating units present. Seven of the eight samples contained the XLLG repeating unit, and whilst the peaks present for suppliers 5 and 7 were small they were clearly observable in both cases, supplier 4, however, showed no observable signal for this repeating unit. The other intact repeating units seen in Figure 26 were also observed in some but not all of the different supplier's samples. XLXG and XXLG, for example, were present in suppliers 1, 3, 6, 7, and 8's samples while samples 2, 3 and 5 contained a peak at a slightly higher molecular weight which may have resulted from a greater degree of acetylation having been achieved as the molecular weight difference was too small to imply the addition of an extra residue. This could also have been true of XXXG, which isn't present, at the 1920Da molecular weight seen in Figure 26, in any of the samples, however, peaks are present at the slightly larger molecular weight of 1995Da.

The sole peak that was present across all eight samples was the lower molecular weight repeating unit fragment seen, in Figure 26, at 1704Da. Beyond this each of the samples contained a series of lower molecular weight structures, most notably a cluster of peaks present between 1000Da and 1350Da in all samples, and a varying amount of lower molecular weight peaks than that. These smaller peaks may have represented a greater variety of lower molecular weight repeating unit fractions contained within the different suppliers, however, another possibility is that they were an artifact of the ionisation process, with the breakup of carbohydrates leaving structures to be detected which are not present in the sample tested prior to ionisation.

By contrast, the mass spectrometry data collected indicates there were differences in the repeating units present, although this may be partially mitigated by peaks corresponding to the repeating units seen previously that had a slightly higher degree of acetylation and consequently a different molecular weight. The mass spectra recorded also indicated the presence of a greater number of lower molecular weight variants, however the absence of any signals matching these structures in the NMR spectra recorded suggests many may have resulted from the breakup of carbohydrates during the ionisation process.

4.2.8 – Quantitative Analysis of Different Suppliers' Samples via Integration

Initial quantitative analyses, performed using the same scripts as for the SEC-separated test case repeating unit samples, on the eight different supplier samples, prior to their hydrolysis and separation, indicated a degree of variation for the proportion of L-arabinose-based contaminant present, Figure 43 and Appendix 3.3. Subsequent quantitative analyses were then performed on the hydrolysed samples to determine the proportions of each repeating unit. From the spectra recorded it could be observed that the additional peaks were present at a far lower intensity than those for the target carbohydrate, therefore integral measurements were also recorded for these signals. The integrals data collected was subsequently converted to indicate what percentage each carbohydrate or residue comprises the total integral intensity for the regions tested. Both the anomeric and CH_2 group regions were used for the measurement of repeating unit residue integrals (Figure 44 and Figure 45 respectively) whilst the L-arabinose containing contaminant proportions were determined solely from the anomeric region.

The presence of contaminant compounds within a starting material sample could result in varying consequent properties on an industrial scale if present at different levels between samples, hence the ability to characterise the proportion of both target carbohydrates any contaminants present is important. Therefore, an initial quantitative analysis of the different suppliers' xyloglucan polysaccharide samples focussed on the L-arabinose based contaminant indicating its presence in a wide range of proportions, between 0.2 and 10.3% of the residues contained within the samples. Supplier 1 contained the least L-arabinose based contaminant, with 0.2% of the mixture, being considerably purer than the remaining samples. By contrast, two of the suppliers' samples, 2 and 4, contained the L-arabinose based contaminant as greater than 10% of the total sample at 10.2 and 10.3% respectively, Figure 43. Similarly, supplier 6 which provided the test-case sample used during method development, albeit a different batch, indicated a relatively high proportion of L-arabinose based contaminant, 7.1%. Having identified the proportion of each sample comprising L-arabinose based contaminant, an amount was taken, and the xyloglucan polysaccharide hydrolysed, and spin fractionated to remove any of this contaminant. The oligosaccharides produced were subsequently used for further quantitative analysis to identify any structural differences to the target carbohydrate across the different suppliers' samples.

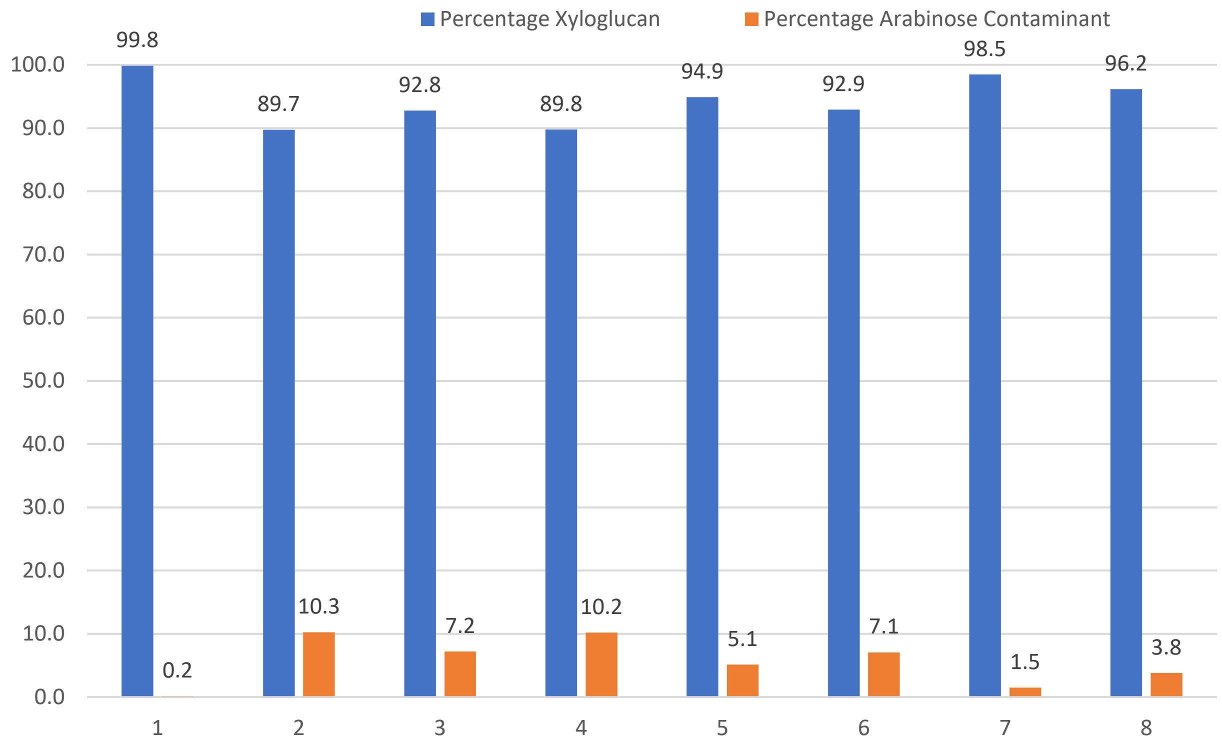


Figure 43: Percentage values for the two carbohydrates present, xyloglucan – blue – and L-arabinose-based polymer – orange – within samples from eight different suppliers.

As noted above the major peaks in the eight different supplier samples tested were present in the same proportions. The non-reducing end D-glucose and two internal D-glucose residue peaks, form the greatest proportion of residues within the anomeric region, Figure 44, comprising 30% of the residues present with the reducing end D-glucose α and β anomers comprising a further 5% and 8% respectively. Two additional major constituents of the samples tested were the terminal and further linked internal D-xylose residues which can guide initial suppositions regarding repeating unit composition given the three-fold higher further linked, 15%, D-xylose present compared to the terminal, 5%, D-xylose alternative. This suggests that there is either a greater than equal proportion of side chains present which contain D-galactose residues, or a less than equal proportion of the non-galactose containing side chains. Consequently, the proportion of repeating units present would be skewed such that either XLLG is present at higher levels than the other repeating units, or XXXG is present at a reduced level. The peaks from the four additional sections measured are present at a far lower intensity than those assigned to the xyloglucan repeating units, Figure 44, confirming the supposition made from the observable peak heights. Three of these four additional peaks can be identified in suppliers 2, 4, and 5's samples, whilst the fourth is contained in suppliers 2 and 4.

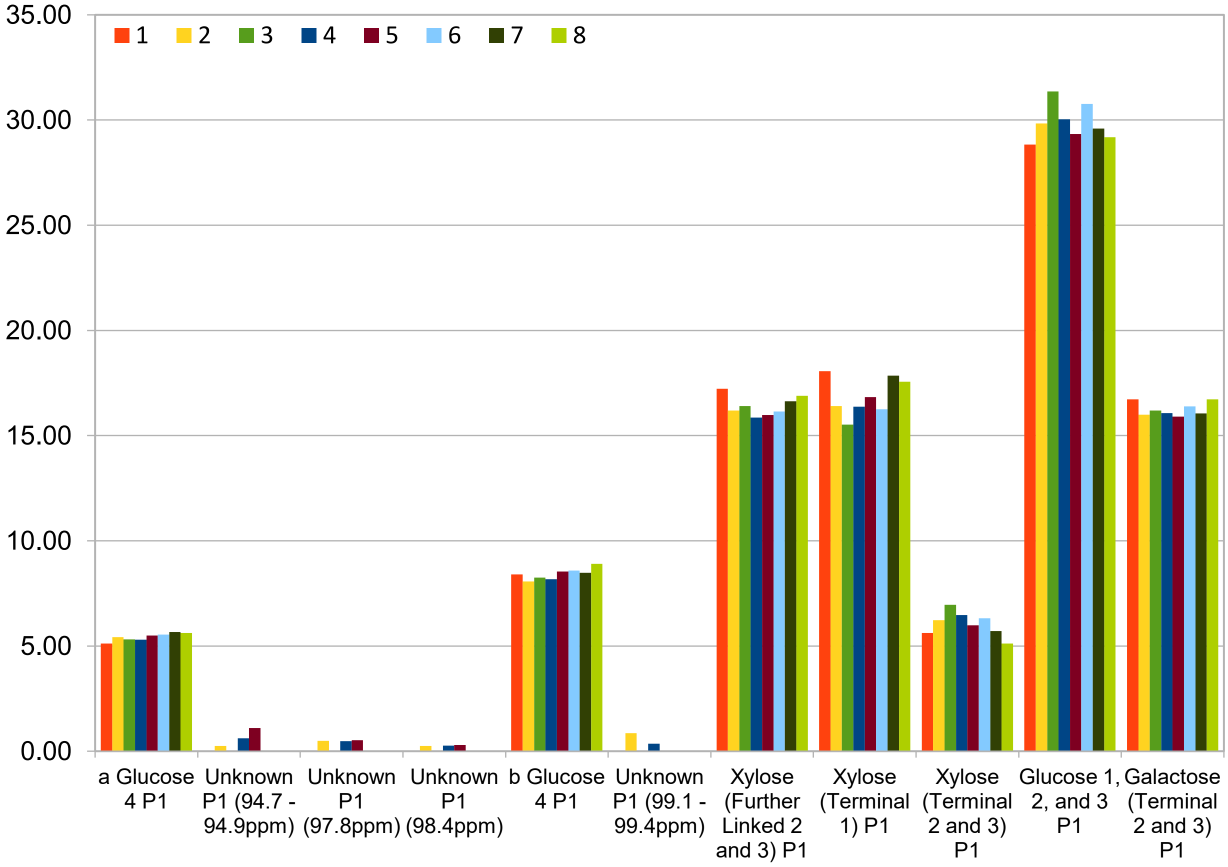


Figure 44: Percentage values for each of the major peaks present within the anomeric region for eight different suppliers' xyloglucan samples alongside those for four additional peaks.

The level of variability across the different suppliers' samples for each residue type differs depending on which peak is being recorded as seen in Figure 44. Those for the α and β D-glucose 4 or D-galactose 2 and 3 residues show very little variation in the percentage that residue comprises the total anomeric region remaining at around 5%, 8%, and 16% respectively. In contrast, the D-glucose 1, 2, and 3 and terminal D-xylose 1 residues show considerable variation with a 2.5% difference between supplier 1, 18%, and supplier 3, 15.5%, samples for the terminal D-xylose 1 peak. A similar variation of 2.6% occurs in the D-glucose 1, 2, and 3 peak between the supplier 1, 28.8%, and supplier 3, 31.4%, samples. Whilst many of the percentages recorded for the major peaks correspond to each other as expected, such as terminal D-galactose 2 and 3 intensities matching those for further linked D-xylose 2 and 3 residues, others are not so clear cut. There is, for example, a greater than expected amount of D-glucose 4 present compared to D-glucose 1, 2, and 3, although this is most likely accounted for by the repeating unit fragments being composed of three backbone D-glucose residues rather than the four seen for intact repeating units.

As with the anomeric region, some of the percentages for peaks within the CH_2 group region appear to differ from that expected. The theoretical number of residues present in the four repeating units and two repeating unit fragments would lead to the expectation of a slightly higher percentage of D-glucose, 22 residues, than D-xylose and D-galactose, 21 residues. However, as can be seen in Figure 45, between 50% and 55% of the residues present are identified as either D-xylose or D- galactose. The ratio of D-glucose residues containing a side chain, 30% of those present, compared to those without one, 13%, however, matches repeating units where the first three D-glucose residues have at least a D-xylose present in the side chain while the fourth residue in the repeating unit is unbound at its CH_2 group. The slight discrepancy between the recorded value for D-glucose residues with an unbound CH_2 group, around 13%, and the expected value, 10%, is likely also owing to the proportion of repeating unit fragments present in the samples, being composed of three backbone residues rather than the four from the intact units.

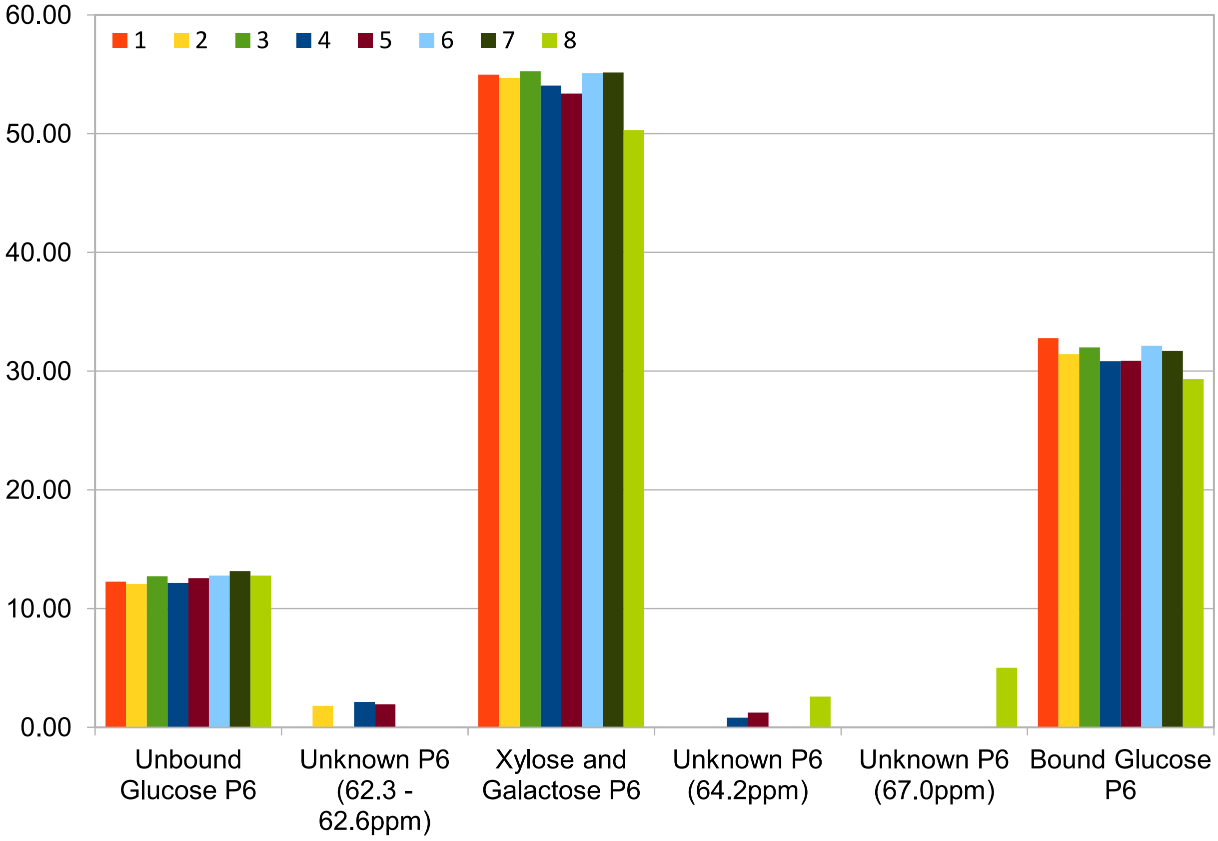


Figure 45: Percentage values for each of the major peaks present within the CH_2 group region for eight different suppliers' xyloglucan samples alongside those for the three additional peaks present.

There is little variability across the different suppliers regarding the percentage of each residue type within the CH_2 region, similar to that seen for most of the peaks within the anomeric region. The most notable difference being the D-xylose and D-galactose peaks from the supplier 8 sample which is 5% less abundant than in the other seven. This may be due to the contaminant peaks seen at 64.2ppm and 67.0ppm which are present at far greater intensity than any of the other additional peaks. Other samples with additional peaks present, suppliers 2, 4, and 5, all have slightly reduced percentages for the D-glucose, D-xylose and D-galactose residues, as would be expected, although not reaching the 5% decrease seen for supplier 8. Two of these suppliers' samples also contain the additional peak at 64.2ppm with all three containing the additional four peaks between 62.3 and 62.6ppm, Figure 45.

Quantitative analysis of the different suppliers was performed using scripts, provided by Dr. Rudi Grosman, for the measurement of integral intensity. These have shown the similarities and differences between samples from different suppliers for both the major peaks comprising the xyloglucan repeating units as well as the less abundant additional peaks observed and whose relationship with the repeating units remains unknown. The data generated from both the assignment of different supplier samples' spectra and the subsequent integral measurement can be used to attempt a reconstruction of the polymer from which these oligosaccharides were derived. As with the test case sample the different suppliers each appear to have a greater proportion of XLLG, XLXG, and XXLG repeating units than they do XXXG, either caused by increased abundance of the former or reduced abundance of the latter. These data can also be combined with those produced during size-dependant xyloglucan test case sample separation to corroborate one another and provide greater confidence that the conclusions being drawn are accurate.

4.3 – Discussion:

The aim of this chapter was to develop a carbohydrate structure characterisation method with NMR spectra collected at a resolution with peaks present at their natural line width. Further to this the analysis of a series of different suppliers' samples was targeted both as a means of validating the method created and to determine the composition of samples which could, if later applied industrially, be used as starting materials in modification reactions or included directly in commercial applications.

The desired method for carbohydrate structural analysis has been developed and used to generate spectra at their natural line width. From this method and using scripts for the quantifiable analysis of resulting spectra a considerable amount of data has been generated and subsequently used to make suppositions about the composition of the xyloglucan polysaccharides examined. The data collected enabled the conclusion that xyloglucan's repeating units are not present in equal proportions with a greater amount of XLLG, or a reduced amount of XXXG, indicated using size-exclusion chromatography-separated samples for both mass spectrometry and integral intensity data analysis. Within these samples, fragments were identified which were believed not to be repeating units present in the polysaccharide, rather they have resulted from the hydrolysis or separation processes required for measurement of NMR and MS. Whilst these fragments appear to be caused by either hydrolysis or separation, they are present uniformly each time the same xyloglucan sample was hydrolysed by xyloglucanase, which potentially suggested that this effect was the result of the enzyme. Alternatively, it is possible that when hydrolysed to the tetrasaccharide backbone repeating unit structures, certain residues were susceptible to spontaneous further degradation independent of the enzyme. The peaks for these repeating unit fragments are an example of those which are seen to differ across the eight different suppliers' samples tested. Whilst these eight samples were seen to comprise the same four repeating units, a series of additional low intensity peaks were identified that could correspond to variants of the residues comprising the starting polysaccharide.

The method developed built on previous techniques which have used either solely 1D NMR experiments (Tuomivaara et al., 2015; Hotchkiss et al., 2015) or low resolution 2D experiments (Komatsu and Kikuchi, 2013; Ding et al., 2016; Arruda et al., 2015) for analysis. The 2D spectra recorded in these prior analyses give similar results to those seen when measuring the polysaccharide in this research, including the lower than expected intensity for the non-reducing end and internal D-glucose peak, and the presence of L-arabinose within the sample. However, there appears to have been little attempt to characterise xyloglucan using the method of hydrolysis, separation and high resolution NMR that has been employed in this instance with equally limited evidence of 2D experiments beyond 2D ^{1}H,^{13}C-HSQCs, 2D ^{1}H,^{1}H-COSYs, and 2D ^{1}H,^{1}H-TOCSYs having been used. Coupling the additional 2D techniques employed here on hydrolysed oligosaccharides with the integral intensity based data analysis techniques applied has resulted in more specific assignments, such as identifying individual D-glucose anomeric peaks containing different side chains and in some cases with differing side chains on neighbouring residues, rather than assigning all D-glucose anomerics as either one or two peaks (Ding et al., 2016; Arruda et al., 2015).

Whilst the development of the method outlined in this chapter represents progress for the structural characterisation of carbohydrate samples, alternative nmr pulse programs, pulse program combinations and computational analysis tools can be used for carbohydrate structure characterisation (Brodaczewska et al., 2018; Kupče et al., 2021; Kapaev and Toukach, 2018), whilst other methods can be employed to replace certain steps. One of these alternatives is the application of diffusion NMR in place of the physical, size-dependent, separation techniques favoured here. DOSY based NMR experiments could be used in place of spin fractionation for the analysis of the xyloglucan and L-arabinose-based contaminant polymer proportions within a sample. Coupling these analyses with a standard curve using calibrated samples would allow for molecular weights of both components to be calculated (Viel et al., 2003; Sinnaeve, 2014). Further to this, the separation of repeating units performed here in a size dependent manner could again be replaced by diffusion NMR, in this case 3D DOSY-HSQC or 3D HSQC-iDOSY (Dal Poggetto et al. 2018; Vitorge and Jeanneat, 2006), and analysed using recently developed computational software, such as MAGNATE (Dal Poggetto et al. 2018). These two methods would be natural points of progression for the research carried out here and could, particularly for repeating unit analysis, represent a time saving compared to size exclusion chromatography.

In conclusion, a high resolution structure characterisation method has been developed, tested on a complex carbohydrate, and used for the analysis of eight samples provided by different suppliers. Whilst further method development could incorporate diffusion NMR based separation methods, in place of physical size-exclusion chromatography based separation, the method used here is effective in its current form. Further applications of the developed method targeted towards the structural characterisation of modified samples will be tested and, if necessary, optimised in Chapter 5. Specifically, this will examine both the sites and degree of sulfation on time course samples where both existing and novel methods have been employed for the modification process.

Chapter 5 – Structural Characterisation of Modified Polysaccharides: Application of Novel Sulfation Methods

5.1 – Introduction

The pattern of sulfate groups present on polysaccharide chains can have a major impact on their biological and interaction properties. For example, heparin, a widely prescribed anticoagulant, relies on a high degree of sulfation for its pharmaceutical activity while different sulfation patterns effect its ability to interact with members of the Fibroblast Growth Factor family (Onishi et al., 2016; Rabenstein, 2002; Ornitz and Itoh, 2015; Xu et al., 2012; Li et al., 2016). Altering the sulfation pattern of carbohydrates can therefore affect their ability to interact with proteins and generate distinct biological activities as a result. Increasing or decreasing the degree of sulfation for both cellulose and the 3,6-anhydrogalactose ring containing carrageenan samples for example, generates varying anticoagulant activities; polysaccharides having a higher degree of sulfation proving better suited to this application (Wang et al., 2010; de Araújo et al., 2013).

Commonly used methods for carbohydrate sulfation can be split into two principal groups, those using sulfate containing acids, such as sulfamic or chlorosulfonic acids, under extreme reaction conditions (Novikova et al., 2007) and those using gentler conditions with a sulfur trioxide complex as the sulfating agent (de Araújo et al., 2013; Xu et al., 2016). Whilst sulfur trioxide complex-based reactions are considered 'greener' than their acid-based alternatives, the commonly used solvents, dimethylformamide and pyridine, are an obstacle to performing sulfation reactions in an environmentally friendly manner. Another problem when performing pyridine sulfur trioxide complex in dimethylformamide sulfation reactions relates to the required conversion of the carbohydrate into a hydrophobic salt that would be soluble in the organic solvents used. Water insoluble and gel forming polymers would be particularly difficult to convert to these salts owing the potential for cation exchange column blockages, amongst other factors. These issues can be mitigated by using a recyclable solvent which is also capable of dissolving gel forming and water insoluble carbohydrates in their native salt forms, here one of the prevailing ionic liquids, 1-butyl-3-methylimidazolium chloride, was chosen.

Ionic liquids are an increasingly diverse class of solvents which offer a recyclable, and hence more environmentally friendly system for chemical carbohydrate modification while the advent of variants specifically tailored to individual tasks has led to their use in multiple applications (Vekariya, 2017; Olivier-Bourbigou et al., 2010). The ability of ionic liquids to dissolve cellulose (Swatloski et al., 2002; Isik et al., 2014) allows for the modification of the most abundant natural carbohydrate more feasibly in ways previously impossible owing to its crystalline nature. Application of ionic liquids to polysaccharide sulfation follows the same pattern as for the traditional protocols where the anionic polysaccharide is coupled with either acid or sulfur trioxide complex based sulfating agents (Gericke et al., 2009; Chopin et al., 2015). Whilst previously published methods have used ionic liquids to dissolve cellulose, they have retained dimethylformamide as a solvent comprising at least a third of the total reaction volume for modification when coupled with both acids (Wang et al., 2009) and sulfur trioxide complexes (Gericke et al., 2009; Chopin et al., 2015). The aim of the research described in this chapter is twofold, first to develop a method which can produce sulfated carbohydrates using as few as possible environmentally damaging compounds. Second, to test the method developed in the previous results chapter in order to determine its capability for analysing both the sites and degree of sulfation. NMR provides an indirect detection method for monitoring sulfation reactions where the addition of these highly electronegative functional groups causes a considerable shift to the peak for sulfate bound CH groups in both the ^{1}H and ^{13}C dimensions.

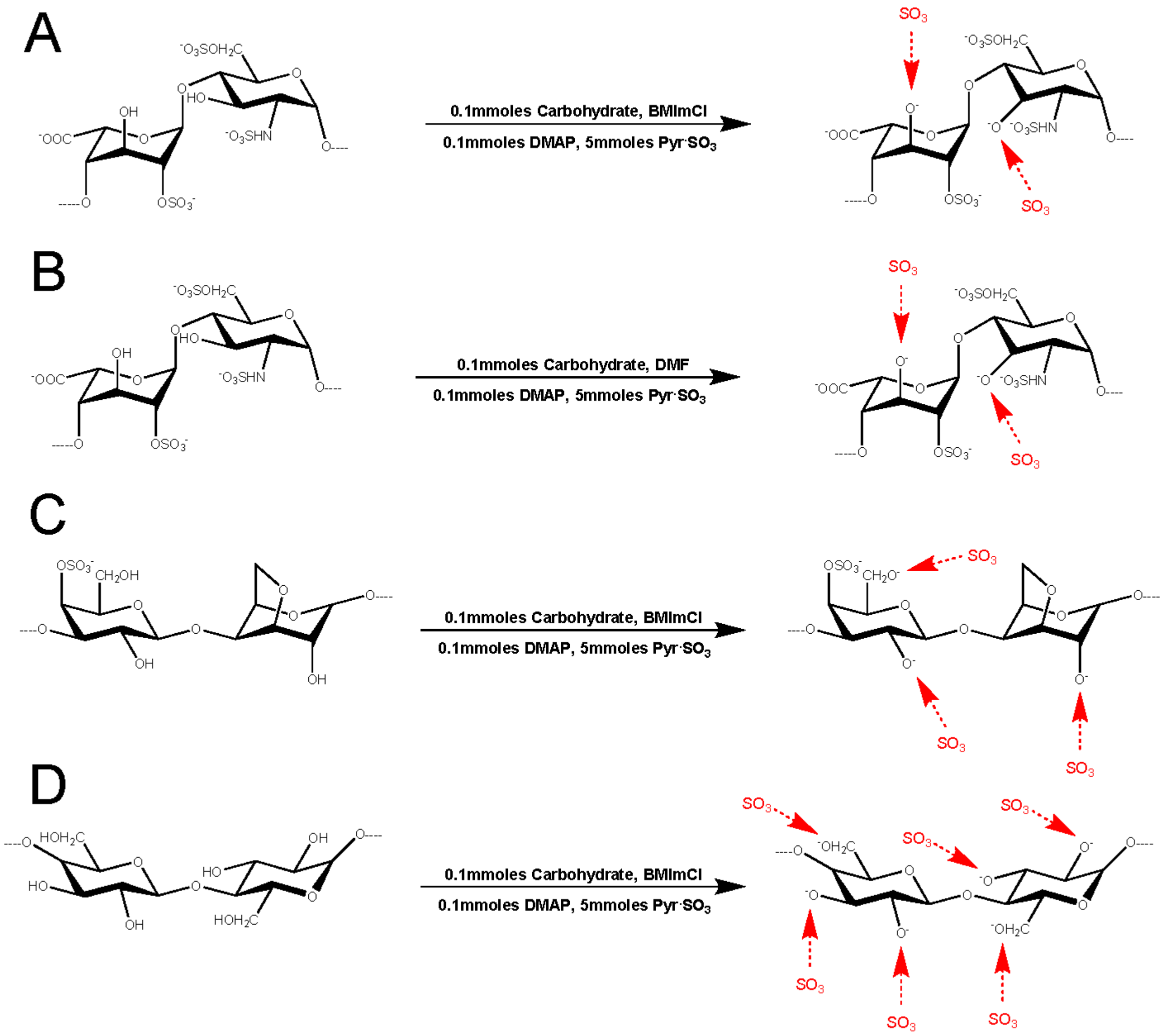


Figure 46: Reaction schemes and locations of potential sulfate group modification sites for four of the five carbohydrates tested with the 1-butyl-3-methylimidazolium chloride and dimethylformamide methods shown for heparin, A and B, as well as κ-carrageenan, C , and cellulose, D. The SO_3 groups shown in red indicate the potential sites of sulfation.

5.2 - Results

5.2.1 – Optimisation of a Sulfation Method in 1-Butyl-3-Methylimidazolium Chloride

Having optimised a system for the analysis of carbohydrates, a series of variably sulfated samples were produced, to which this analysis method was applied after initial screening. The sulfation reactions performed on the plant-based carbohydrates, cellulose, xyloglucan and κ- and ι-carrageenans, alongside the biopharmaceutical, heparin used an adapted pyridine sulfur trioxide, hereafter Pyr·SO_3, complex based method with solvation in ionic liquids. Previous Pyr·SO_3-based methods using ionic liquids have focussed on the inclusion of 1-butyl-3-methylimidazolium chloride, BMImCl, for the dissolution of carbohydrates that are insoluble in dimethylformamide, DMF, whilst incorporating DMF in the solvent mixture upon addition of the sulfating agent (Chopin et al., 2015). The experiments performed here, however, removed DMF from the reaction mixture completely to determine whether ionic liquids alone could be used for successful carbohydrate sulfation, irrespective of their solubility in the traditional organic solvents. The conditions required for these reactions, including 4-dimethylaminopyridine, DMAP, and Pyr·SO_3 concentrations as well as the ionic liquid solvent used, were therefore investigated to determine whether such reactions would be possible.

Reaction condition optimisation protocols used the red algae derived carbohydrate κ-carrageenan, initially in 1-butyl-3-methylimidazolium tetrafluoroborate, $BMImBF_4$ at 85°C. One of the key factors to be optimised, Pyr·SO_3 concentration, was tested within a range of 0.5 to 50 molar equivalents, relative to the repeating unit disaccharide concentration. These samples also contained 0.5 molar equivalents of the catalyst, 4-Dimethylaminopyridine (DMAP), again relative to the repeating unit disaccharide concentration. The non-anomeric regions of the spectra recorded for the starting material and seven Pyr·SO_3 concentrations tested, Figure 47, exhibited no observable difference between the starting material and when 0.5 molar equivalents Pyr·SO_3 was used. An increased concentration of 1, 2.5, or 5 molar equivalents Pyr·SO_3 generated a series of newly emerged observable peaks, however, these did not relate to the three hydroxyl group sulfation targets, coloured red, blue and green. Also observed in the spectra for 2.5 molar equivalents Pyr·SO_3 was a peak representative of the galactose-4-sulfate residue in which the primary hydroxyl group had been sulfated. This peak was seen to increase in intensity with increased Pyr·SO_3 concentration until the final concentration tested, at 50 molar equivalents. Further to this, 25 and 50 molar equivalent concentrations of Pyr·SO_3 gave rise to sulfation at position two hydroxyl groups from both the galactose-4-sulfate and 3,6-anhydrogalactose residues present within the polymer.

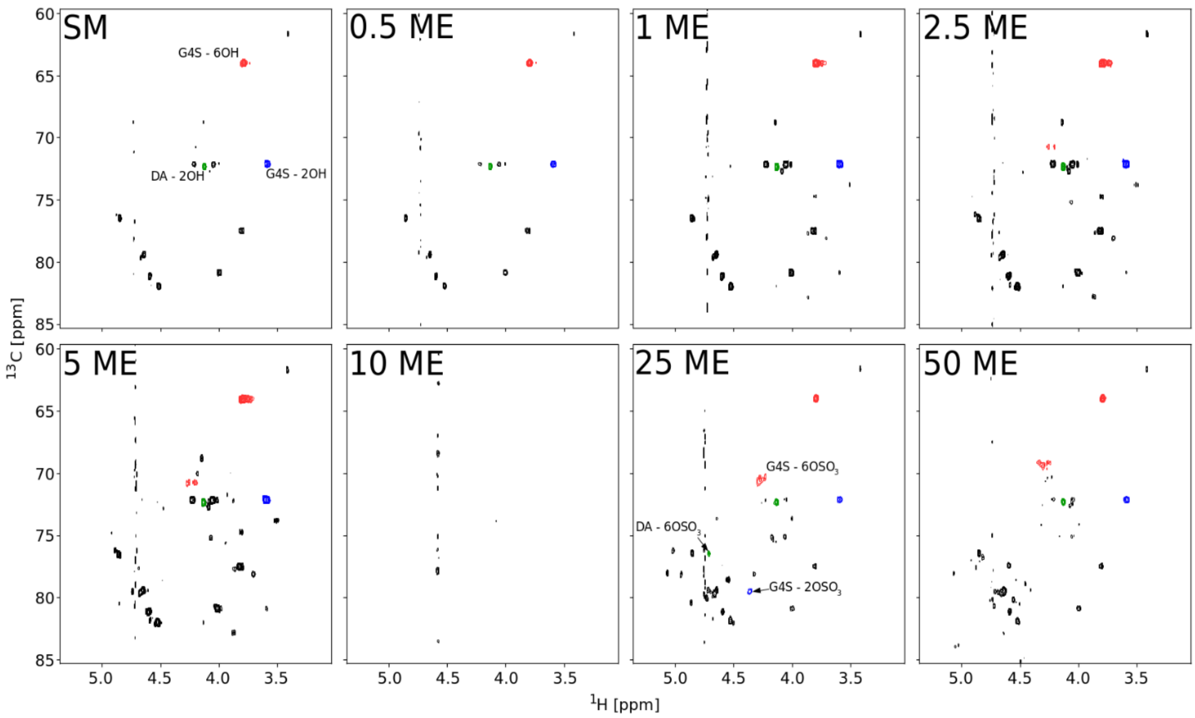


Figure 47: κ-Carrageenan in BMImBF$_4$ reaction PyrSO$_3$ optimisation spectra for the starting material and Pyr·SO$_3$ concentrations between 0.5 and 50 molar equivalents with the galactose-4-sulfate positions six, red, and two, blue, and 3,6-anhdrogalactose position two, green, hydroxyl and sulfate group peaks indicated. G4S indicates the galactose-4-sulfates residues and DA, the 3,6-anhydrogalactose residues within each dimer. The modified carbohydrate present in the 10ME sample has been lost, this was most likely during the extraction and washing processes post modification.

Measurement of integral intensities, courtesy of scripts provided by Dr. Rudi Grosman, for samples containing the three sulfated targets or their native hydroxyl groups was therefore used to quantitatively determine the proportion of both modified and unmodified groups present within the polysaccharide. The proportion of the three hydroxyl groups which were sulfated at each Pyr·SO$_3$ concentration, Figure 48, corroborated some of the observations made from the qualitative analysis of the spectra. The general pattern observed in Figure 47 of increased sulfation with increased Pyr·SO$_3$ concentration was present for each of the three sulfation sites. 6-O-sulfation of galactose-4-sulfate increased, from 3.6% to 53.7% between 1 and 50 molar equivalents of Pyr·SO$_3$, as did its position two sulfation, which increased from 2.6% sulfated in the starting material to being 59.4% sulfated having been reacted with 25 Pyr·SO$_3$ molar equivalents. Similarly, the sole potential sulfation site at position two of the 3,6-anhydrogalactose residue increased from fluctuating between 0% and 1.2% to having 17.5% of these residues include a position two sulfate group at 25 molar equivalents of Pyr·SO$_3$. However, there are instances where this general trend did not hold true with a decreased proportion of position two galactose-4-sulfate hydroxyl group sulfation from 59.4 to 42.6% and the absence of position two sulfated 3,6-anyhdrogalactose residues, down from 17.5% at 25 molar equivalents, in the 50 molar equivalents sample. As can been seen in Figure 47 and Figure 48 there is also a complete absence of peaks in the spectrum for 10 molar equivalents of

Pyr·SO_3. It is not known what caused this anomaly however one possibility could be the presence of residual water, while another is sample loss during the extraction and washing procedures. Despite the unexpected absence of 2-O-sulfated 3,6-anhydro-α-D-galactose in the 50 Pyr·SO_3 molar equivalents sample this concentration generated the highest level of sulfation and as such was selected as the optimised concentration for subsequent sulfation reactions. This was consistent with a general pattern of increased sulfation having occurred with increased Pyr·SO_3 concentration.

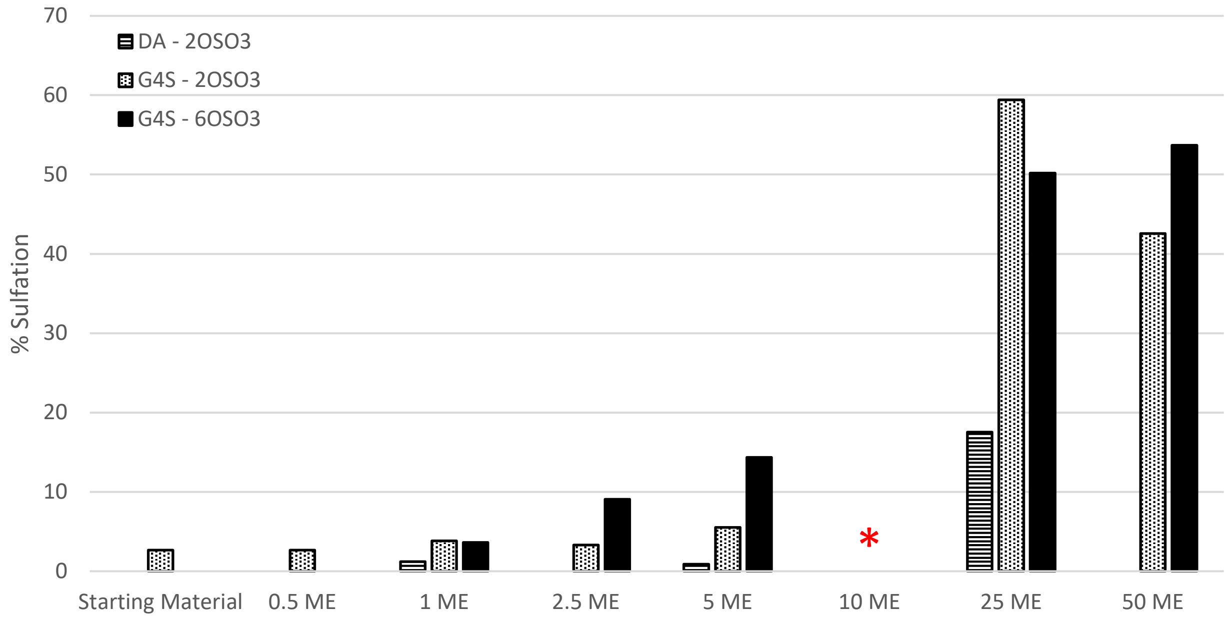


Figure 48: κ-Carrageenan in $BMImBF_4$ reaction $PyrSO_3$ optimisation of the sulfation reaction, with samples containing 0.5 molar equivalents DMAP. The bars with different fills represent galactose-4-sulfate positions six (black), two (spotted), and 3,6-anhdrogalactose position two (striped) sulfate group proportions, derived from integral intensity measurements. The red star for 10 molar equivalents of Pyr·SO_3 indicates the lack of integrals collected owing to an absence of peaks within this spectrum. G4S indicates the galactose-4-sulfates residues and DA, the 3,6-anhydrogalactose residues within each dimer.

Having determined that 50 molar equivalents of pyridine sulfur trioxide complex was optimal for the sulfation of κ-carrageenan in $BMImBF_4$, the DMAP catalyst concentration was optimised for a range of concentrations between 0.1 and 5 molar equivalents. As previous studies identified that DMAP acted as a catalyst, promoting and accelerating the carbohydrate sulfations performed, (Chopin et al., 2015; Wang et al., 2015), it was added to the reactions, shown in Figure 49 for κ-carrageenan, performed during this research. Analysis of the observable peaks within the spectra collected, Appendix 4 Figure 2, indicated lower DMAP concentrations were optimal owing to the apparent extent of the modification achieved, when reactions were performed with 5 molar equivalents Pyr·SO_3. These additional peaks included one for position 6-O-sulfated galactose-4-sulfate at DMAP concentrations between 0.1 and 2.5 molar equivalents with second emerging for 2-O- and 6-O-sulfation of galactose-4-sulfate sulfation between 0.1 and 1 DMAP molar equivalents. There were no peaks observable for the sulfated position two of the 3,6-anhydrogalactose residues across the range of concentrations tested. From these reactions a concentration of 0.1 DMAP molar equivalents was optimal and consequently was used for subsequent modification reactions.

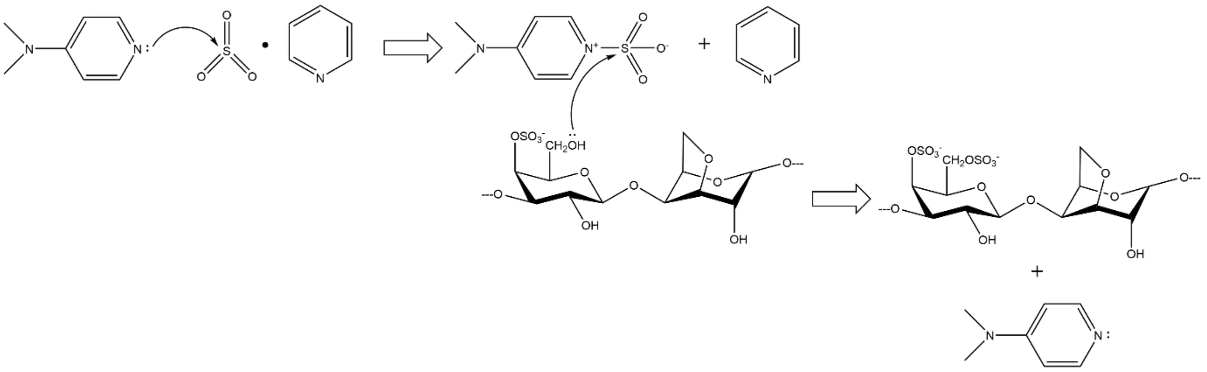


Figure 49: Reaction mechanism of the DMAP catalysed pyridine sulfur trioxide sulfation of κ-carrageenan.

As with Pyr·SO_3 concentration optimisation, the integral intensity measurement for the starting material and six DMAP concentrations, Figure 50, provided a quantitative analysis of the reactions performed. From these integrals the proportion of position six sulfated galactose-4-sulfate residues decreased with increasing DMAP concentration from 18.3% at 0.1 molar equivalents to 7.7% at 5 molar equivalents. A similar decreasing trend was also present for the position two sulfated galactose-4-sulfate residues which dropped from 6.2% to 3.4% between 0.1 and 2.5 molar equivalents DMAP, however, in this instance there was then a dramatic rise in the proportion of position two sulfated galactose-4-sulfate residues, to 7.4%, in the 5 molar equivalents DMAP sample. In contrast to these, increased DMAP concentrations generated increased sulfation of the 3,6-anhydrogalactose position two hydroxyl group at 2.5 and 5 molar equivalent concentrations. rising from 0% at 1 molar equivalent to 1.9% at 5. As can be seen from Figure 50, and has been described above, whilst there were broad trends toward decreased sulfation at multiple sites when the molar equivalents of DMAP were increased, this trend was not as apparent as that which was observed with increased pyridine sulfur trioxide molar equivalents, Figure 48. This may have indicated that varying the amount of DMAP present is not as important as varying the amount of pyridine sulfur trioxide. However, it could also have indicated that there may be additional factors affecting the reaction, other than the amount of DMAP or pyridine sulfur trioxide present, which remained unknown. In spite of this increase at higher concentration, and the increase observed for the galactose-4-sulfate position two sulfation at 5 molar equivalents, the lowest concentration of DMAP, 0.1 molar equivalents, tested was carried forward to be used in the time course experiments.

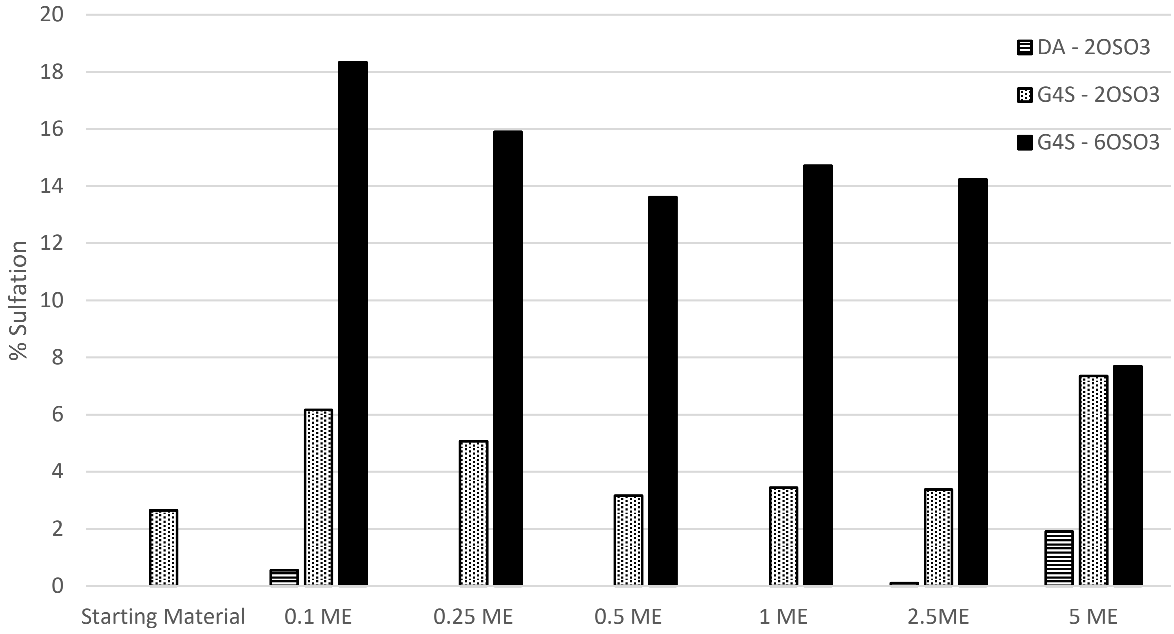


Figure 50: κ-Carrageenan sulfation, using 5 molar equivalents Pyr·SO_3, in BMImBF$_4$ reaction DMAP optimisation with galactose-4-sulfate positions six, black, and two, spotted, and 3,6-anhdrogalactose position two, striped, sulfate group proportions, derived from integral intensity measurements. G4S indicates the galactose-4-sulfates residues and DA, the 3,6-anhydrogalactose residues within each dimer.

The final reaction condition optimisation component tested was the ionic liquid solvent to be used. Whilst reactions performed in BMImBF$_4$ generated sulfated material during optimisation, a commonly used alternative, BMImCl, was tested for comparison at three Pyr·SO_3 concentrations, 2.5, 25, and 50 molar equivalents, with 0.1 molar equivalents DMAP also present. The spectra collected indicated that there was considerably greater sulfation for the reactions performed in BMImCl. Complete conversion of the galactose-4-sulfate hydroxyl groups and partial sulfation of that at position two of the 3,6-anhydrogalactose residues was evident in BMImCl at 25 and 50 molar equivalent concentrations. This was in contrast to BMImBF$_4$ where only partial conversion of the galactose-4-sulfate position six hydroxyl was evident in the spectra, Appendix 4 Figure 3.

Analysis of the integrals recorded indicated that whilst sulfation had run to completion for galactose-4-sulfate position six after 24 hours with 25 molar equivalents Pyr·SO_3, sulfation at position two of the residue had not. 87% and 97% of the residues had been sulfated at this position when using 25 and 50 molar equivalents Pyr·SO_3 respectively. By comparison the integrals recorded at the same Pyr·SO_3 concentrations from reactions performed in BMImBF$_4$ indicated that galactose-4-sulfate residues only contained 43% and 54% sulfation at positions two and six respectively when using 50 molar equivalents Pyr·SO_3, Figure 51. The same increased degree of sulfation was observed for position two of the 3,6-anhydrogalactose residue between BMImBF$_4$ and BMImCl with 18% of residues, in BMImBF$_4$, and 53%, in BMImCl, sulfated at this position when 25 molar equivalents Pyr·SO_3 were used. Increasing the concentration of Pyr·SO_3 to 50 molar equivalents further increased the proportion of 3,6-anhydrogalactose residue sulfated at position two to 64% in BMImCl, while in $BMImBF_4$ the proportion of sulfated residues fell to 0%. Whilst the degree of sulfation seen in Figure 51 indicated that BMImCl was the optimal solvent to use from the two that were tested, it also serves to highlight the importance of increased $Pyr{\cdot}SO_3$ molar equivalents within the reaction mixture.

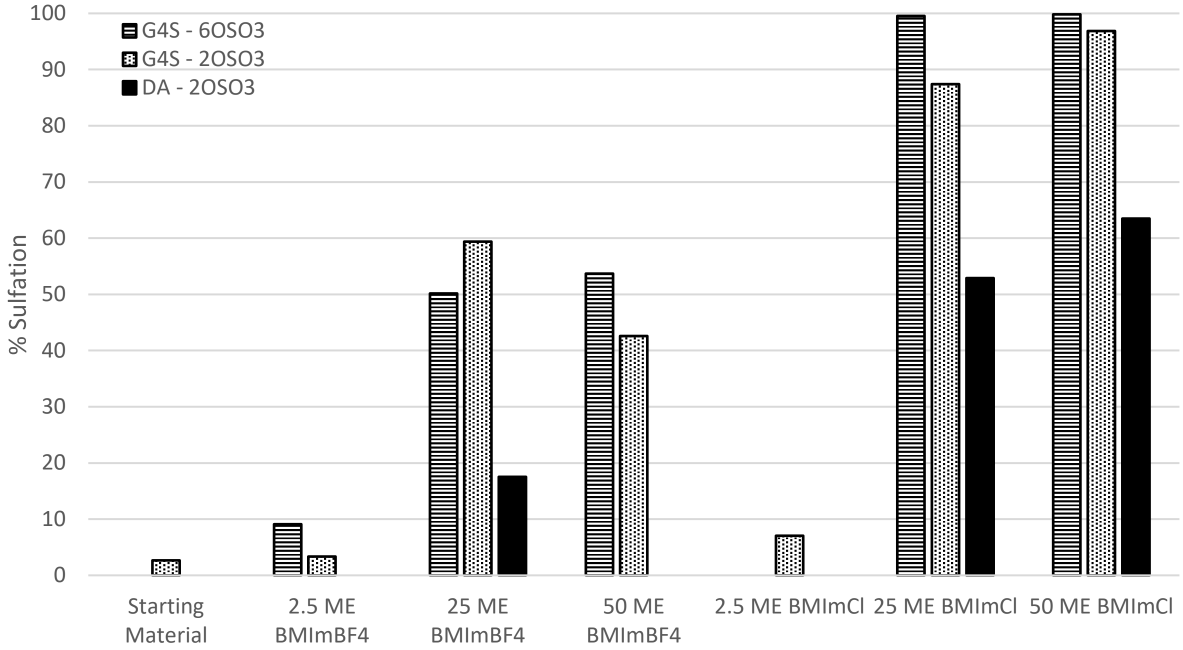


Figure 51: Comparison of κ-Carrageenan sulfation reactions in $BMImBF_4$ and BMImCl with three concentrations of $Pyr{\cdot}SO_3$ used for each; these being 2.5, 25, and 50 molar equivalents.

Optimisation of reaction conditions for the sulfation of κ-carrageenan in ionic liquids, and in the absence of traditional organic solvents, identified the optimal composition of a reaction mixture that could be used for subsequent modification reactions. These optimised conditions were composed of the $Pyr{\cdot}SO_3$ sulfating agent present at 50 molar equivalents and the DMAP catalyst present at 0.1 molar equivalents compared to the carbohydrate's repeating unit disaccharide in the reaction solvent BMImCl. In order to determine the applicability of this method across multiple carbohydrates, and the order in which their hydroxyl groups are sulfated, and hence their reactivity in BMImCl, a series of time course reactions were performed for the plant-based carbohydrates, cellulose, xyloglucan, and κ- and ι-carrageenans.

5.2.2 – Sulfation of Gel Forming and Insoluble Polysaccharides in 1-Butyl-3-Methylimidazolium Chloride

Having generated a series of optimised conditions for 24 hour κ-carrageenan sulfation reactions, these were applied as a time course reaction to four plant-based carbohydrates, cellulose, xyloglucan, and κ- and ι-carrageenans, which represented instances of varying branching and initial degree of sulfation. The spectra recorded on the thirty minutes and one, two, three, four, six, and twenty-four hours timepoints, Figure 52, for the κ-carrageenan sulfation reaction, Figure 46, showed partial sulfation of the galactose-4-sulfate position six hydroxyl group to an increasing degree between thirty minutes and three hours, albeit owing to a error, believed to have occurred during sample modification or extraction, there was no carbohydrate present in the one-hour time point. From the four, six and twenty-four-hour timepoints it was seen that the sulfation at this location had run to completion. A similar pattern was observed for the position two hydroxyls of both this galactose-4-sulfate and 3,6-anhydrogalactose with a gradually increased degree of sulfation up to the four-hour timepoint at which point the reaction had run to completion, with no further changes observed.

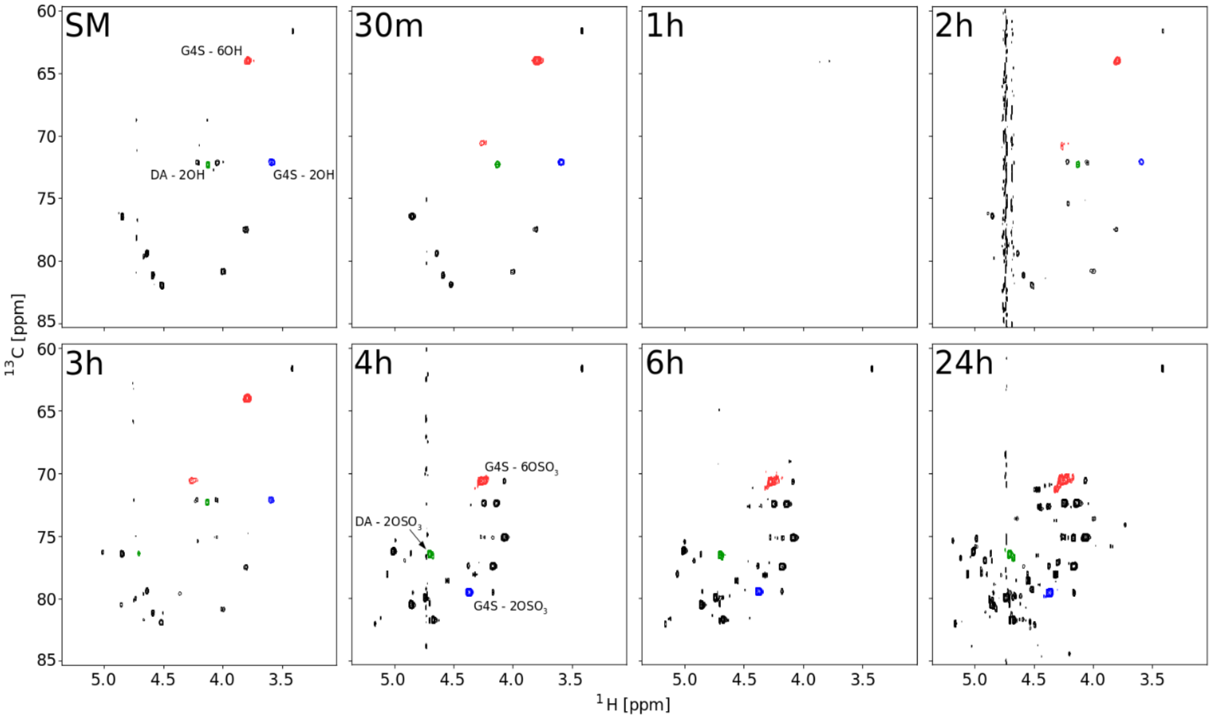


Figure 52: κ-Carrageenan time course, using the optimised conditions in BMImCl, reaction spectra for the starting material and thirty minute as well as one, two, three, four, six, and twenty-four hour timepoints with the galactose-4-sulfate (G4S) positions six, red, and two, blue, and 3,6-anhdrogalactose (DA) position two, green, hydroxyl and sulfated hydroxyl group peaks indicated. As previously explained, the lack of signals in a region, here 1h, is due to sample loss, most likely during sample extraction and washing.

The observations made from the spectra recorded, Figure 52, for both the galactose-4-sulfate hydroxyl groups were supported by integral intensity data generated from these spectra, Figure 53. Here it was seen that the position six hydroxyl increased from being sulfated in 0% of the residues to 19% after three hours with a subsequent dramatic increase to 99% sulfation after four hours, which further increased to 100% after 24 hours. However, the integrals did show a fall in the proportion of sulfated residues between two and three hours, going from 27% to 19% in a manner which was not expected. Similarly, the galactose-4-sulfate position two hydroxyl showed a gradual increase to its degree of sulfation between the starting material and the three-hour time point, reaching 26% before increasing dramatically to 96% sulfation after 4 hours. Whilst the same dramatic increase was seen for the 3,6-anhydrogalactose residue between the three, 7% sulfation, and four, 66% sulfation, hour timepoints, the integrals recorded suggest this is the highest degree of sulfation reached. The contrast between the integrals recorded for both the sulfated and hydroxyl group peaks from this residue position and what can be observed in the spectra, Appendix 4 Figure 4, relates to the presence of higher intensity peaks which flanked the unsulfated position two hydroxyl group peak position. These overlapping peaks distorted the integral intensity recorded for the position two hydroxyl group, and consequently reduced the proportion of residues reported as being sulfated at this position. It is believed that this position was fully sulfated and that this would have been the result reported had it not been for the overlapping peaks in the measurement region.

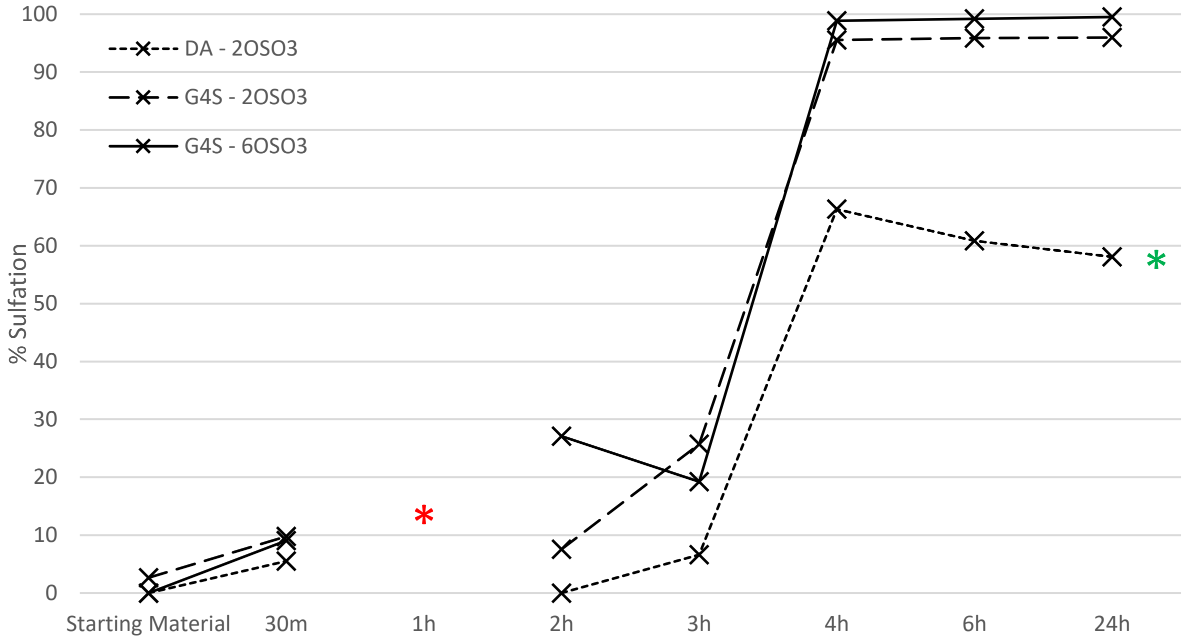


Figure 53: κ-Carrageenan in BMImCl time course reaction for the sulfation of 3,6-anhydrogalactose position two, small dashes, and galactose-4-sulfate position two, large dashes, and six, complete line, hydroxyl groups. The red star for the one hour timepoint indicates an absence of peaks within this spectrum while the green star after the 3,6-anhydrogalactose residue trace indicates that the integrals for this hydroxyl group have been be affected by an overlapping signal contributing to the unsulfated hydroxyl group measurement.

As was seen for κ-carrageenan, sulfation reactions performed on ι-carrageenan, Figure 46, indicated partial sulfation of the galactose-4-sulfate position six hydroxyl group after thirty minutes. However, rather than the sulfated variant having been present at a low level until the four hour timepoint where it dramatically increased to be present in nearly all instances of this residue, as was the case for κ-carrageenan, in ι-carrageenan peaks for both the sulfated and unsulfated hydroxyl groups remain throughout the time course, Figure 54. Evident in the spectra recorded was an increase in the degree of sulfation based on the observable intensities of the two peaks relative to each other. Similarly, for ι-carrageenan the peak representative of the unsulfated galactose-4-sulfate position two hydroxyl had all but completely diminished by the 30 minute timepoint with an observable peak for the sulfated hydroxyl group present in the expected position which increased in intensity throughout the timepoints measured.

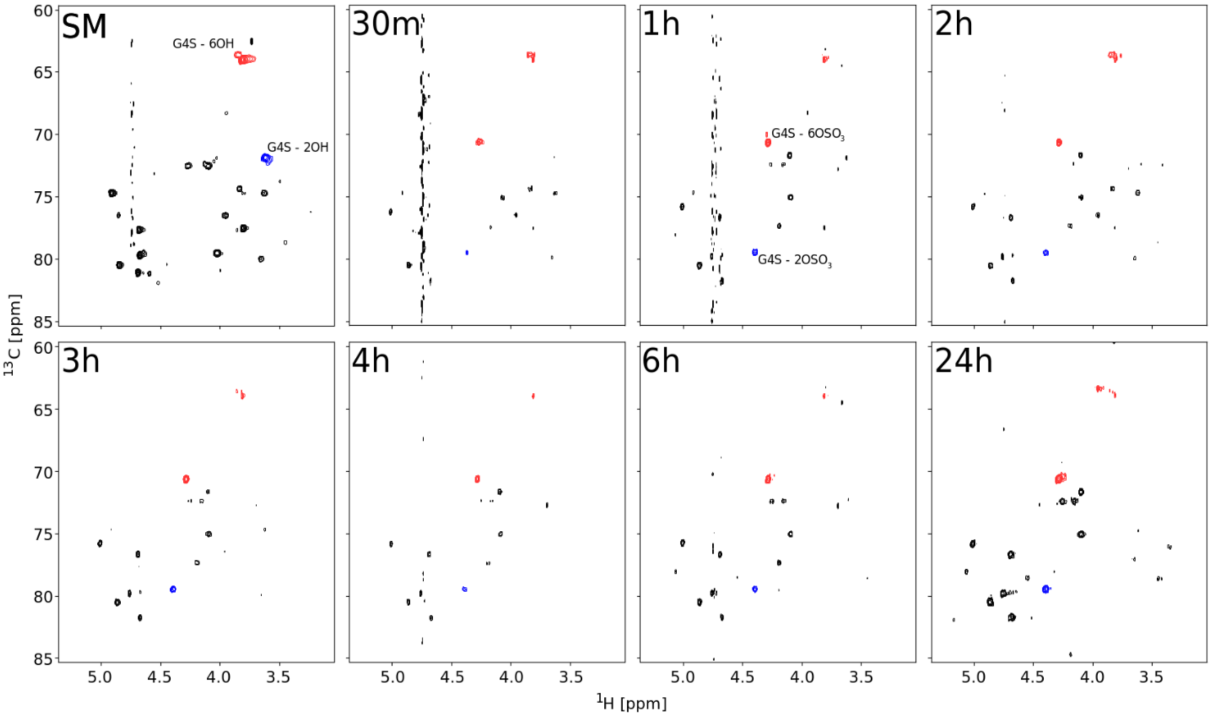


Figure 54: ι-Carrageenan time course, using the optimised conditions in BMImCl, reaction spectra for the starting material and thirty minute as well as one, two, three, four, six, and twenty-four hour timepoints with the galactose-4-sulfate (G4S) positions six, red, and two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.

Integral intensity analysis, Figure 55, supported these observations with the proportion of sulfated hydroxyl groups at galactose-4-sulfate position six increasing to 49% after thirty minutes followed by a generally increased degree of sulfation until the four hour time point, with 76% sulfation, before it declined slightly to 70% after twenty-four hours. This is matched by position two of the same residue which increased to be 87% sulfated after thirty minutes before the sulfation level fluctuated between this and 61%, after four hours, and subsequently increased to be completely sulfated after 24 hours.

Some conclusions regarding the order of reactivity for the different hydroxyl groups present across the two carrageenans were made from the data collected. These included that both the κ-carrageenan galactose-4-sulfate 2-O- and 6-O-hydroxyl groups were modified simultaneously in BMImCl. These groups therefore showed no observable difference in their relative reactivity, in contrast to the expected preference for the primary 6-O-hydroxyl group. The same however could not be concluded for the ι-carrageenan galactose-4-sulfate position two and six hydroxyl groups, where position two appeared to be sulfated more readily across a shorter time frame, again in contrast to the expected order of reactivity given one is a primary hydroxyl and the other a secondary. Similar conclusions have been drawn on the relative reactivity of these two positions between the two starting materials in BMImCl as those present in ι-carrageenan reached a higher degree of sulfation over a far shorter time frame. After thirty minutes 87% of the position two and 51% of the position six hydroxyl groups in ι-carrageenan were sulfated compared to 10% and 9% respectively for κ-carrageenan. However, the near total sulfation of these positions occurred far quicker in κ- than ι-carrageenan, with 96% and 99% sulfation of the κ-carrageenan galactose-4-sulfate position two and six hydroxyl groups after four hours. In comparison, the ι-carrageenan position two hydroxyl group reached a similar level of sulfation after twenty-four hours and the highest degree of sulfation at position six was 76%. The slow initial rate of κ-carrageenan modification, up to 3 hours, followed by a sudden switch to having been fully sulfated appeared atypical and, while the cause remained unknown, possible explanations could have included incomplete dissolution of the carbohydrate samples during the initial time points. An alternative explanation could have been that the gel forming nature of this carbohydrate prevented access to the sites of modification, however, as ι-carrageenan is also gel forming and was highly sulfated within an hour this is unlikely to be the case.

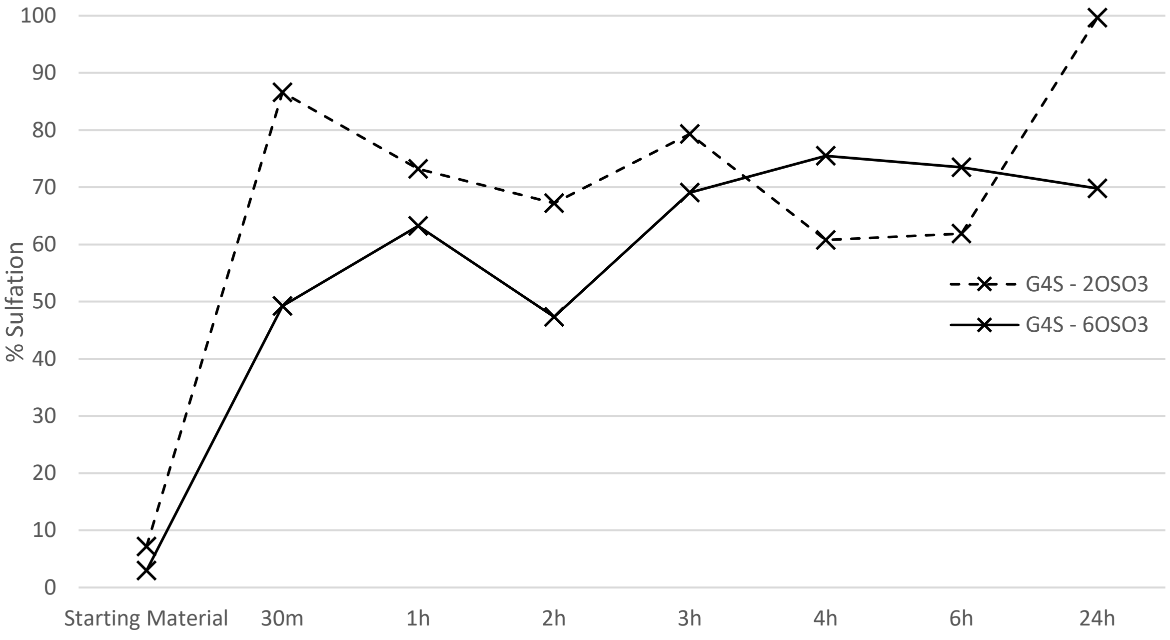


Figure 55: ι-Carrageenan in BMImCl time course reaction for the sulfation of galactose-4-sulfate position two, dashes, and six, complete line, hydroxyl groups.

The determinations made for the relative reactivities of 3,6-anhydrogalactose 2-O-sulfation compared to 2-O- and 6-O-sulfation of galactose-4-sulfate in κ-carrageenan samples was affected by the presence of an overlapping peak in the 3,6-anhydrogalactose 2-OH measurement area. This therefore affected the integral intensities recorded and from these the proportion of sulfated hydroxyl groups, as the reaction had apparently reached completion at 65% sulfation. However, analysis of the observable peaks present in the spectra recorded indicated near total sulfation of this position after four hours, in a manner which closely tracked the sulfation of the galactose-4-sulfate hydroxyl groups and would have indicated the same relative reactivity.

Application of the optimised conditions for Pyr·SO_3 carbohydrate sulfation in BMImCl on non-sulfated starting material polysaccharides such as cellulose, led to its modification, Figure 46. Unsulfated and insoluble in the traditional organic solvents used for sulfation reactions, cellulose provided a prime candidate to investigate the viability of this reaction. Cellulose proved to be soluble in the ionic liquid, 1-Butyl-3-Methylimidazolium Chloride, used, which evidenced this solvent's suitability for the modification of traditionally insoluble carbohydrates. That cellulose modification had occurred was evident due to the presence of signals in a spectra recorded, measured on the polysaccharide. This modification occurred 30 minutes, hereafter denoted 'm', after solid cellulose powder was added to BMImCl, suggesting 6-O-sulfation alone was sufficient to convert the cellulose polymers into a water soluble state. From the spectra recorded, Figure 56, partial sulfation of the position six hydroxyl group was evident at the thirty minute timepoint. A lack of observable sulfation at either of the position two or three, secondary, hydroxyl groups suggested the expected order of reactivity, where the primary, position six, hydroxyl would have been modified preferentially to the secondary hydroxyl groups, held true for cellulose sulfation in BMImCl.

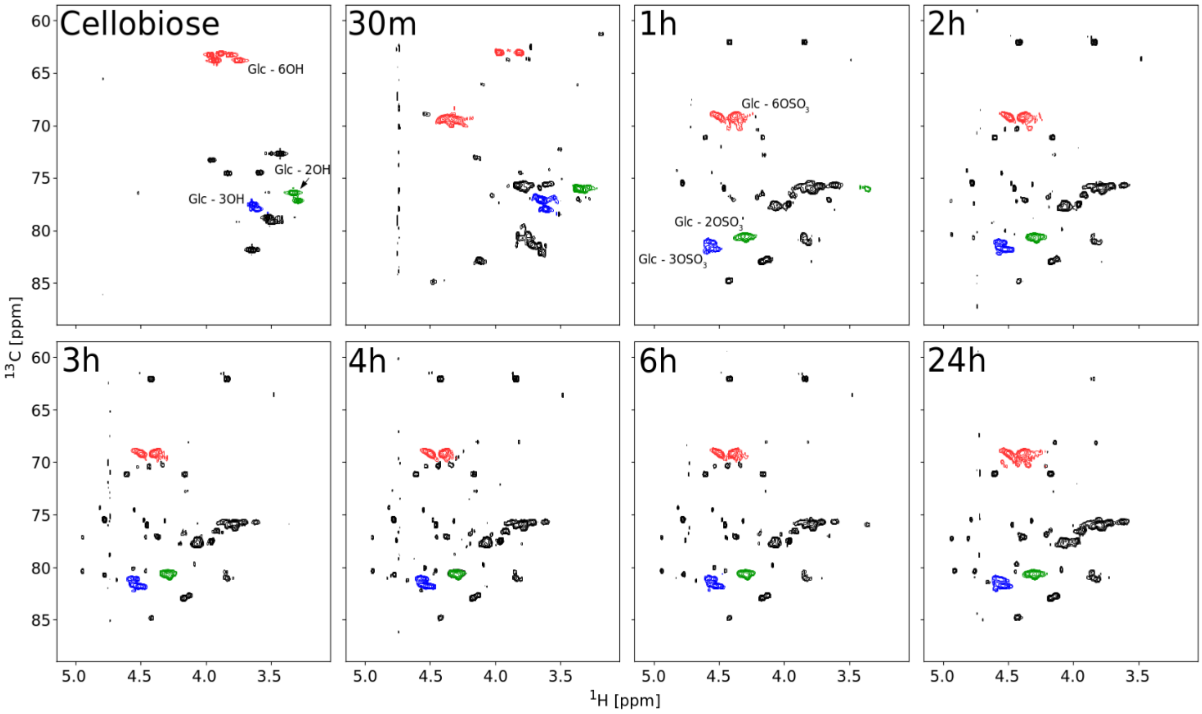


Figure 56: Cellulose time course, using the optimised conditions in BMImCl, reaction spectra with the glucose (Glc) positions six, red, and three, blue, and two, green, hydroxyl and sulfated hydroxyl group peaks indicated. Polysaccharide cellulose's constituent disaccharide, cellobiose, was used in this figure owing to the polymer's insolubility in water.

The conformity observed with the expected order of reactivity was further highlighted in the one hour timepoint where the position six hydroxyl group was fully sulfated as were both the position two and three hydroxyl groups simultaneously, again as would have been expected for two equatorial secondary hydroxyls. The spectra for the remaining time points, Figure 56, showed the pattern of peaks as was present after one hour, which indicated both that sulfation had run to completion after one hour and that no desulfation was exhibited with prolonged heating in the ionic liquids.

Integral intensity values measured, **Error! Reference source not found.**, corroborate the visible evidence of sulfation present in the spectra recorded. From these integrals it was determined that after a thirty minute reaction time 65% of the position six hydroxyl groups had been sulfated, in contrast to 1% and 2% sulfation of the position two and three hydroxyl groups respectively. The clear preference for the primary hydroxyl group observed from the spectra recorded, Figure 56, thus having been corroborated quantitatively. Similarly, the dramatic increase in sulfated position two and three hydroxyl groups between thirty minutes and one hour was evident from the integrals recorded, with 84% and 86% of these residues sulfated respectively. This again corroborated the visual analysis of the spectra recorded and allowed for the conclusion that neither was modified preferentially to the other.

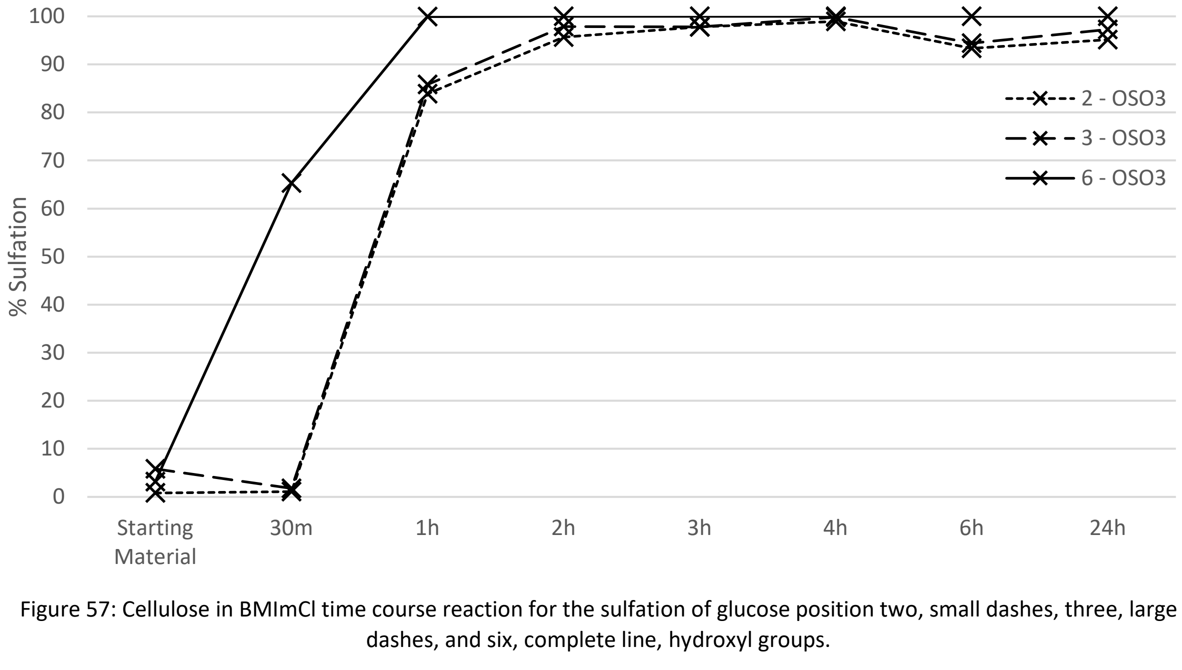


Figure 57: Cellulose in BMImCl time course reaction for the sulfation of glucose position two, small dashes, three, large dashes, and six, complete line, hydroxyl groups.

Integrals recorded for sulfated position six showed that it had been totally sulfated after an hour whilst those for sulfated position two and three hydroxyl groups indicated that they reached their greatest degree of sulfation after four hours, with 99% and 100% of residues sulfated respectively. This, therefore, suggested a one-two reaction with the first stage being the modification of the primary hydroxyl group, which was followed by modification of the secondary hydroxyls as the second stage of the modification reaction. This could be seen to follow the expected order of hydroxyl group reactivity where the primary hydroxyl group would be modified first, and the equatorial secondary hydroxyls then modified at the same rate.

The demonstration that carbohydrates other than carrageenans, could be modified using the optimised reaction conditions for Pyr·SO_3 sulfation in BMImCl, suggested an attempt be made to modify the branched hemicellulose, xyloglucan, Figure 46. Whilst prior sulfation of xyloglucan had not been attempted, and consequently the assignments of sulfated hydroxyl group positions were unknown, it was determined from a visual analysis of the spectra that sulfation had been successful at a number of locations owing to the characteristic appearance of peaks shifted downfield in both the ^{1}H and ^{13}C dimensions, Figure 58.

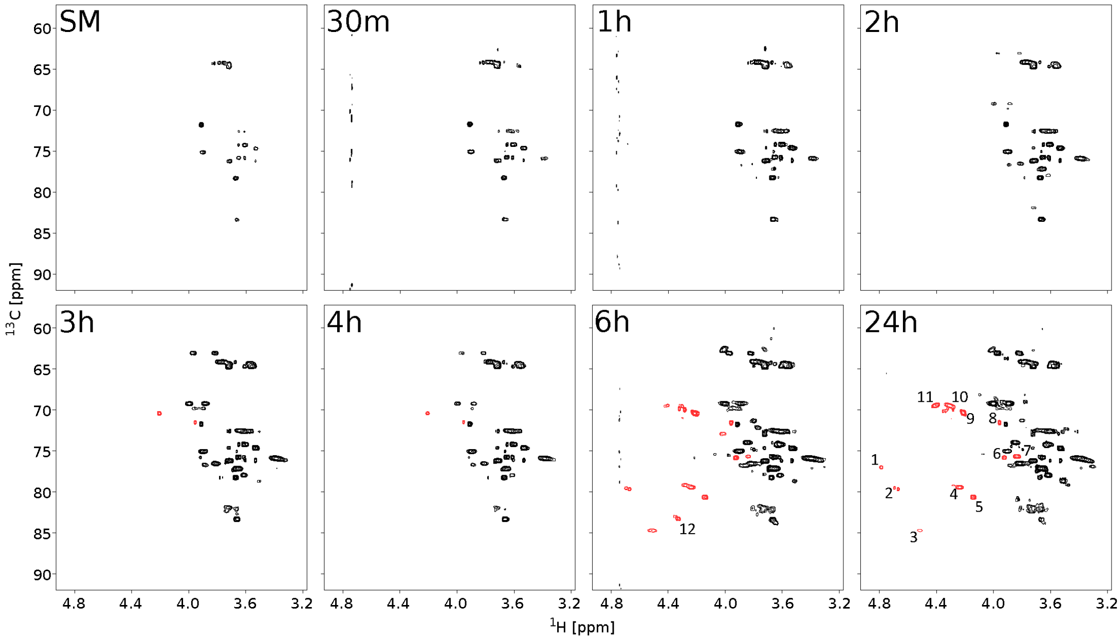


Figure 58: Xyloglucan time course, using the optimised conditions in BMImCl, reaction spectra for the starting material and thirty minute as well as one, two, three, four, six, and twenty-four hour timepoints with the newly emerging sulfated hydroxyl group peaks indicated in red.

From a visual examination of the spectra measured it was observed that sulfation had not occurred significantly until the three hour timepoint, when two novel peaks emerged, denoted 8 – at 3.956ppm/71.50ppm – and 9 – at 4.211ppm/70.36ppm which remained the sole novel peaks present at the four hour timepoint. Both the spectra for the six and twenty-four hour timepoints exhibited a considerable number of novel peaks compared to earlier samples, with peaks denoted, 2 – at 4.668ppm/79.63ppm, 3 – 4.517ppm/84.66ppm, 4 – 4.236ppm/79.42ppm, 5 – 4.141ppm/ 80.60ppm, 6 – 3.925ppm/75.80ppm, 7 – 3.834ppm/75.67ppm, 10 – 4.306ppm/69.46ppm, 11 – 4.404ppm/69.46ppm, and 12 – 4.334ppm/83.18ppm, each newly identified in the six hour timepoint spectrum and that denoted 1 – 4.778ppm/77.85ppm in the twenty-four hour timepoint spectrum. However, peak 12, which was clearly observable in the six hour timepoint spectrum, was wholly absent at twenty four hours.

Analysis of the integral values recorded highlighted that peak 12 was not the only signal to decline in intensity between the six hour and twenty-four hour time point spectra, Figure 59, although it was one of the more extreme. Peaks 2, 3, 5, 6, 7, and 9 also generated lower integral intensities in their twenty-four hour sample compared to that at six hours, while peak 5 generated approximately the same value. This indicated that, unlike cellulose, certain sites of sulfation were prone to subsequent desulfation over extended periods of time, whilst others, peaks 1, 4, 8, 10, and 11 are more stable over a twenty-four hour period.

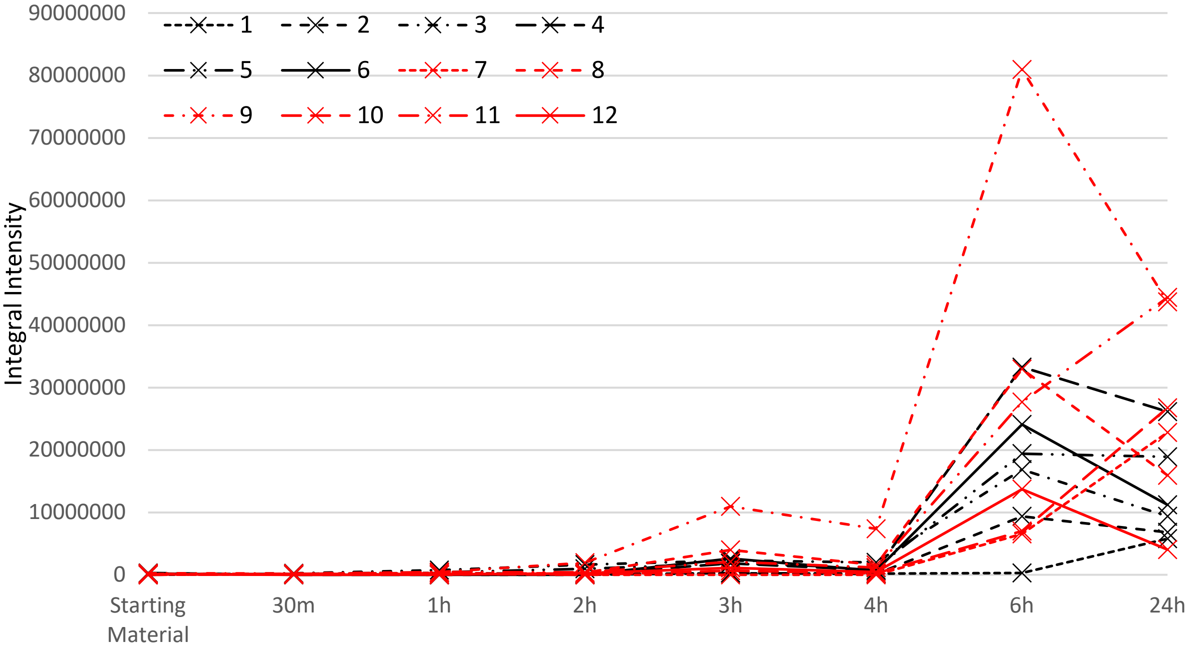


Figure 59: Xyloglucan in BMImCl time course sulfation reaction with peaks 1, black small dashes, 2, black medium dashes, 3, black medium dashes and single spot, 4, black large dashes, 5, black large dashes and double spot, 6, black complete line, 7, red small dashes, 8, red medium dashes, 9, red medium dashes and single spot, 10, red large dashes, 11, red large dashes and double spot, and 12, red complete line, indicated.

Time course reactions performed on four plant-based carbohydrates generated a series of sulfated samples using the optimised conditions for Pyr·SO3 based sulfation in BMImCl in a manner which suggested transferability across multiple carbohydrates. The four polysaccharides used indicated that these reaction conditions were viable for both charged and neutral carbohydrates, as well as those that were linear or branched. Whilst analysis of the sites of sulfation were possible for cellulose, and κ- and ι-carrageenan, this was not completed for xyloglucan, consequently this could be investigated further should this research be progressed at a later point. Having tested the method developed it required comparison against the traditionally used protocols. Therefore, heparin was sulfated using both the traditional method, modification of the tetrabutylammonium salt with solvation in DMF, and the method optimised above, without conventional organic solvents, using both the sodium and tetrabutylammonium salts to identify whether this preparative step could have affected the sites and degree of sulfation for an already highly sulfated polysaccharide.

5.2.3 – Comparison of the Pyridine Sulfur Trioxide Complex based Sulfation Methods for Multiple Heparin Salts in 1-Butyl-3-Methylimidazolium Chloride or Dimethylformamide

Having defined an optimised series of reaction conditions and performed time courses of these reactions on the plant-based carbohydrates, cellulose, xyloglucan, and κ- and ι-carrageenans, a comparison of this protocol against a traditionally used carbohydrate sulfation method was required. Therefore, the biopharmaceutical carbohydrate, heparin, as its tetrabutylammonium salt, was sulfated using the standard Pyr·SO_3 method in dimethylformamide at 85°C, 55°C, and 35°C, Figure 46. The carbohydrate sulfation method developed, using BMImCl as the solvent, Figure 46, was also performed on heparin, as both its sodium and tetrabutylammonium salts, to enable this comparison.

Tetrabutylammonium heparin sulfation in DMF at 85°C was seen, both via visual examination of the spectra recorded, Appendix 4 Figure 4, and quantitative analysis of the integral intensities measured, Figure 60, to have produced heparin sulfated at position three of both the glucosamine and iduronic acid residues. However, partial – 10% – glucosamine de-N-sulfation was observed after 30 minutes and remained present at a similar level in all samples up to the twenty four hour time point. No further sulfation or desulfation was evident after 30 minutes, hence the reaction was said to have run to completion at this time. As was previously observed during the κ-carrageenan time course, the region that was used to measure integral intensity values for one of the peaks, unsulfated iduronic acid position three, contained a portion of another peak. The presence of this portion of another peak within the measurement area increased the integral intensity measured and resulted in an erroneously low proportion of sulfated iduronic acid position three being reported in Figure 60, (marked with a green star).

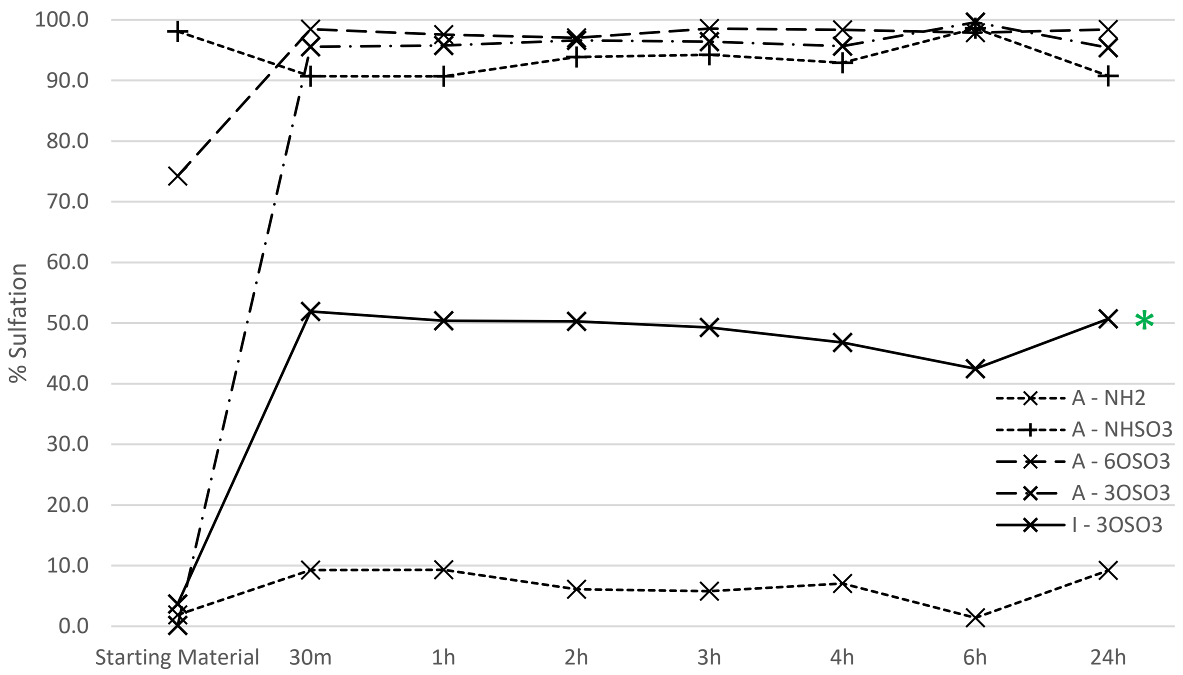


Figure 60: Tetrabutylammonium heparin, 85°C, sulfation in DMF time course with glucosamine position two NH_2 and $N(H)SO_3$ variants indicated by small dashes with X and + markers respectively. Glucosamine $3OSO_3$ and $6OSO_3$ variants are indicated by large dashes with an X marker and large dashes with a dot in between and an X marker respectively. The iduronic acid $3OSO_3$ variant is indicated by a solid line with an X marker, and the green star on this line indicates that this was affected by a contribution from the unsulfated hydroxyl group measurement of an overlapping signal.

The same modifications of glucosamine and iduronic acid position three hydroxyl groups, seen in the 85°C reaction, were identified in the thirty minute time point of the sulfation reaction performed at 55°C. However, following both a visual examination of the spectra collected, Appendix 4 Figure 5, and quantitative analysis of the integral intensities measured, Figure 61, differences between these reactions were evident. During the 55°C reaction, whilst the most obvious shift seen for the sulfation of glucosamine position three was that of the peak for position three itself, a minor shift was also observed for the peak representing the $N(H)SO_3$ group present at position two of this residue. Whilst this glucosamine position two peak shifting upon glucosamine position three sulfation had also been seen during the higher temperature reaction, the reaction performed at 55°C indicated a greater proportion of desulfated position two, NH_2, groups were formed. This desulfation resulted in a clearly visible peak in each of the 55°C reaction timepoints, in contrast to the single low intensity peak present in the thirty minute timepoint of the 85°C reaction.

Further to this, it can be seen from the spectra recorded that sulfation of the iduronic acid position three hydroxyl group was only partial at 55°C, with peaks observable for both the sulfated and unsulfated variants in each timepoint. Additionally, the 55°C reaction three hour timepoint showed evidence of glucosamine position six desulfation which was not seen at the higher temperature. Peaks for other atom positions within a repeating unit containing desulfated position six were therefore visible alongside those seen for a persulfated repeating disaccharide. This suggested that the sample produced was non-homogenous with at least two repeating units present, that which was persulfated and that which had been desulfated at position six. The cause of this non-homogeneity within the 55°C reaction's three hour time point was likely that a small proportion of residual water was present, forming sulfuric acid, which partially desulfated the previously modified carbohydrate.

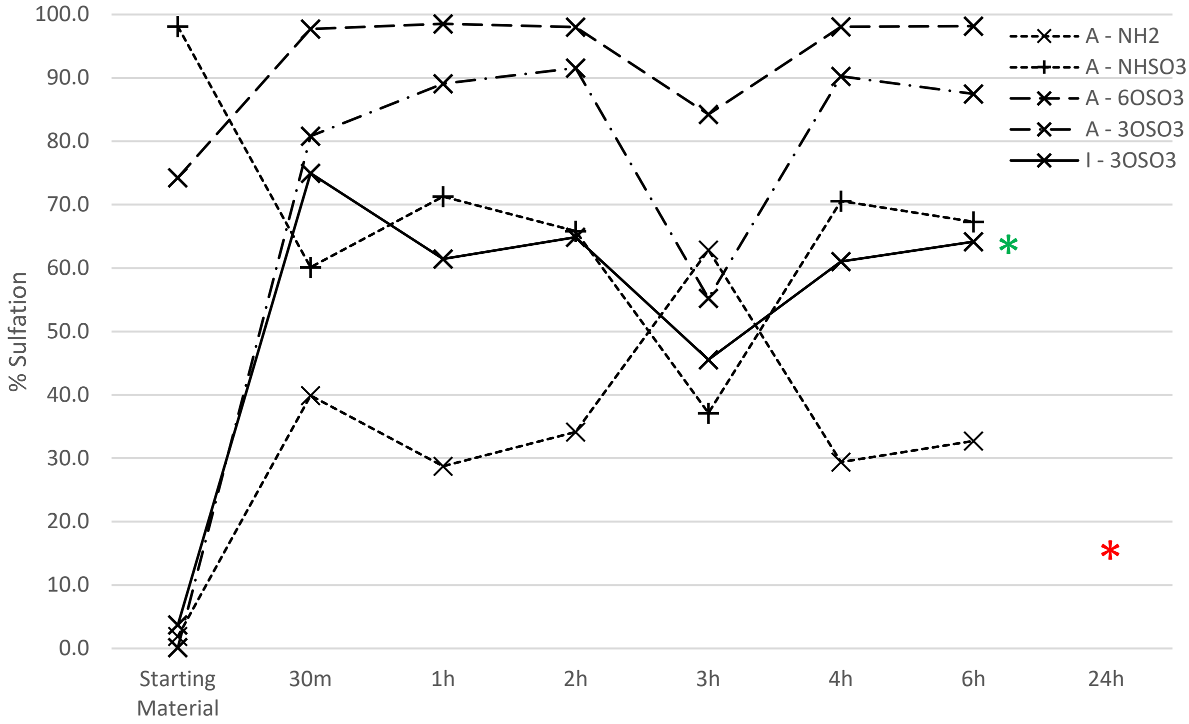


Figure 61: Tetrabutylammonium heparin, 55°C, sulfation in DMF time course with glucosamine position two NH_2 and $N(H)SO_3$ variants indicated by small dashes with X and + markers respectively. Glucosamine $3OSO_3$ and $6OSO_3$ variants are indicated by large dashes with an X marker and large dashes with a dot in between and an X marker respectively. The iduronic acid $3OSO_3$ variant is indicated by a solid line with an X marker, and the green star on this line indicates that this was affected by a contribution from the unsulfated hydroxyl group measurement of an overlapping signal. The red star for the twenty-four hour timepoint indicates the lack of integrals collected owing to an absence of peaks within this spectrum.

The combination of glucosamine position two partial desulfation, across the range of samples measured, and position six desulfation, in the three hour timepoint, generated a less clear cut picture in the integral intensity measurements. As at higher temperature, the 55°C reaction glucosamine positions three and six integrals showed a high proportion of sulfated residues in the thirty minute timepoint, 81% and 100% respectively – comparable to the 96% and 99% sulfation levels from the 85°C reaction. However, the proportion of both of these atom positions that were sulfated dropped considerably in the three hour timepoint, with partial glucosamine position six desulfation observed upon visual examination of the spectrum. A decreased proportion of iduronic acid position three sulfation was also observed within the three hour time point. The reported proportion of sulfated iduronic acid position three was already erroneously low, as has previously been explained, and the reduced integral intensity value recorded for sulfated iduronic acid position three in the three hour time point served to further reduce this proportion. Integral intensity measurements from the glucosamine position two atom variants corroborated the observation made during visual spectrum examination that the proportion of desulfation at this position was greater during the 55°C reaction, where it peaked at 40%, than at 85°C, with a peak of 9%.

Whilst the heparin DMF sulfation reactions performed at 55°C and 85°C generated broadly similar spectra and integral intensities measurements, albeit with certain discrepancies previously highlighted, visual spectrum examination suggested the 35°C reaction differed considerably from both of these, Appendix 4 Figure 6. The thirty minutes, one hour, and two hour spectra, for example, contained a peak for N-sulfated glucosamine present when both the iduronic acid and glucosamine position three hydroxyls are unsulfated. This was present alongside a peak which suggested partial glucosamine de-N-sulfation had occurred, which later became complete position two desulfation between three and four hours. Complete sulfation of the iduronic acid position three and near complete sulfation of the glucosamine position three hydroxyl groups was then observed in the six and twenty-four hour timepoint spectra. This was coupled with complete sulfation of the glucosamine position six hydroxyl and a mixture of sulfated and desulfated glucosamine position two variants, suggesting that this position was initially desulfated before partial resulfation occurred in the later time points. The overall result of this being that the same pattern of peaks was present in the 35°C reaction six hour time point as was observed after thirty minutes for the reactions run at both 55°C and 85°C.

The visual observations were again coupled with a quantitative analysis from the integral intensities measured, Figure 62, which suggested that, as was observed from the spectra, glucosamine position six sulfation remained high throughout the reaction, having risen from 74% in the starting material to 98% after twenty-four hours. This was coupled with iduronic acid and glucosamine position three sulfation which rose from 4% to 74% and 0.1% to 83% respectively between the starting material and twenty-four hour timepoint samples, with a particularly dramatic increase observed after the three hour time point where these residue positions were 25% and 12% sulfated respectively. There was also a spike in the percentage of iduronic residues sulfated at position three within the one hour timepoint, which was evident from the integrals measured, Figure 62. A peak was present in the right area of the one hour timepoint spectrum but absent in either of those for thirty minutes or two hours, thus, matching the integral intensity data. Integral intensities also corroborated the relationship observed during the visual spectrum examination between the N-sulfated and N-desulfated glucosamine variants, however, this proved to be less extreme than was indicated in the spectra, with the sulfated variant never having been completely absent, as the lowest level it reached was 22%. In the twenty-four hour timepoint, the level of sulfation at glucosamine position two had risen to 54%, with the desulfated variant consequently accounting for the remaining 46%.

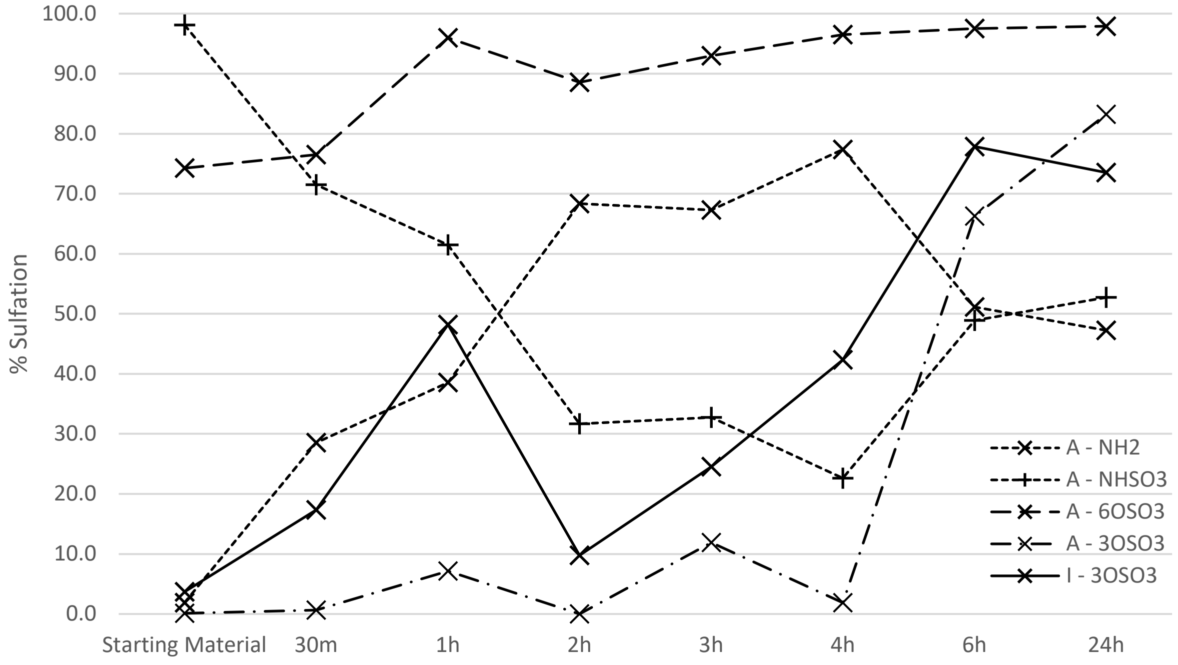


Figure 62: Tetrabutylammonium heparin, 35^{o}C, sulfation in DMF time course with glucosamine position two NH$_2$ and N(H)SO$_3$ variants indicated by small dashes with X and + markers respectively. Glucosamine 3OSO$_3$ and 6OSO$_3$ variants are indicated by large dashes with an X marker and large dashes with a dot in between and an X marker respectively. The iduronic acid 3OSO$_3$ variant is indicated by a solid line with an X marker.

The spectra recorded for the two reactions run in BMImCl, Appendix 4 Figures 7 and 8, suggested that Pyr·SO$_3$ sulfation performed on the tetrabutylammonium heparin salt successfully sulfated both the iduronic acid and glucosamine position three hydroxyl groups after 30 minutes. The spectrum recorded for this time point was observed to have the same pattern of peaks as that recorded for the 55^{o}C reaction run in DMF. There was no evidence of further sulfation or desulfation of this carbohydrate between one and four hours, however, the spectrum recorded for the twenty-four hour timepoint showed near complete glucosamine N- and 6-O-desulfation. The peaks for sulfated iduronic acid and glucosamine position three hydroxyls were also absent in this timepoint, however, while a peak was observed for the desulfated glucosamine, none was present for desulfated iduronic acid therefore it was unclear whether this position had truly been desulfated. This desulfation may also have been present in the six hour timepoint, however, the peaks recorded were of inadequate intensity to make any observations. Pyr·SO$_3$ sulfation of sodium heparin run in BMImCl, however, did not occur. The spectra recorded showed a pattern of peaks identical to that for the starting material until the three hour timepoint, with glucosamine positions two and six then having been desulfated, without evidence of sulfation at either the iduronic acid or glucosamine position three hydroxyls.

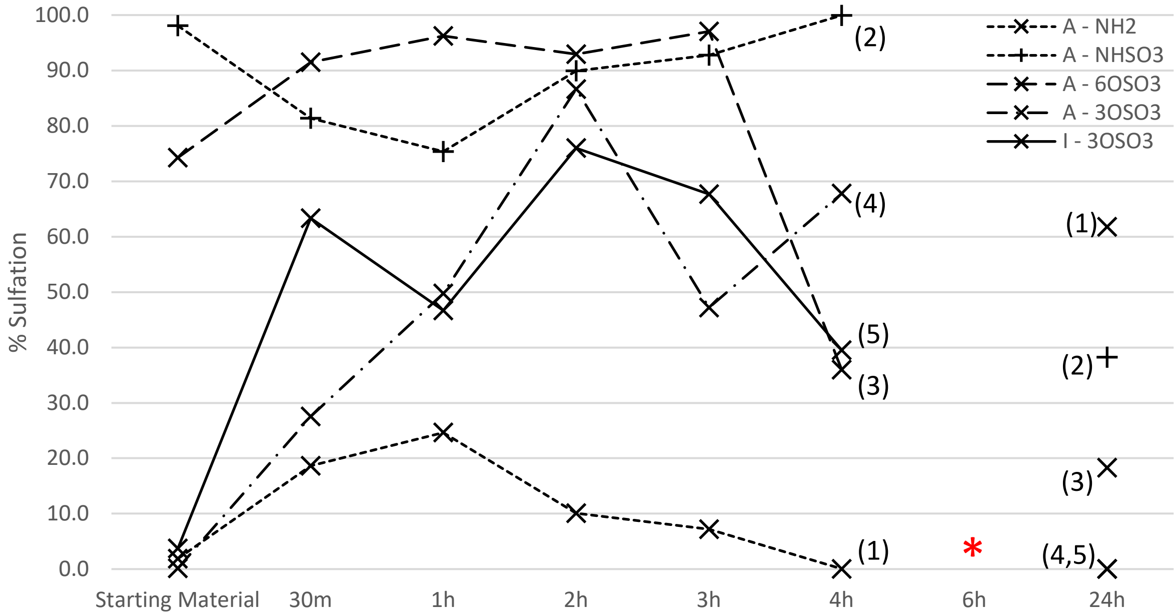


Figure 63: Tetrabutylammonium heparin, 85°C, sulfation in BMImCl time course with glucosamine position two NH_2 and $N(H)SO_3$ variants indicated by small dashes with X and + markers respectively. Glucosamine $3OSO_3$ and $6OSO_3$ variants are indicated by large dashes with an X marker and large dashes with a dot in between and an X marker respectively. The iduronic acid $3OSO_3$ variant is indicated by a solid line with an X marker. The red star for the six hour timepoint indicates the lack of integrals collected owing to an absence of peaks within this spectrum. The numbers in brackets at the twenty four hour time point correspond to those given the same number for the thirty minute to four hour time points.

Analysis of the integral intensity data, Figure 63 and Figure 64, recorded on the two $Pyr{\cdot}SO_3$ sulfation reactions run in BMImCl corroborated the observations made during the visual examination of the spectra collected. The integral intensity values recorded for the tetrabutylammonium heparin reaction showed an initially high proportion of glucosamine sulfated at position six, which peaked at 97% , that then fell drastically to end at 18% sulfated in the twenty four hour time point. Also evident from the integrals recorded was that the proportion of position three sulfated glucosamine and iduronic acid residues increased after thirty minutes, then further increased up to two hours, with 87% and 76% sulfation respectively, before they both fell to having been 0% sulfated after twenty-four hours. The observed high level of sulfated glucosamine position two amine groups was also borne out from the integrals recorded, having risen to 92% after four hours, as was its decline in the twenty-four hour sample, where it was reduced to 38%, with the desulfated variant accounting for the remaining 62%, Figure 63.

The integral intensity values for the sodium heparin salt showed the expected lack of change for any of the relevant atom positions until the four hour timepoint, where the proportion of position two desulfated glucosamine residues rose, from 2% in the three hour sample, to 61%. This was followed by the glucosamine position six desulfation, observed in the spectrum for the twenty-four hour timepoint, where level of sulfation fell to 13%, from 66% in the six hour sample. Also evident from the integrals recorded was the lack of sulfation at either of the position three hydroxyl groups, Figure 64.

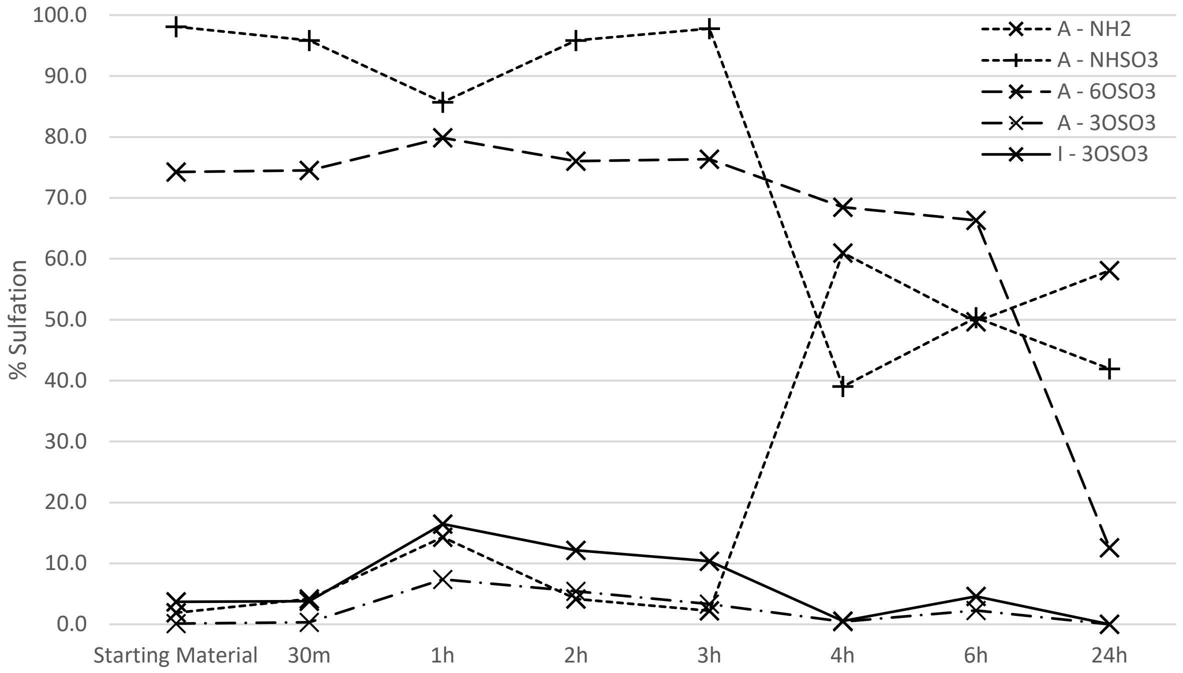


Figure 64: Sodium heparin, 85ºC, sulfation in DMF time course with glucosamine position two NH_2 and $N(H)SO_3$ variants indicated by small dashes with X and + markers respectively. Glucosamine $3OSO_3$ and $6OSO_3$ variants are indicated by large dashes with an X marker and large dashes with a dot in between and an X marker respectively. The iduronic acid $3OSO_3$ variant is indicated by a solid line with an X marker.

Pyr·SO_3 sulfation time courses, run in either DMF or BMImCl, which examined the modification of heparin as either its tetrabutylammonium or sodium salts allowed for the comparison of the two methods. The traditional tetrabutylammonium heparin sulfation reaction methods generated sulfation at both glucosamine and iduronic acid position three atoms at all three temperatures, with increased degree and rate of sulfation seen with increased temperature. However, there were some irregularities within he samples produced, as well as the consequent spectra and integral intensity values recorded, which would require further experimentation to explain conclusively.

As with the traditional heparin sulfation methods, Pyr·SO_3 based tetrabutylammonium heparin sulfation reactions, run in BMImCl, exhibited sulfation of both the glucosamine and iduronic acid position three hydroxyl groups after thirty minutes, before peaking, in the 2 hour time point, at 87% and 76% respectively. However, where these reactions differed from those of the traditional methodology was that these 3-O-sulfated residues were both subsequently 3-O-desulfated with simultaneous glucosamine N- and 6-O-desulfation also having occurred. The Pyr·SO_3 based sodium heparin sulfation reactions, run in BMImCl, were however completely different to those of the traditional sulfation protocol. These experiments failed, as they generated spectra which exhibited no evidence of either sulfation or desulfation during the initial time points and only glucosamine N- and 6-O-desulfation was observed in the later time points. This could have suggested that heparin as its sodium salt was not properly solvent accessible, however, an alternative explanation could relate to the previously reported co-ordination of cations by multiple oxygens within the heparin repeating disaccharides (Hughes et al., 2017). This could have prevented the sodium ions replacement by 1-butyl-3-methylimidazolium, meaning the reaction wasn't possible.

Therefore, the comparison of these two methodologies highlighted two key factors, the first being that the Pyr·SO_3 based sulfation method run in BMImCl is not universally applicable to all carbohydrate cation salts in its present form. The second key factor engendered from this comparison was that, for certain carbohydrates, shorter reaction times are required in order for the modification to be successful. This leads to an overall conclusion therefore, that while the method that was developed functioned well on the range of gel forming and water insoluble carbohydrates initially tested there is considerable further optimisation required until this protocol can be considered universally applicable. This optimisation should also include a detailed analysis of why the tetrabutylammonium heparin samples, that had been sulfated in BMImCl in the early time points, were desulfated in the later reaction time points.

5.2.4 – Fourier Transform Infrared Spectroscopy of Selected Modified Samples

As the NMR spectra and, from these, the integral intensity values recorded indicated that sulfation had been successful for most of the carbohydrates tested, a selected series of samples were further analysed by Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) Spectroscopy to confirm that these samples had been modified. The samples selected for analysis by ATR-FTIR spectroscopy were the heparin, κ-Carrageenan, cellulose, and xyloglucan starting materials as well as their respective twenty-four hour sulfation time points, with that of the 85°C traditional sulfation method, i.e. that in DMF, selected to represent sulfated heparin. NMR analysis had shown these twenty-four time pints for cellulose, κ-carrageenan and heparin to have been fully sulfated, while that for xyloglucan had not. Unlike NMR, FTIR detects sulfate groups directly through characteristic S-O, S=O, O-S-O stretching and bending modes, which also tend to be strong signals.

FTIR spectra for the oversulfated heparin samples exhibited a slight increase in the peak at 1200cm^{-1} indicative of the S=O stretch vibration in comparison to that of the starting material, **Error! Reference source not found.**. In each the three replicates this peak was closer in intensity to the peak for C-C and/or C-O bonds at 1000cm^{-1} than was the case in the starting material. Further, the peak at 750cm^{-1}, representative of S-O stretch vibrations increased relative to the heparin starting material sample, although in the first two replicates this appeared to be covered by a general increase in absorbance values recorded for this region. The limited increase in the absorbance recorded for both the S=O and S-O stretch vibrations between oversulfated heparin and starting material heparin samples was as expected given the highly sulfated nature of the starting material.

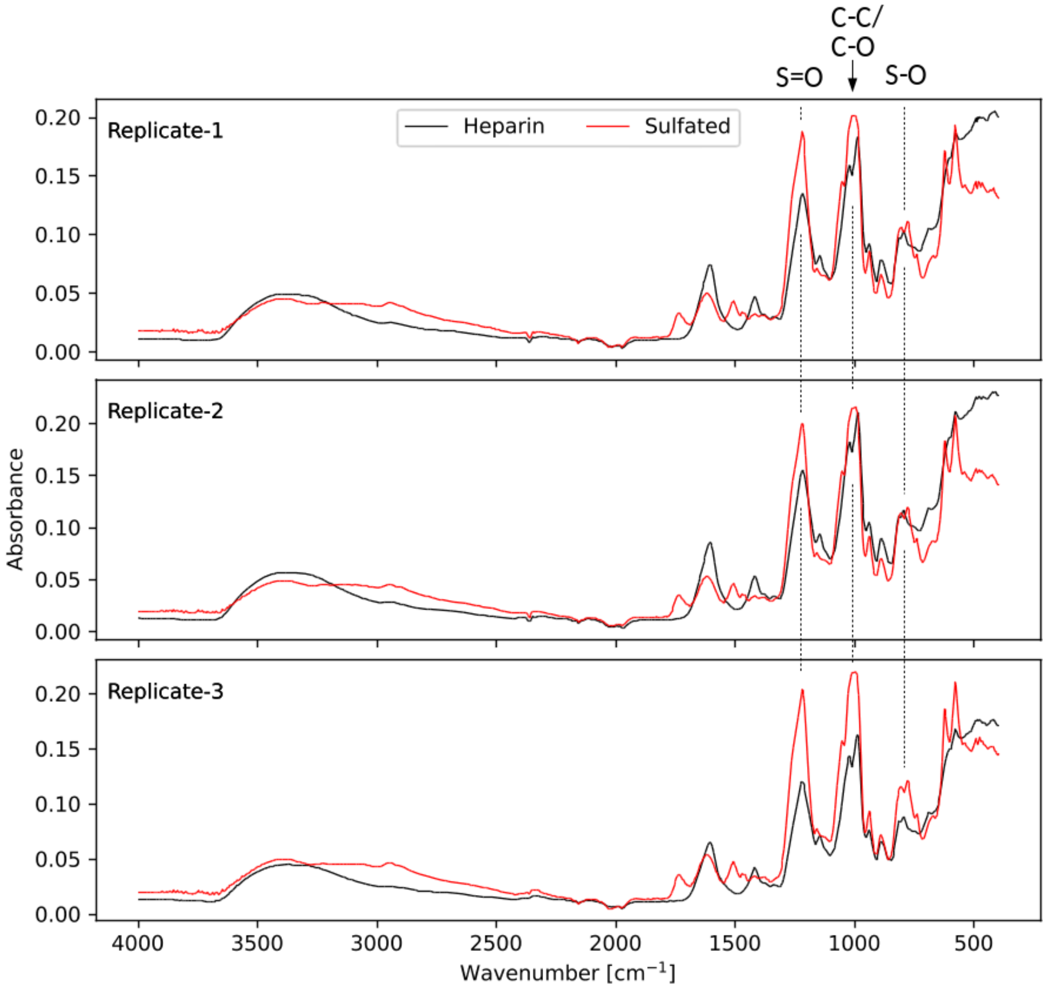


Figure 65: ATR-FTIR spectra for the three heparin (black) and oversulfated heparin (red) samples

In comparison to the modest increases seen for the oversulfated heparin samples compared to the spectra recorded for the starting material, the sulfate peaks for the κ-carrageenan samples increased dramatically at both positions, Figure 66. Despite a general reduction in the absorbance readings for the starting material in comparison to the fully sulfated variants, the peaks present for the S=O and S-O stretch vibrations, at 1200cm^{-1} and 750cm^{-1} respectively, showed a marked increase. These spectra also contained a peak for C-C and/or C-O stretch vibrations, at 1000cm^{-1}, which increased, alongside the sulfate bond peaks, between the starting material and modified samples for each of the three replicates. This increase appears to correspond to a shift in the section of the peak which is present at the highest intensity. In the starting material spectrum the C-C and/or C-O peak is most prominent between 1100cm^{-1} to 1050cm^{-1}, whereas in the sulfated spectra it is most prominent in the region of 1050cm^{-1} to 950cm^{-1}. This could correspond to either the C-C or

C-O stretches increasing in intensity to become prominent when previously the other bond type's stretch vibration gave higher absorbance values. The greater difference in absorbance between these κ-carrageenan, modified and starting material, samples compared with those of heparin, was expected given κ-carrageenan had only one sulfate group in the starting material, increasing to four in the modified sample, whereas the heparin starting material contained three starting material sulfate groups.

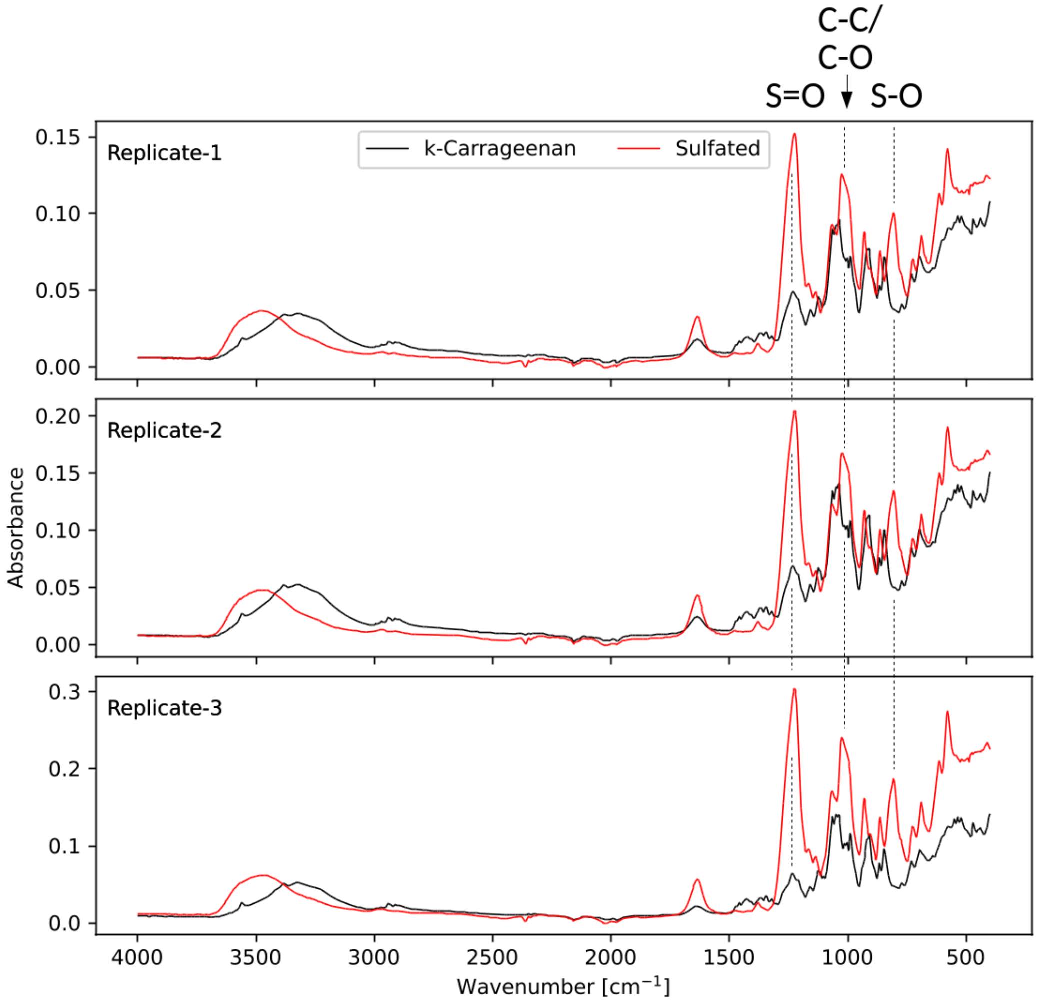


Figure 66: ATR-FTIR spectra for the three κ-carrageenan (black) and sulfated κ-carrageenan (red) samples.

The ATR-FTIR spectra recorded for fully sulfated cellulose samples compared against those of the starting material also showed an increase in the S=O and S-O stretch vibration peaks, with that for S=O having increased from being negligible to being one of the most prominent present, Figure 67. The S-O peak also increased, but to a lesser extent than the S=O peak. Additional peaks could also be observed in the sulfated cellulose spectra at 1500cm^{-1} and 1600cm^{-1} potentially indicative of unidentified functional groups which require further analysis to characterise, including by an expanded range of high resolution NMR techniques. These also could represent more complex vibrational modes potentially resulting from hydrogen bonding and intramolecular interactions, or alterations in the other vibrational modes present. Also present was an increase in the spread of absorbances observed between 3500cm^{-1} and 2500cm^{-1}, likely representative of the hydroxyl groups. Whether these hydroxyl groups were present due to a higher level of residual water in the sample, were representative of cellulose hydroxyl groups which had not been sulfated, or resulted from changes to other, more complex, vibrational modes would also require a further examination using an expanded series of NMR methods performed on oligosaccharide variants of the sample.

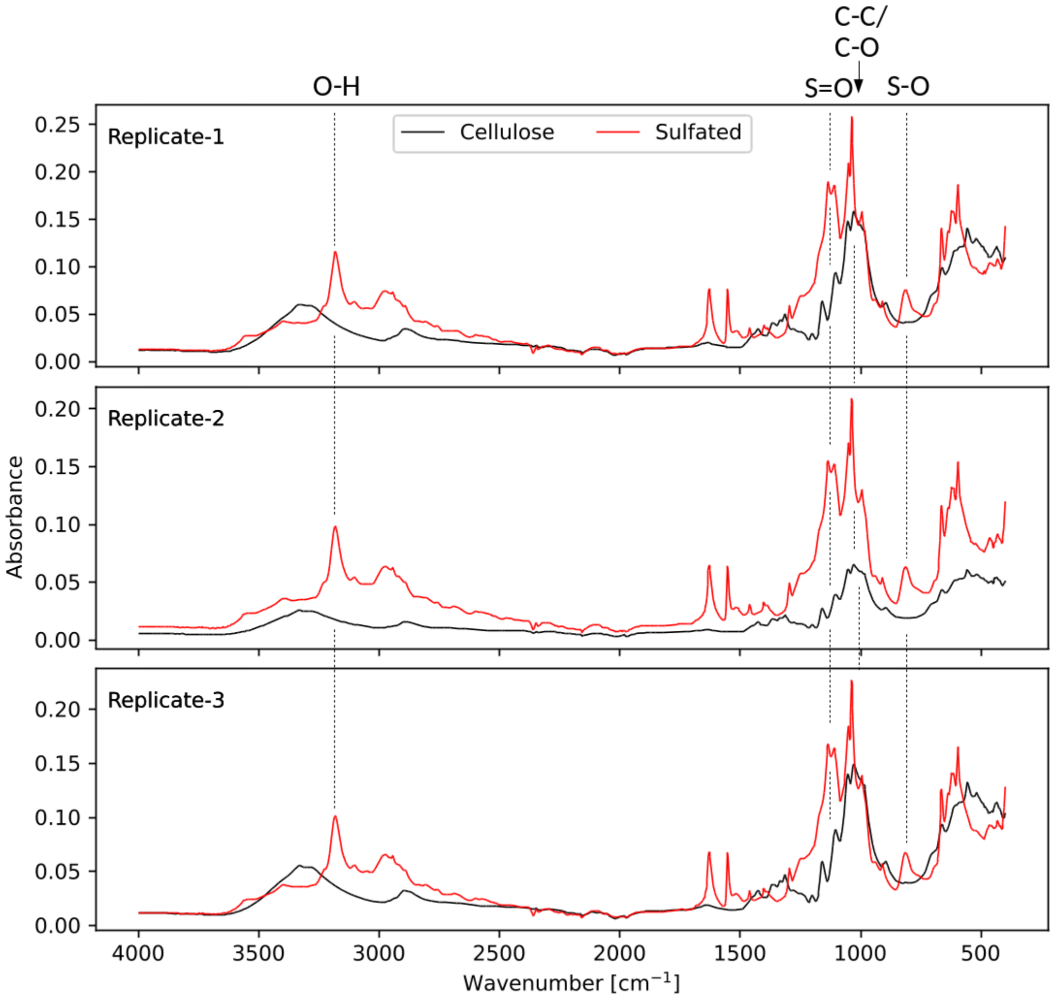


Figure 67: ATR-FTIR spectra for the three cellulose (black) and sulfated cellulose (red) samples.

The final samples examined via ATR-FTIR were those of the xyloglucan starting material and twenty four hour sulfation time point, Figure 68. The spectra recorded contained a peak at 1300cm^{-1} potentially representative of S=O stretch vibrations, however, they did not include a peak at 750cm^{-1} for S-O stretch vibrations. The cause of this discrepancy likely related to the fact that the degree of sulfation was far lower for xyloglucan than the other carbohydrates, coupled with the fact that the S-O peaks were consistently lower than those for S=O across the samples tested by ATR-FTIR. A series of additional peaks were present at 1500cm^{-1} and 1600cm^{-1} in the twenty-four hour sulfation sample as were seen in sulfated cellulose, and further additional peaks were present for the sulfated xyloglucan between 1100cm^{-1} and 400cm^{-1}. As no assignments were made for these peaks, with the exception of that at 1100cm^{-1} being attributed to C-C and/or C-O stretch vibrations, further analysis would be required via additional NMR experiments, potentially alongside alternative techniques such as Raman spectroscopy. Pairing FTIR and NMR was particularly informative as their methods complement each other during functional group identification, such as ATR-FTIR being capable of identifying S=O, S-O, and O-H vibrations which are invisible to NMR.

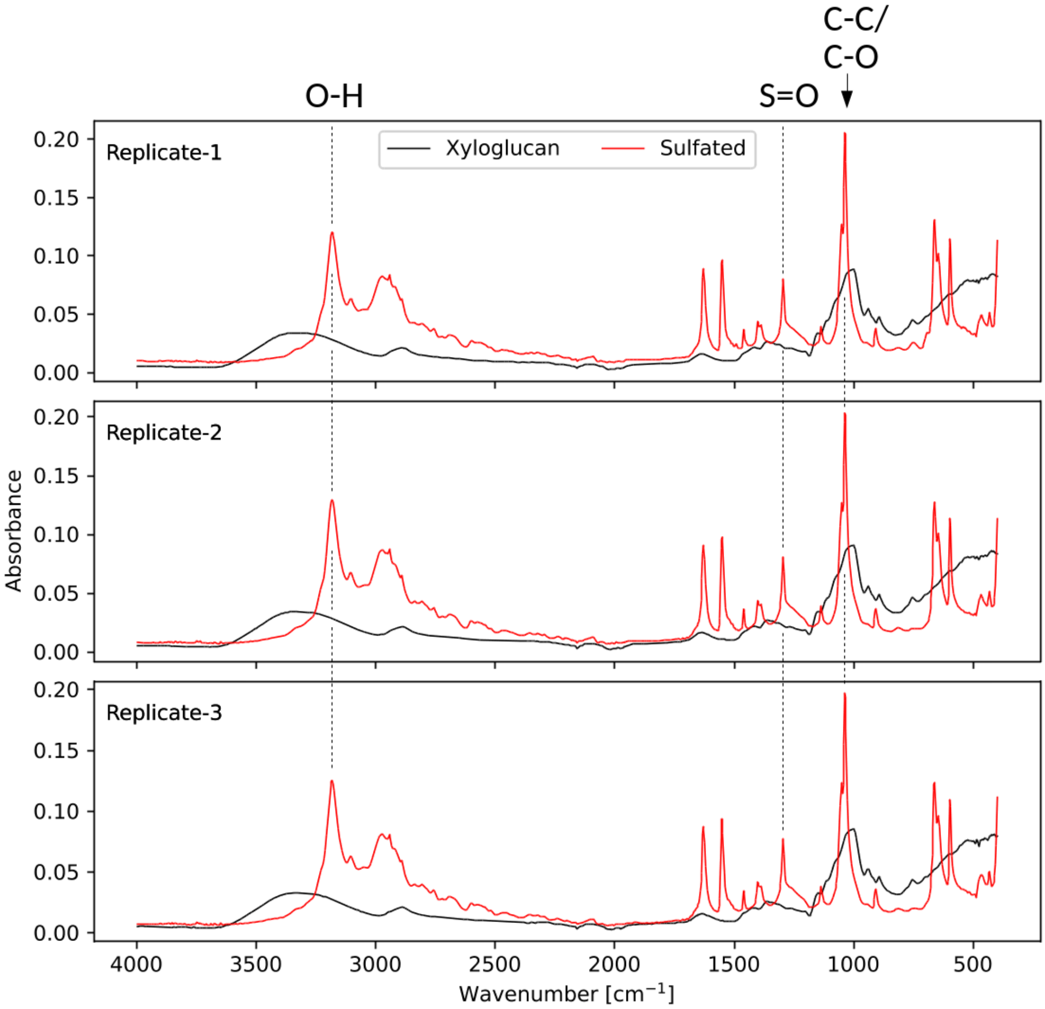


Figure 68: ATR-FTIR spectra for the xyloglucan (black) and twenty four hour xyloglucan sulfation timepoint (red) samples.

The data recorded by ATR-FTIR spectroscopy corroborates that seen by NMR spectroscopy in most cases, with increased S=O and S-O stretches evident for the heparin, κ-carrageenan, and cellulose samples. However, whilst a peak was present for the xyloglucan sample at 1300cm^{-1} which could have been representative of S=O bonds this was not as clear as for the other carbohydrates and consequently this sample would require further analysis by NMR. Other samples could also be subjected to these analysis techniques to provide a series of assignments linked to more specific positions within the structure.

5.2.5 – High Resolution Spectrum Collection and Assignment for Sulfated Samples

Having investigated the sulfation of a series of carbohydrate samples using initial NMR spectroscopy with subsequent corroboration using ATR-FTIR spectroscopy for a select group of these samples, the same select group of samples were used for analysis by the high-resolution NMR protocol developed in chapter four. These experiments were performed in order to generate a series of assignments that were more detailed than those from the screening NMR spectra initially recorded. Additionally, FTIR data analysis highlighted certain inconsistencies that would require further analysis via high resolution NMR spectroscopy, therefore, the samples in which these inconsistencies were observed were also selected for further NMR analysis.

During the first stage of the high resolution NMR analysis protocol the production of oligosaccharide fragments was targeted, with heparinases I and II, κ-carrageenase, cellulase, and xyloglucanase enzymes used for the hydrolysis of heparin, κ-carrageenan, cellulose, and xyloglucan starting materials and twenty four hour sulfation time points respectively. However, of the modified samples, only that for κ-carrageenan could be hydrolysed to the requisite oligosaccharide fragments that would be used for measurement by high resolution NMR. Whilst dependent on the nature of the individual enzymes used, this inability to hydrolyse the twenty four hour sulfation time points indicated sulfation had occurred. The failed hydrolysis most likely resulted from a different carbohydrate structure to that required by the enzyme being present within the active site, with the most likely structural change resulting from the modification reactions performed being hydroxyl group sulfation. Given that the persulfated heparin, cellulose and partially sulfated xyloglucan samples could not be hydrolysed, these samples were not measured using high resolution NMR, while the hydrolysed κ-carrageenan sample was. Analysis of the high resolution spectra recorded on the modified κ-carrageenan sample indicated that the repeating disaccharide structure was not homogeneous throughout the polysaccharide, there being at least two, marginally shifted, peaks present in the spectrum regions corresponding to each atom position, except C-6, throughout the 2-O-sulfated 3,6-anhydrogalactose residue, with the consistent intensity difference present suggesting one major and one minor component, Figure 69.

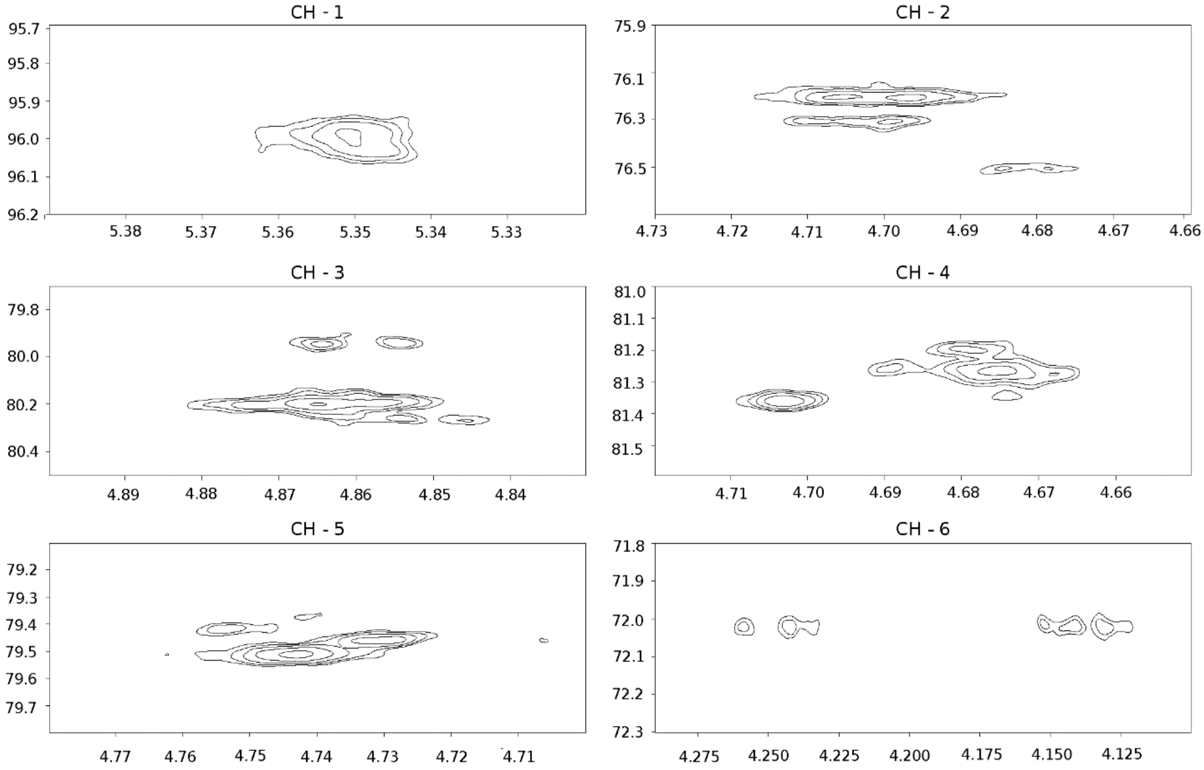


Figure 69: High resolution 2D ^{1}H,^{13}C-HSQC spectrum regions of modified κ-carrageenan with peaks assigned to each the six CH groups in 2-O-sulfated 3,6-anhydrogalactose residues, which were present, at a lower resolution, in the 6h time point from Figure 52.

Similarly, the spectrum regions of the 2-O-, 4-O-, and 6-O-sulfated galactose residues contained at least two peaks at each position, however, for this residue the proportions of the peaks present appeared to be more equal, Figure 70. Four additional peaks are present in the spectrum for κ-carrageenan as well as those shown in Figure 69 and Figure 70 at 5.06ppm/77.68ppm, 4.92ppm/76.65ppm, 4.56ppm/78.19ppm, and 4.32ppm/77.71ppm. These are of a corresponding intensity to the minor peaks seen for both the galactose-4-sulfate and 3,6-anhydrogalactose residues, and are therefore most likely evidence of a second, and possibly third, sulfation pattern present less commonly than that of the major, persulfated repeating disaccharide. These could be assigned by collection of a 2D ^{1}H,^{1}H-COSY spectrum with subsequent data analysis however time did not allow for this. Consequently, this could form an initial point with which to progress the research carried out in this project.

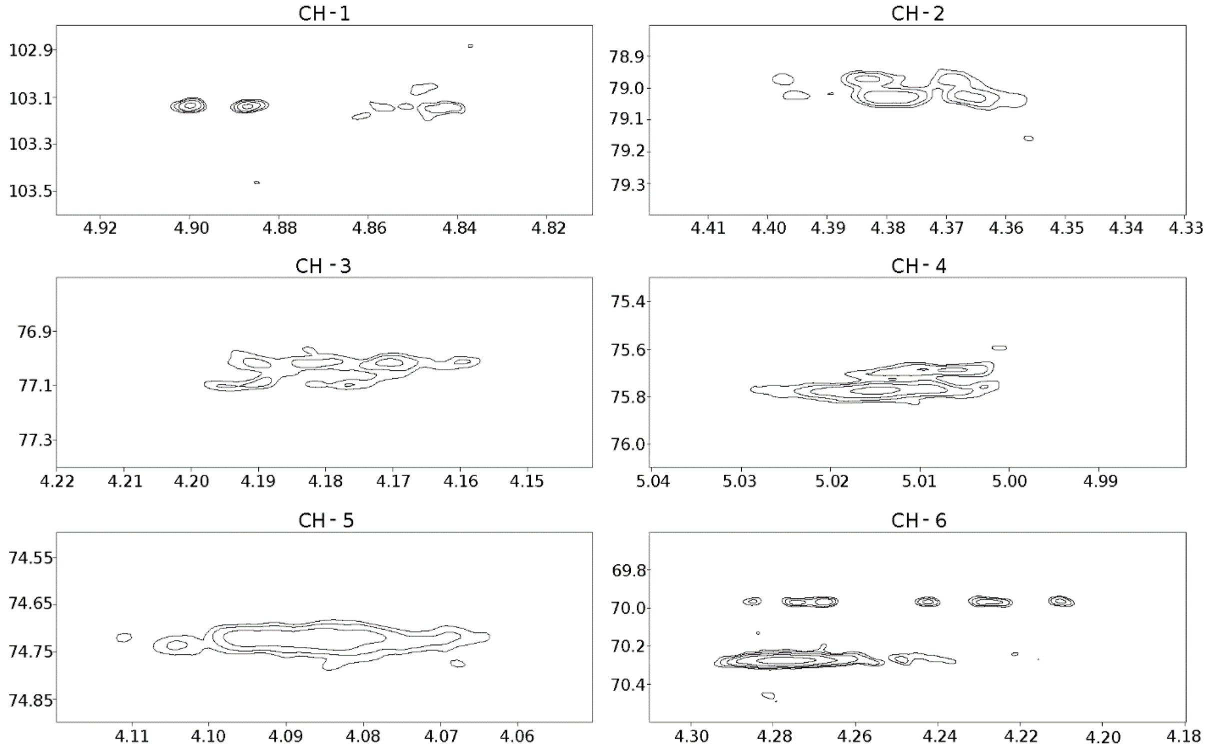


Figure 70: High resolution 2D ^{1}H,^{13}C-HSQC spectrum regions of the modified κ-carrageenan sample with peaks assigned to each of the six CH groups in 2, 4, and 6-O-sulfated β-D-galactose residues, which were present, at a lower resolution, in the 6h time point from Figure 52.

The high resolution NMR spectra collected on the modified κ-carrageenan sample confirmed the conclusion that it was highly sulfated as a result of the pyridine sulfur trioxide sulfation reaction performed in BMImCl. The assignments made for the peaks present within these high resolution spectra match closely with both those of the lower resolution spectra shown in Figure 52 and those of previously published research, which used an alternative – traditional – sulfation protocol (de Araújo et al., 2013). However, multiple peaks were observed for both residue types following hydrolysis to their repeating unit disaccharides. This may have resulted from the polysaccharide having been modified in a non-homogeneous manner with the minor differences between peak shifts masked by the low resolution of the initial screening spectra recorded. Further analyses, centred around the collection and analysis of a 2D ^{1}H,^{1}H-COSY spectrum would therefore be required in order to conclusively determine whether this was the case. Without this spectrum, and potentially others, it couldn't be determined what residue structures the two series of peaks represented. The fact there are no major shifts between the peak locations for each individual atom position precludes one disaccharide structure having sulfate groups present, or absent, at different sites to the other. As the spectra recorded on samples from the previously published research were collected at a lower resolution it cannot be determined whether this heterogeneity was observed, and consequently no comparison between the two sets of results is possible (de Araújo et al., 2013).

5.3 - Discussion

The aim of the research performed in this chapter was to apply the high resolution NMR based structural characterisation technique developed to a series of modified samples. These samples were to be produced using a sulfation method which had been adapted from previous examples using organic solvents in combination with ionic liquids in order to negate some of the problems experienced using traditional sulfation protocols for gel forming and water insoluble carbohydrates. An additional factor in using only BMImCl as the solvent is that it could provide a means by which sulfation patterns other than those expected from the relative order of reactivity could be produced, without resorting to potentially complex protecting group strategies.

The method was adapted from those that use solvent mixtures of ionic liquids in combination with DMF, the traditional organic solvent for sulfation protocols using pyridine sulfur trioxide complexes (Chopin et al., 2015; Gericke et al., 2009; Wang et al., 2007). Eliminating DMF to provide a solvent system composed wholly of BMImCl did enable sulfation of each carbohydrate tested to a greater or lesser degree as seen by screening NMR spectra, as well as by FTIR. The modified carbohydrates produced, when characterised via NMR suggested that cellulose and the two carrageenans had become fully sulfated and NMR assignments matched those of previous sulfation studies (Kowsaka et al., 1991; de Araújo et al., 2013). The assignments made for the heparin samples produced in BMImCl also matched those for persulfation in prior studies (Yates et al., 2000). In this research, 3-O-sulfation increased to its maximum intensity after two hours before declining again. Whilst no assignments could be found for previously sulfated xyloglucan samples, the spectra recorded exhibited the characteristic chemical shift change, moving downfield in both the ^{1}H and ^{13}C dimensions.

Analysis of the order in which the carbohydrates' hydroxyl groups were sulfated suggested that while some followed the order expected given the relative reactivities of primary and secondary hydroxyl groups, others did not. Cellulose, for example, was first sulfated on the primary hydroxyl group, before either of the secondary groups were modified, thus conforming with the expected order of reactivity. However, for both the carrageenan samples, sulfation of the primary hydroxyl group occurred simultaneous to, if not slightly later than, that of the secondary hydroxyl groups, i.e. those at galactose-4-sulfate position two, and 3,6-anhydrogalactose position two for κ-carrageenan. A comparison of the expected sequence of modifications, based off the relative order of hydroxyl group reactivity, versus the experimental results was not made for the heparin samples produced, except that the D-glucosamine and L-iduronate positions three hydroxyl groups were largely sulfated. Similarly, no comparison was made for the expected xyloglucan modification sequence, resulting from the relative order of reactivity, against the experimental results, as comprehensive assignments could not be generated. Repeating the xyloglucan sulfation on a combined repeating unit oligosaccharides sample, to determine the sites of modification and order or reactivity for the hydroxyl groups present would be a key area for progressing this research in a novel manner.

Whilst the method developed for sulfation proved successful, with the persulfation of gel forming carrageenan and water-insoluble cellulose being achieved, there remain elements which could be further developed. The first of these is the continued optimisation of the experimental set-up and reaction mixture composition. Such continued optimisation should include attempting to reduce the proportion of BMImCl within the solvent system. Whilst a combination of ionic liquid and organic solvent, i.e. DMF, was the starting point for optimisation during this research, alternative, less environmentally harmful, solvents could be tested. One such solvent is the biorenewable Dihydrolevoglucosenone, commercially named Cyrene™, which is formed from cellulose and is less environmentally harmful to dispose of than some traditional solvents since no NO_X and SO_X gases are generated during incineration (Sherwood et al., 2014). Whilst this solvent has not yet been used for carbohydrate modification, it has replaced DMF, or alternative organic solvents, in a number of modification methods and could be a viable option to test during carbohydrate sulfation reactions when paired with an ionic liquid (Sherwood et al., 2014; Camp, 2018; Bousfield et al., 2019).

The method developed also requires additional optimisation to reduce the amount of pyridine sulfur trioxide complex present, as the amount being used is far higher than when using organic solvents such as DMF. Whilst a reduced sulfating agent concentration may be afforded by decreasing the proportion of ionic liquid within the solvent mixture this could not be guaranteed, therefore testing to reduce its concentration or replace the pyridine element within the complex with a less environmentally damaging alternative would be important. An alternative to this could be the use of an enzyme compatible ionic liquid to perform sulfation experiments using sulfotransferase enzymes (Moniruzzaman et al., 2010; Zhao et al., 2008). These changes could be combined to generate a more optimised system for the sulfation of polysaccharides with complex repeating units, such as xyloglucan, which in contrast to the three polymers tested with repeating mono- or disaccharides, was not modified on all sites. Finally, when attempting to analyse the modified polysaccharide using high resolution NMR on hydrolysed samples it became apparent that most had not been hydrolysed, therefore running two simultaneous experiments, one of the oligosaccharides and one on the polymer, may be required in future cases.

In conclusion, the method developed proved successful for the modification of heparin, cellulose, and both κ- and ι-carrageenan samples thus providing a protocol by which carbohydrates, particularly those of a gel-forming or water insoluble nature, could be sulfated without prior hydrophobic salt formation steps. Importantly, the method produced samples whose sulfation patterns differed from those obtained with conventional methods, enabling new chemical space to be accessed. Results of both NMR and FTIR analyses indicated successful sulfation, with all of the carbohydrates tested, except xyloglucan, showing evidence of having been persulfated. High resolution NMR experiments performed on κ-carrageenan also indicated a potentially homogeneous repeating structure with some additional peaks requiring further characterisation to determine their assignments, with additional spectra required for a conclusive determination to be made. Despite the success of the techniques developed and used during this research, there remain elements which could be improved, as have been discussed.

Chapter 6 – Protein Interaction Screening of Modified Sulfated Carbohydrates

6.1 – Introduction

Carbohydrate-protein interactions are involved in many biological processes including the signalling pathways that result from Fibroblast Growth Factor (FGF) Receptor complex formation (Onishi et al., 2016; García et al., 2014; Thinakaran et al., 1995; Ornitz and Itoh, 2015; Pellegrini et al., 2000; Schlessinger et al., 2000). The importance of heparan sulfate's interactions with FGFs has long been known, as they have been shown to increase the activity of these FGF signalling complexes (Ornitz and Itoh, 2015). Similar interactions have also been observed with heparin, a highly sulfated variant of the same repeating disaccharide. Of the twenty-two FGF family members the interactions between heparin and FGFs 1 and 2 have been most extensively researched; their, either NMR or x-ray crystallography derived, structures are shown in Figure 71. These structures display the secondary α-helix or β-sheet motifs, formed by the amino acid sequence, alongside the location and orientation of these motifs relative to each other, the consequence of which is the protein surface that enables interaction with other molecules. However, heparin does not interact with the FGF family members alone, with this family representing a small proportion of heparin's protein interactions (Ori et al., 2008; Capila and Linhardt, 2002; Peysselon and Ricard-Blum, 2014; Nunes et al., 2019). An example of an interaction between heparin and a non-FGF family members is that with the Cyclophilin (Cyp) family member Peptidyl-prolyl *cis-trans* isomerase B, hereafter CypB, which has been described during the adhesion of T-lymphocytes (Allain et al., 2002; Hanoulle et al., 2007). Whilst another cyclophilin family member, CypA, has been shown to interact with heparin, little research has been focussed on other members having a high sequence homology to Cyps A or B, such as CypD (Saphire et al., 1999; Capila and Linhardt, 2002). The structures of cyclophilin B and D, are also shown in Figure 71.

As with heparin's interactions with multiple proteins, the FGFs and CypB have been shown to interact with multiple carbohydrates depending, amongst other factors, on their sulfation pattern. FGFs 1 and 2 have the shown to bind other members of the glycosaminoglycan family, notably Chondroitin-4,6-Sulfate, E (Deepa et al., 2002). However a less sulfated mixture of chondroitin-4-sulfate, A, and chondroitin-6-sulfate, C, showed no evidence of interaction with these proteins when tested, likely owing the lower degree of sulfation (Sun et al., 2019; Vessella et al., 2021). Both FGFs 1 and 2 have also been reported to interact with the highly sulfated λ-carrageenan and dextran sulfate oligosaccharides in a manner which stabilises both proteins (Sun et al., 2019), while FGF-2 has been shown to bind the lesser sulfated ι-Carrageenan, in a manner which potentially induced changes to the protein's conformation (Hoffman and Sykes, 1993). Considerably less has been reported on interactions between CypB and carbohydrates other than heparin, other than those with the vaguely termed glycosaminoglycans (Allain et al., 2002; Bukrinsky, 2002; Vanpouille et al., 2007), with none reported for CypD.

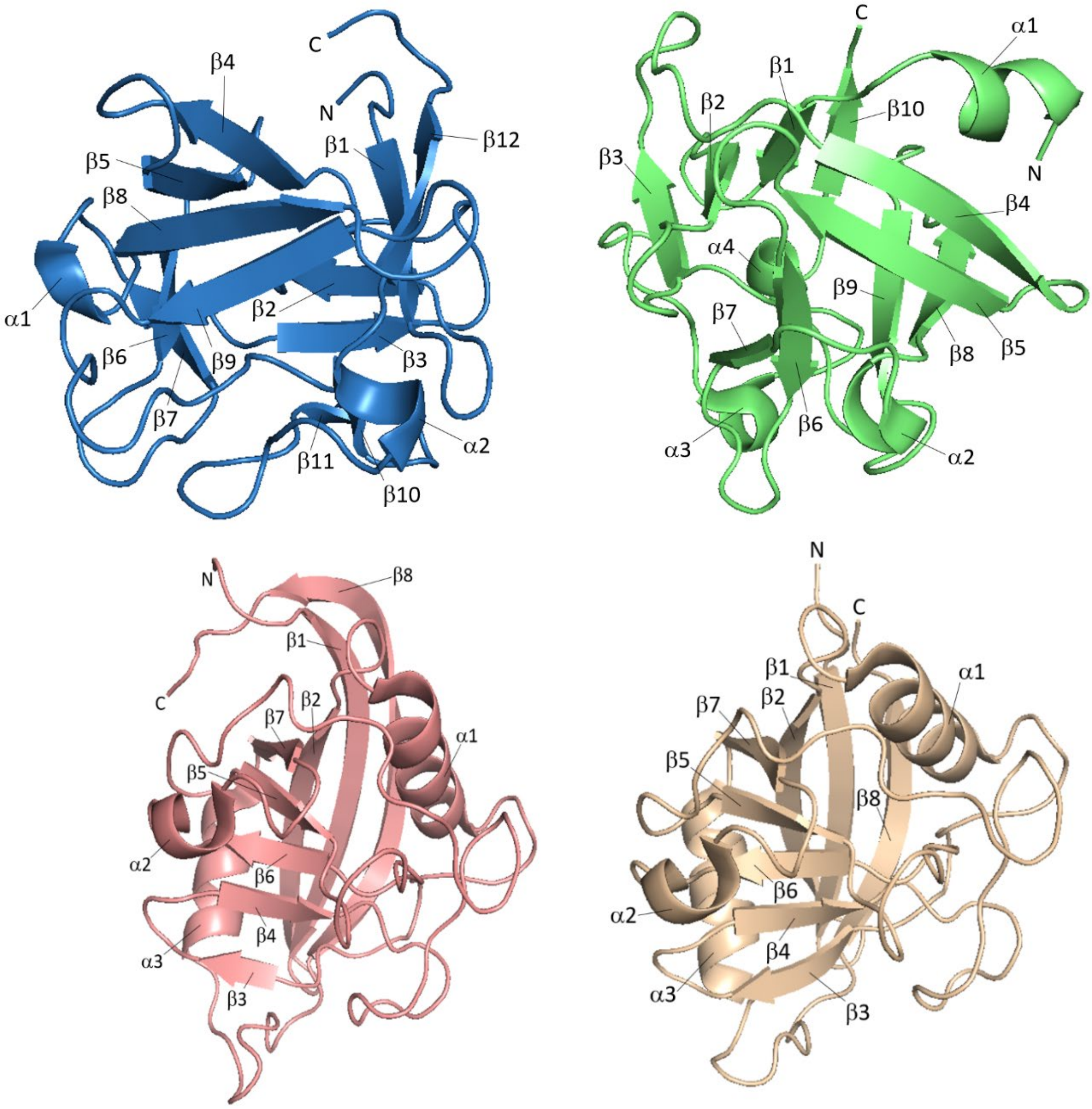


Figure 71: Ribbon diagrams depicting the structures of the four proteins used in this chapter; Fibroblast Growth Factor 1 (blue – PDB:2ERM; Canales et al., 2006), Fibroblast Growth Factor 2 (green – PDB:4OEE; Li et al., 2014), Cyclophilin B (peach – PDB:3ICH; Kozlov et al., 2010), and Cyclophilin D (khaki – PDB:4J5C; Ahmed-Belkacem et al., 2016).

The aim of the research presented in this chapter is to determine, via Differential Scanning Fluorimetry based screening and subsequent Isothermal Titration Calorimetry, whether alterations to the sulfation pattern of four carbohydrate samples affects interactions with FGF-1, FGF-2, CypB, and CypD proteins. These proteins were chosen owing to their biological relevance and the previously reported interactions of heparin and other carbohydrates with FGF-1, FGF-2, and CypB. CypD was included to determine whether the structural homology of the cyclophilin family members equates to interactions with the same carbohydrates.

6.2 – Results

6.2.1 – Protein Expression and Purification

Four plasmids, provided by Professors Dave Fernig and Lu-Yun Lian, encoded proteins either fused to a Tobacco Etch Virus (TEV) protease cleavable N-terminal hexahistidine tag or, for FGF-2, no affinity tag. All four protein constructs were expressed in BL21 STAR^{TM} (DE3) competent cells with consistently increasing optical density readings recorded across repeat expressions prior to induction with IPTG and overnight incubation.

Purification of the four constructs differed depending on the affinity tag present or previously known interacting ability with heparin. CypB and CypD, for example, were both initially purified via immobilised metal affinity chromatography using a nickel ion column, Figure 72, generating high protein yields of considerable, but not complete, purity across a number of elution fractions. The majority of the CypB and CypD proteins isolated by cell lysis appeared within their respective soluble, supernatant, fractions however, the insoluble, pellet, fractions both also generated bands suggesting the presence of target protein.

Nickel ion affinity chromatography based cyclophilin purifications required some optimisation from the previously used protocol to prevent dissociation of nickel ions from the column resin due to the high concentration of DTT in the lysis and purification buffers. However, when reducing the amount of DTT present it was identified that insufficient reducing agent led to the appearance of bands at around 45kDa for both cyclophilins. As a result a final 1mM DTT concentration was used to reduce the proteins' disulfide bonds and prevent dimer formation whilst not affecting the integrity of the column.

Chromatograms and corresponding SDS-PAGE gels for the nickel ion affinity purifications are shown in Figure 72 for cyclophilins B and D respectively. These gels enabled the determination of target protein containing elution fractions, 5-13 for CypB and 7-13 for CypD, indicated by the red boxes, which correspond to the broad peaks present in both chromatograms. During CypB purifications the peak that corresponded to the target protein was dominant within the resulting chromatograms. However, a minor peak was also present representative of the impurities eluted from the column during purification that, in some fractions, co-eluted with the target protein. In contrast, during CypD purifications the peak which corresponded to impurities within the filtered lysate, fractions 4-6, was considerably larger than that for the target protein. There was however greater separation between the fractions containing eluted target protein and impurities for CypD than occurred for CypB, Figure 72.

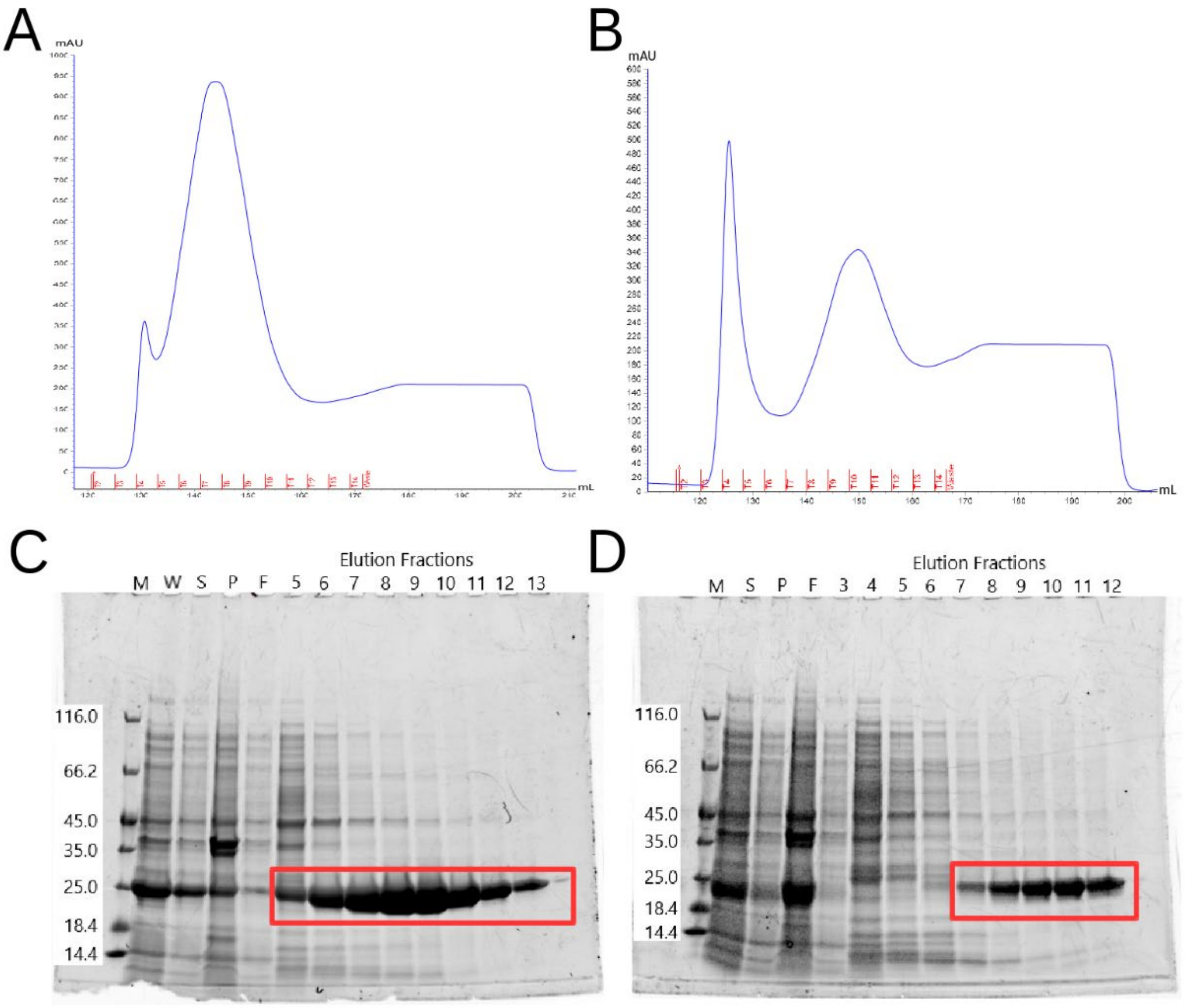


Figure 72: Chromatograms and SDS-PAGE gels generated from the nickel ion affinity based purification of cyclophilins B (A and C) and D (B and D). The target cyclophilin bands are shown within the red box at 25kDa for the appropriate elution fractions. M = Protein Mw Marker, W = Whole Cell Lysate, S = Supernatant, P = Pellet, and F = Flow through while numbered wells indicate the elution fraction.

To further purify the eluted CypB and D fractions size-exclusion chromatography (SEC) was employed. Prior to this, however, the hexahistidine tags were cleaved via overnight incubation with TEV protease and reverse purified across a nickel ion affinity column. These reverse purifications generated samples of improved purity compared to those which resulted from the initial affinity chromatography step. However, the TEV protease cleavage reactions did not run to completion with some uncleaved CypB and D proteins observable on subsequent SDS-PAGE gels and corresponding chromatograms (data not shown). Also present within these elution fractions were the cleaved hexahistidine tag and the majority of the impurities observed during the initial nickel ion chromatography steps. Having produced samples of both CypB and CypD without their N-terminal affinity tag the final stage of purification was the size-dependent separation of target proteins from any remaining impurities. These purifications, shown in Figure 73 for CypsB and D, resulted in the elution of a series of fractions containing solely the target proteins, all remaining impurities having been previously eluted.

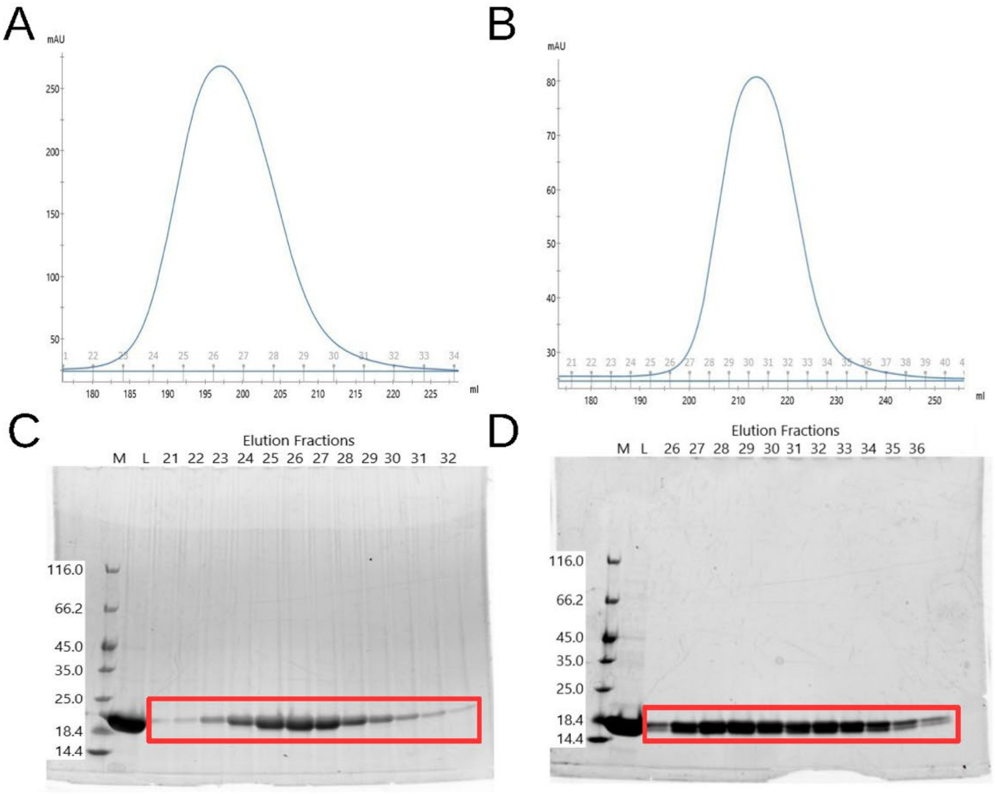


Figure 73: Chromatograms and SDS-PAGE gels generated from the size exclusion-based purification of cyclophilins B (A and C) and D (B and D). The target cyclophilin bands are shown within the red box at 20kDa and 18kDa respectively for the appropriate elution fractions. M = Protein Mw Marker and L = Loaded Sample while numbered wells indicate the elution fraction.

The fractions purified via nickel affinity and size exclusion chromatographies, of demonstrated purity, were combined and concentrated to 100μM samples. Representative samples for both proteins were then measured via 1D ^{1}H NMR spectroscopy in order to test the proteins had correctly folded. The spectra measured for CypB, and CypD are shown in Figure 74 and indicated the samples produced contain folded protein with the peak patterns exactly matching spectra recorded for previously purified instances of both proteins.

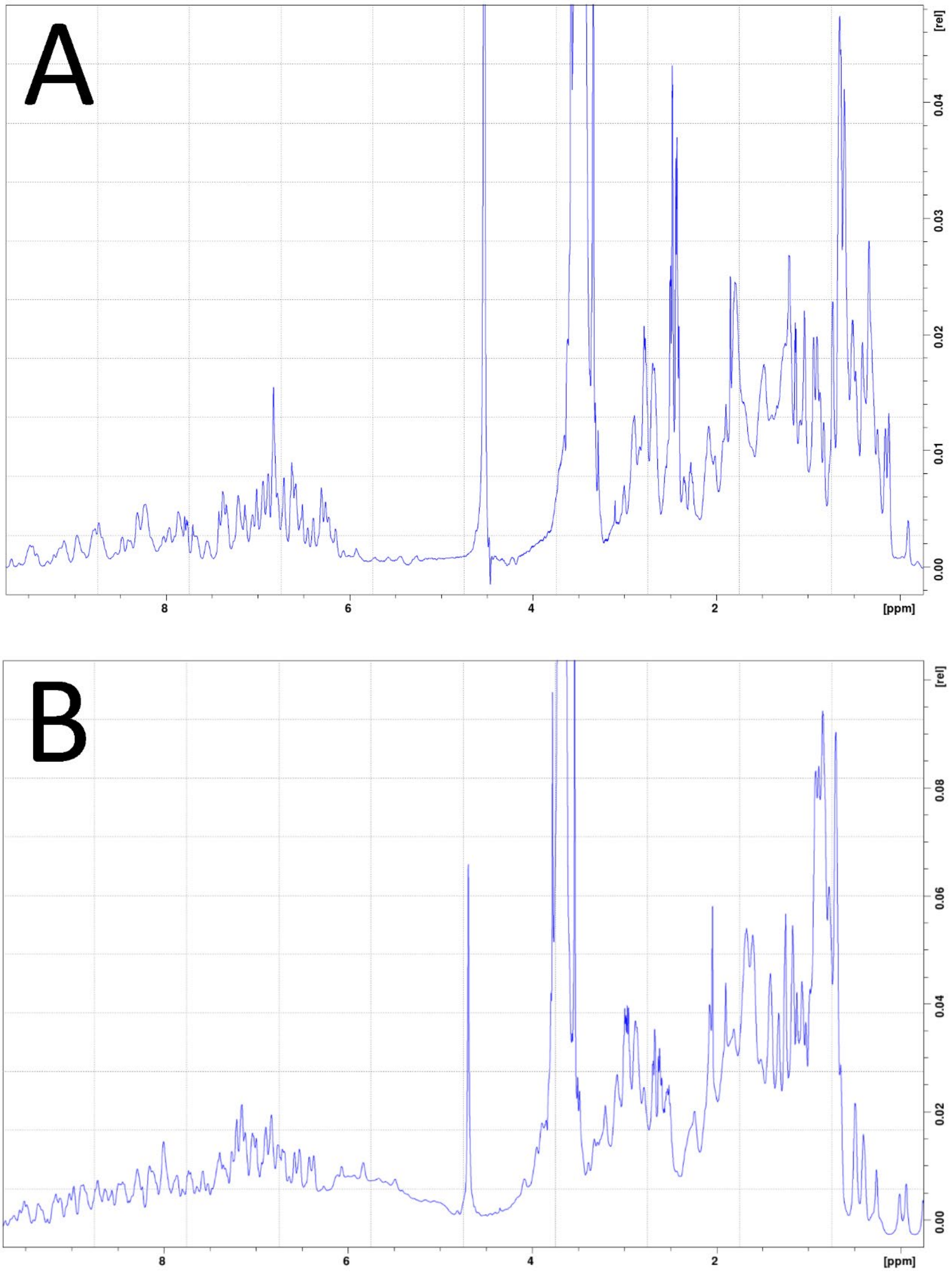


Figure 74: Sections of the 1D ^{1}H NMR spectra of Cyclophilins B (A) and D (B) unlabelled protein samples exhibiting the all of the peaks present within the respective spectra, those regions at the extremes of the ppm range measured, where peaks were not present, have been excluded.

Unlike the Cyclophilin samples, FGFs 1 and 2 were initially purified using a Heparin HiTrap column owing to their previously characterised interacting ability. In both cases affinity chromatography resulted in the elution of a number of moderately pure protein samples as shown in Figure 75. The purity of the proteins produced could have been improved by increasing the initial sodium chloride concentration during the elution gradient, from 150mM, to 600mM. However, this could have compromised the amount of target protein recovered, owing to the increased potential for the protein to be lost into the wash sample. As further purification steps were to be used, it was determined that the potential for loss of protein was an unacceptable risk. While the FGF-1 purification proved successful with a high yield generated, that derived from the FGF-2 purification was considerably lower. Following heparin affinity chromatography the FGF-1 hexahistidine tag was cleaved by TEV protease and removed by a reverse nickel ion affinity chromatography step.

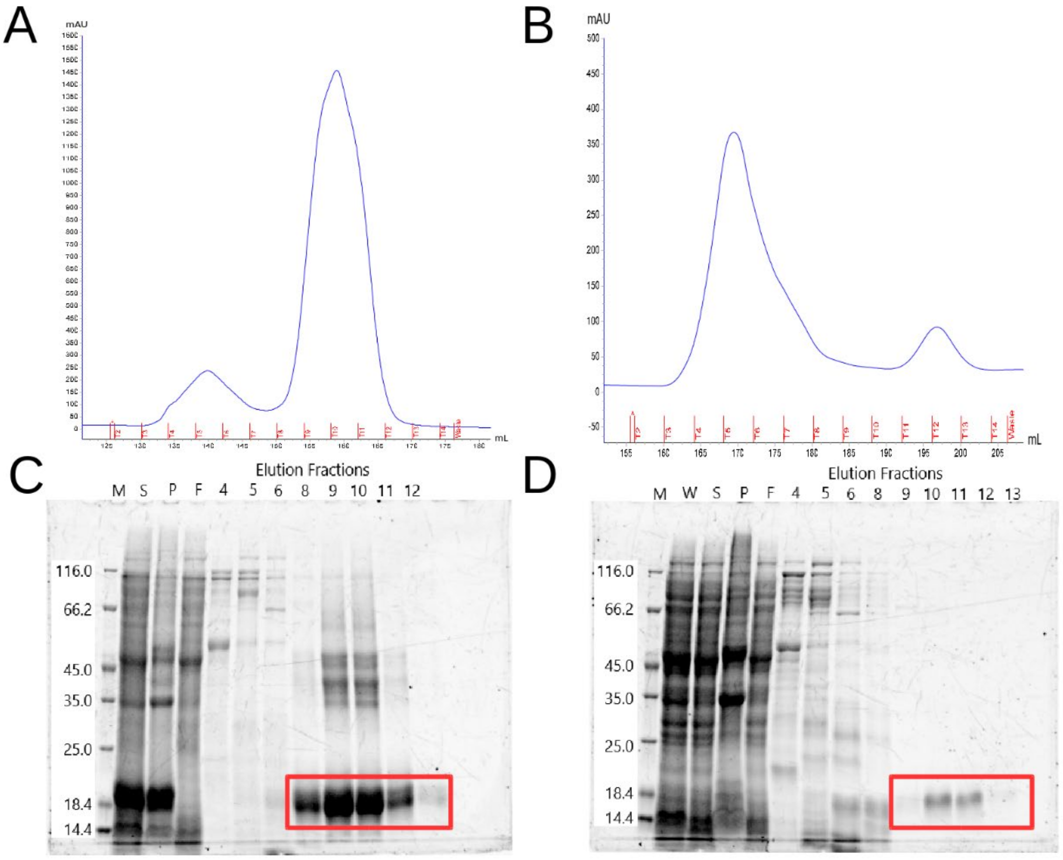


Figure 75: Chromatograms and SDS-PAGE gels generated from the heparin affinity based purification of FGFs 1 (A and C) and 2 (B and D). The target FGF bands are shown within the red box at 18kDa for the appropriate elution fractions. M = Protein Mw Marker, W = Whole Cell Lysate, S = Supernatant, P = Pellet, and F = Flow through while numbered wells indicate the elution fraction.

The purity of the FGF-1 and 2 elution fractions post-affinity chromatography was inadequate for sample production; therefore they were concentrated, combined into one fraction per protein and purified via SEC. These purifications generated a series of eluted fractions of high purity with the SDS-PAGE gel run on FGF-2, indicating no other protein in significant quantities, which is supported by the corresponding chromatogram, Figure 76. In contrast to this, there are a series of weak bands at 35kDa running across the SDS-PAGE gel for FGF-1, also Figure 76. However, given the relative intensity of the two series of bands present and the fact that both proteins had eluted in the same SEC fractions, with the only other peak observable in the chromatogram corresponding to impurities, no further purification steps were added. When measured via 1D ^{1}H NMR spectroscopy the resulting spectrum, Figure 77, matched that which would be expected from a lower molecular weight folded protein with identical peaks to spectra previously recorded for FGF-1.

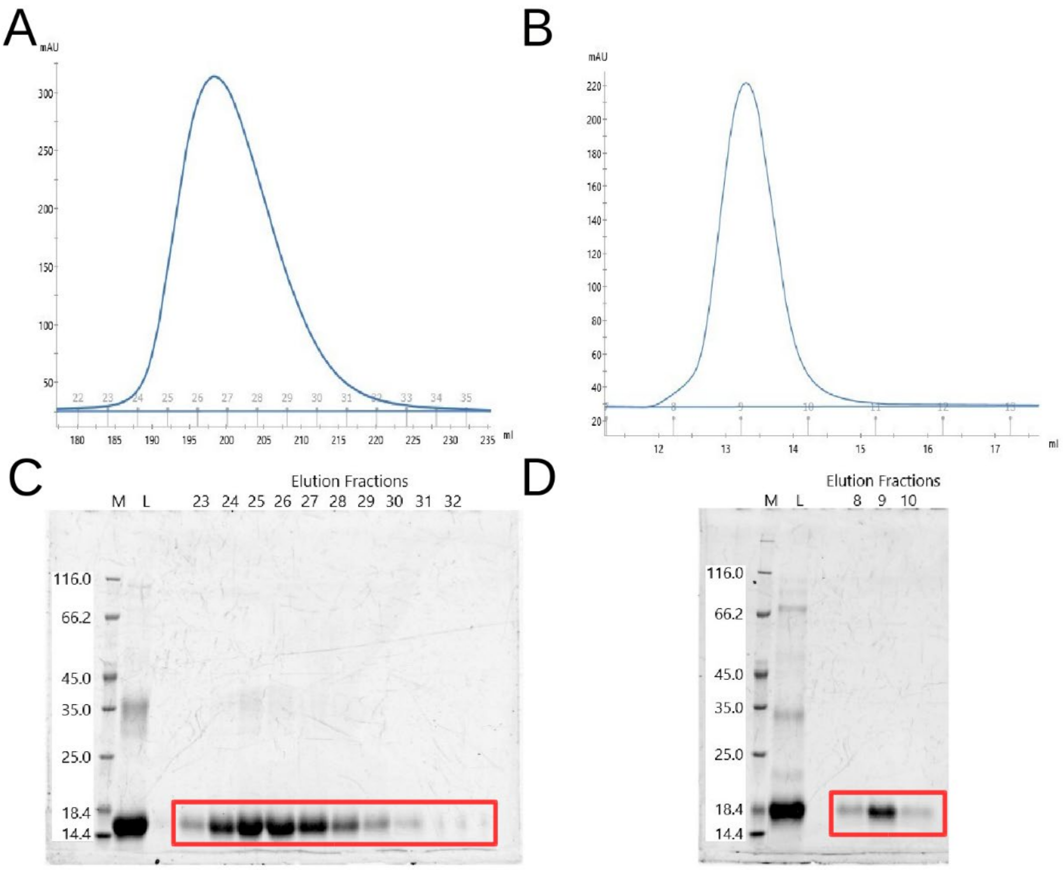


Figure 76: Chromatograms and SDS-PAGE gels generated from the size exclusion-based purification of FGFs 1 (A and C) and 2 (B and D). The target FGF bands are shown within the red box at 16kDa and 18kDa respectively for the appropriate elution fractions. M = Protein Mw Marker and L = Loaded Sample while numbered wells indicate the elution fraction.

The reduced yield of FGF-2 elicited from its expression and purification did not warrant further optimisation given the small amount of protein that was required for DSF screening assays. However, the low yield did necessitate a change in SEC column size and elution protocol that resulted in equally pure protein fractions. The low FGF-2 yield observed resulted from the use of a construct that expressed the complete protein, in contrast to FGF-1 where the first twenty amino acids had been removed. Using the full sequence retained a series of 5’ terminal stem loops in the mRNA that were not present in that encoding FGF-1, which likely served to disrupt the translation process. Whilst this effect may have been reduced by the addition of an affinity tag, there was not one present on the FGF-2 construct used, however, as the yield generated was sufficient to carry out the further experiments required for this protein, no further action was required. A representative fraction from this peak was then concentrated and measured via 1D ^{1}H NMR which also indicated the presence of folded protein, matching spectra from previous FGF-2 purifications, Figure 77.

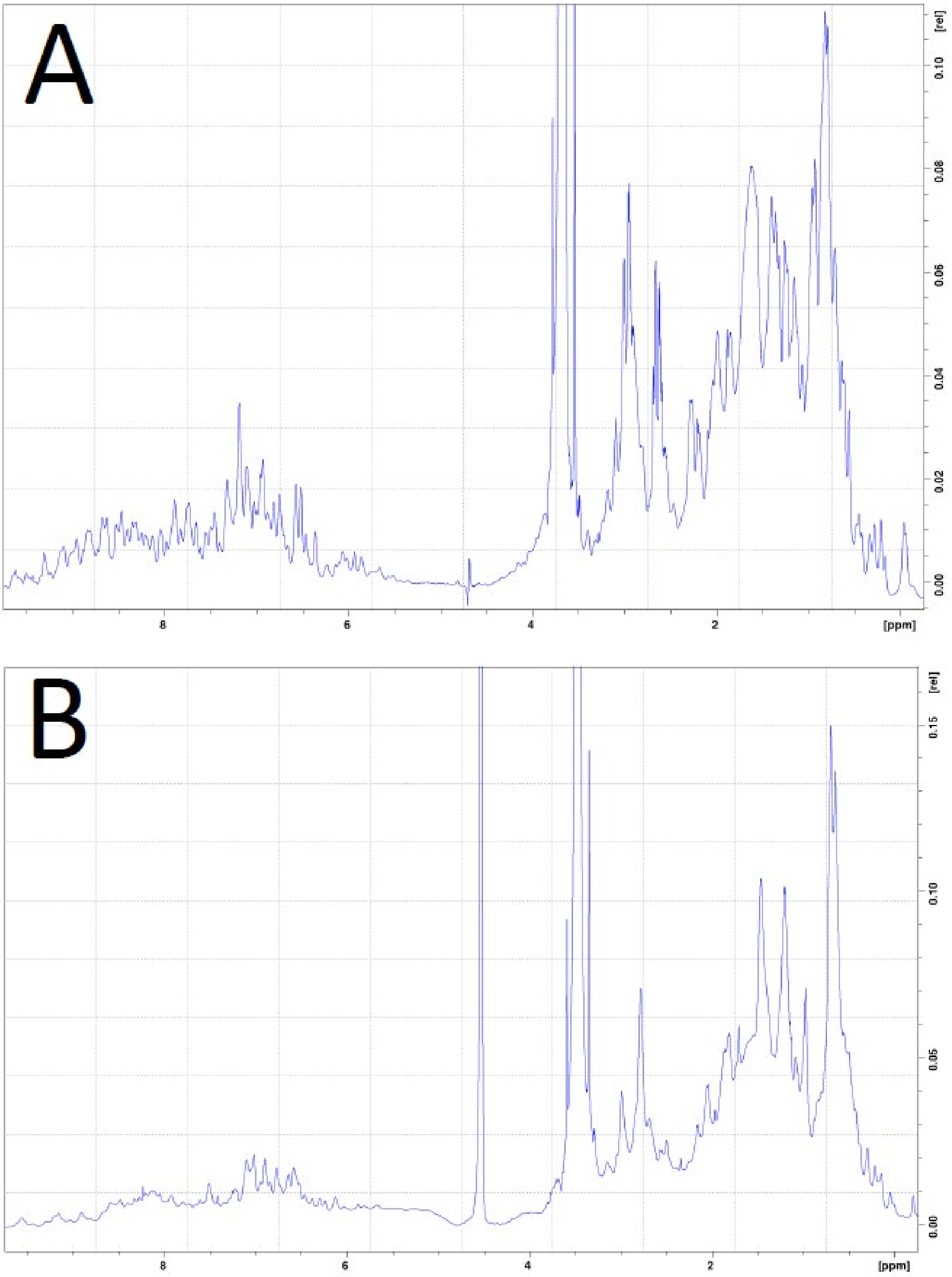


Figure 77: Sections of the 1D ^{1}H NMR spectra of FGF-1 (A) and -2 (B) unlabelled protein samples exhibiting the all of the peaks present within the respective spectra, those regions at the extremes of the ppm range measured, where peaks were not present, have been excluded.

In conclusion, each of the four proteins when expressed and purified generated high purity samples that were subsequently used for both DSF and ITC experiments. However, of the four proteins produced only three generated a high yield, with that for the fourth, FGF-2, considerably reduced. Despite being inadequate for certain biophysical analysis techniques, the amount generated was sufficient to progress with both the DSF screening assays and ITC titrations.

6.2.2 – Differential Scanning Fluorimetry Protein-Carbohydrate Interaction Screening:

The FGF 1 and 2 as well as cyclophilin B and D proteins produced were used to screen for interactions by DSF. This technique was employed as it required low sample concentrations whilst being high throughput, capable of screening 96 carbohydrate and protein combinations in one batch. Use of DSF enabled comparison between the results generated in this project and those reported in previous publications for the stability of FGF-1 and FGF-2 when present alone and in combination with a range of carbohydrates (Uniewicz et al., 2011; Xu et al., 2012; Sun et al., 2019). These prior results detailed a melting temperature increase when combined with a selection of unmodified carbohydrates as well as modified heparin sulfation patterns (Uniewicz et al., 2011; Xu et al., 2012; Sun et al., 2019). In contrast to this, the lack of previously reported data for cyclophilin-carbohydrate interactions made such a comparison impossible. The melting temperature data recorded from the DSF experiments performed here was used to indicate changes to protein stability on carbohydrate addition and from this suggested whether an interaction had occurred. However, further experiments, using an alternate biophysical interaction analysis technique were required in order to support the results generated by DSF. A table of the melting temperatures and their standard deviations is shown in Appendix 5.

6.2.2.1 – DSF Analysis of the FGFs and Cyclophilins Alone:

The DSF experiments measured, in triplicate, for FGF-1, FGF-2, CypB, and CypD when present without a carbohydrate each produce a curve showing a sigmoidal increase in normalised fluorescence with increasing temperature, suggestive of protein unfolding. The temperature points shown in Figure 78 and the remaining figures in this section are those between the minimum and maximum fluorescence values recorded, hence only presenting the unfolding curve from the data collected. Those temperature points before that generating the minimum normalised fluorescence or after that generating the maximum normalised fluorescence were excluded during the early stages of computational data analysis.

Computational analysis of the data used to generate these normalised fluorescence curves enabled the determination of melting temperatures for FGF-1, FGF-2, CypB, and CypD, by using their V_{50}, to give 48.4°C ± 0.05°C, 54.6°C ± 0.07°C, 55.1°C ± 0.03°C, and 50.3°C ± 0.05°C respectively. The similarity of the calculated melting temperatures indicated that it was reasonable to use these proteins for comparison with carbohydrate containing samples, as well as for comparisons between themselves. The temperatures recorded in these experiments for FGF-1 and FGF-2 were similar to those reported in previous publications, being 1.1°C and 1.4°C lower respectively, further supporting their suitability for use in the DSF experiments performed in this project (Sun et al., 2019). The replicability of these denaturation experiments was good with low standard deviations reported for the four proteins, each being less than a tenth of a degree, which was minimal compared to the differences between melting temperatures reported.

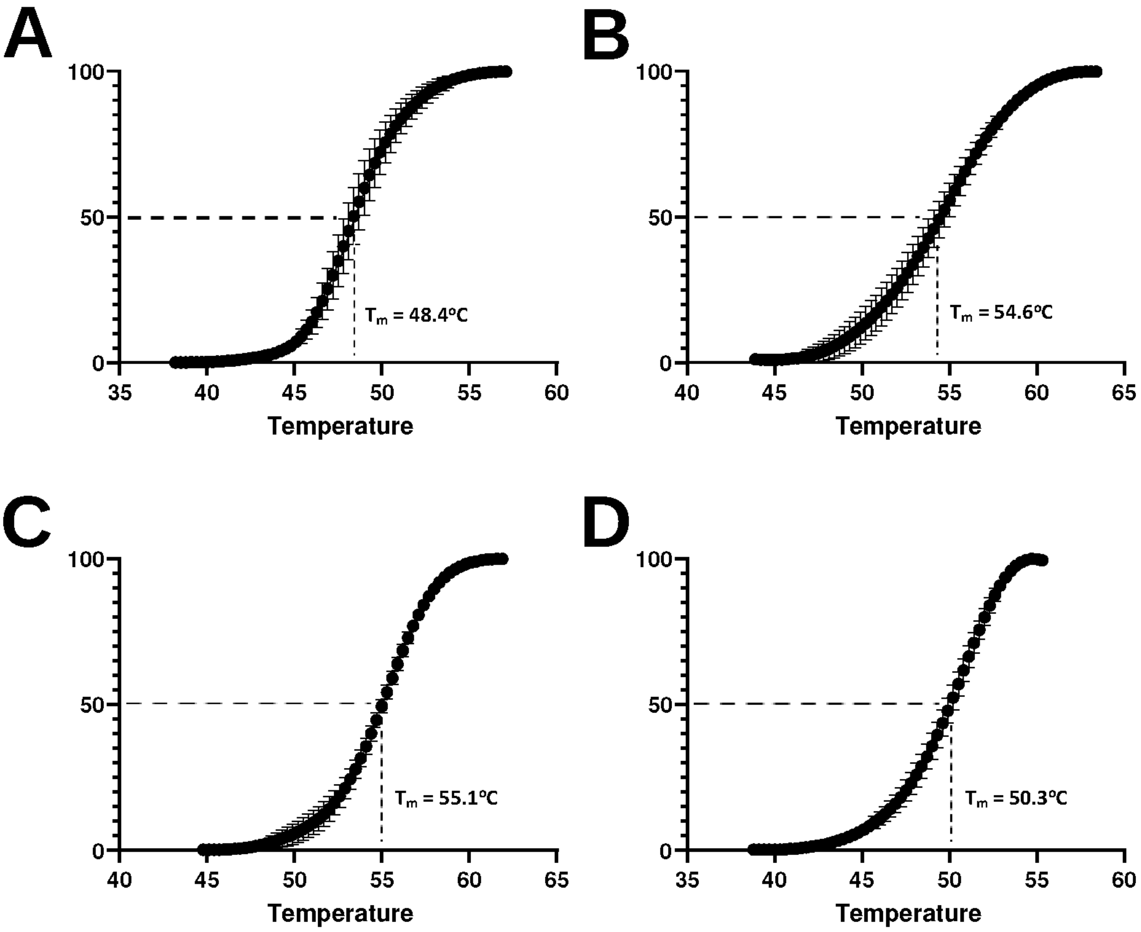


Figure 78: Normalised fluorescence of 5μM FGF-1 (A), FGF-2 (B), CypB (C), and Cyp D (D) when present without carbohydrate.

6.2.2.2 – Effect of Natural Carbohydrates on FGF and Cyclophilin Stability:

The addition of unmodified polysaccharide samples had a considerable effect on the normalised fluorescence curves for each of the proteins tested, Figure 79. The FGF-1 and FGF-2 unfolding curves shifted to higher temperatures, Figure 79A and B, on the addition of both heparin, purple, and ι-carrageenan, blue, indicating greater protein stability. That of FGF-1, Figure 79A, also shifted to higher temperatures when combined κ-carrageenan, green, however with FGF-2, Figure 79B, the observed shift was towards the lower temperatures. The unfolding curve shifts observed were more extreme with increasing degree of sulfation, as that of heparin, with three sulfate groups, shifted further than either ι-carrageenan, with two, or κ-carrageenan, with one. The unfolding curve of neither protein was seen to shift on the addition of unmodified xyloglucan, yellow.

By contrast, carbohydrate addition shifted the CypB and CypD unfolding curves to lower temperatures, potentially indicative of decreased protein stability. The CypD unfolding curve shift, Figure 79D, was most extreme when combined with heparin, purple, while a minor effect was observed with ι-carrageenan, blue. Neither κ-carrageenan, green, nor xyloglucan, yellow, generated an observable unfolding curve shift. The Cyp B unfolding curve shift, Figure 79C, varied most extremely on the addition of κ-carrageenan, green, with ι-carrageenan, blue, and heparin, purple, doing so to a lesser extent. As with CypD, no effect was seen for Xyloglucan, yellow.

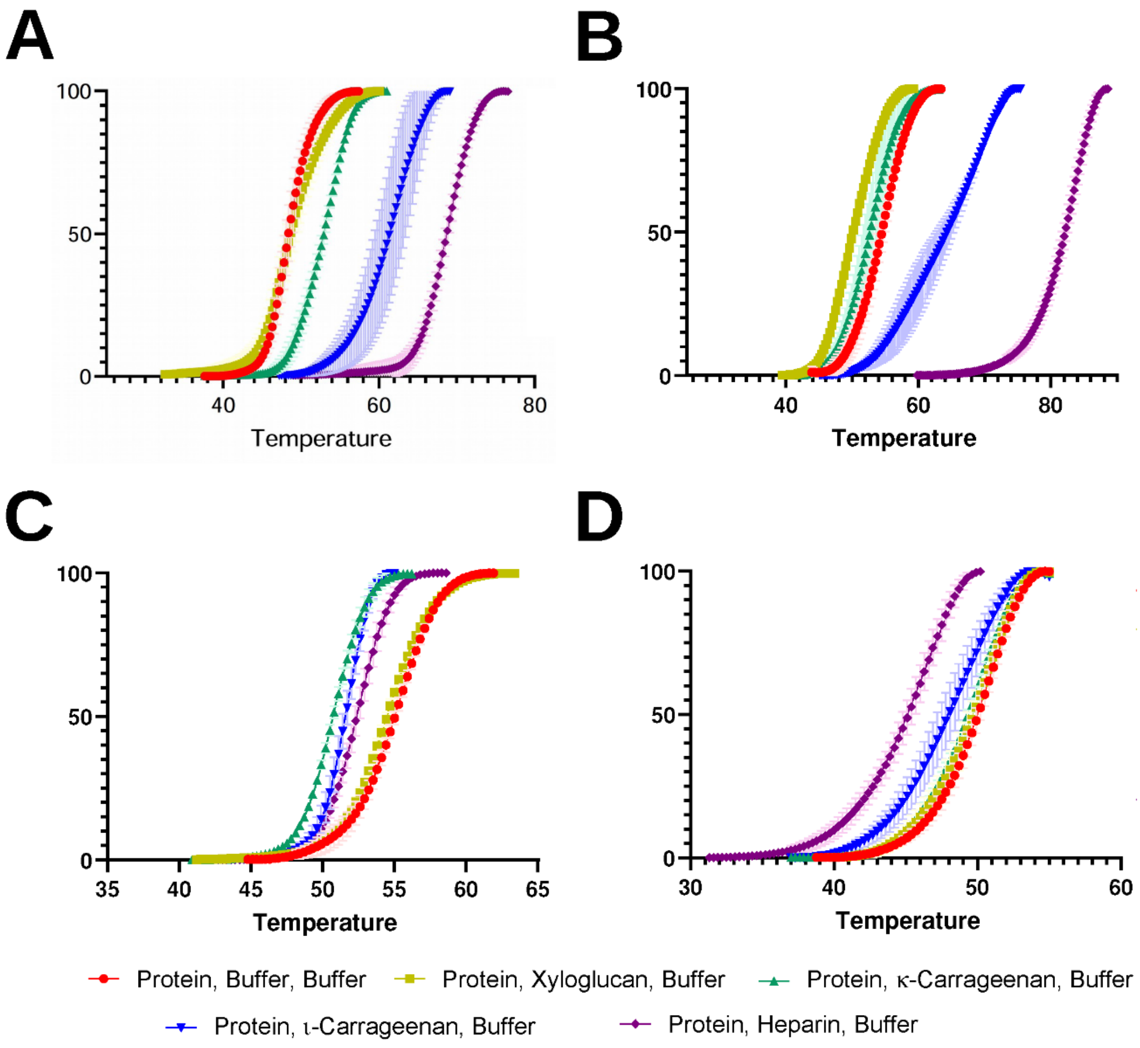


Figure 79: Stabilisation, or destabilisation, effect of 0.5mg/mL Xyloglucan (yellow), κ-carrageenan (green), ι-carrageenan (blue), heparin (purple), and buffer (red) on the melting of 5μM FGF-1 (A), FGF-2 (B), CypB (C), and Cyp D (D). The shading present around the points measured, and particularly prevalent for the ι-carrageenan samples, corresponds to the error bars for each of those measurements.

The changes observed from the normalised fluorescence (unfolding) curve shifts, Figure 79, were quantified by analysis of the proteins' melting temperatures and, from these, the stabilising effect of the carbohydrate present, relative to that of heparin, Figure 80. As mentioned above heparin had the greatest effect on both FGFs 1 and 2, increasing their melting temperatures by 20.4°C and 28.4°C, to 68.8°C and 83.0°C respectively. The corresponding FGF 1 and 2 melting temperature increases of 13.2°C and 11.4°C, respectively, when ι-carrageenan was added equated to a stabilising effect that was 65% of that of heparin on FGF-1 and 40% on FGF-2. Similarly, the melting temperature increase derived by κ-carrageenan was 4.5°C for FGF-1, corresponding to 23% of the heparin stabilising effect. Compared to the FGFs, Cyclophilin melting temperature differences were far less extreme on the addition of an unmodified polysaccharide. The greatest melting temperature variations exhibited for CypB and CypD in this project were a 4.3°C decrease, which occurred when κ-carrageenan was added to CypB, giving a 50.8°C melting temperature – an effect that was 125% of that of heparin, and a 4.5°C CypD meting temperature decrease on the addition of heparin, giving a melting temperature of 45.9°C.

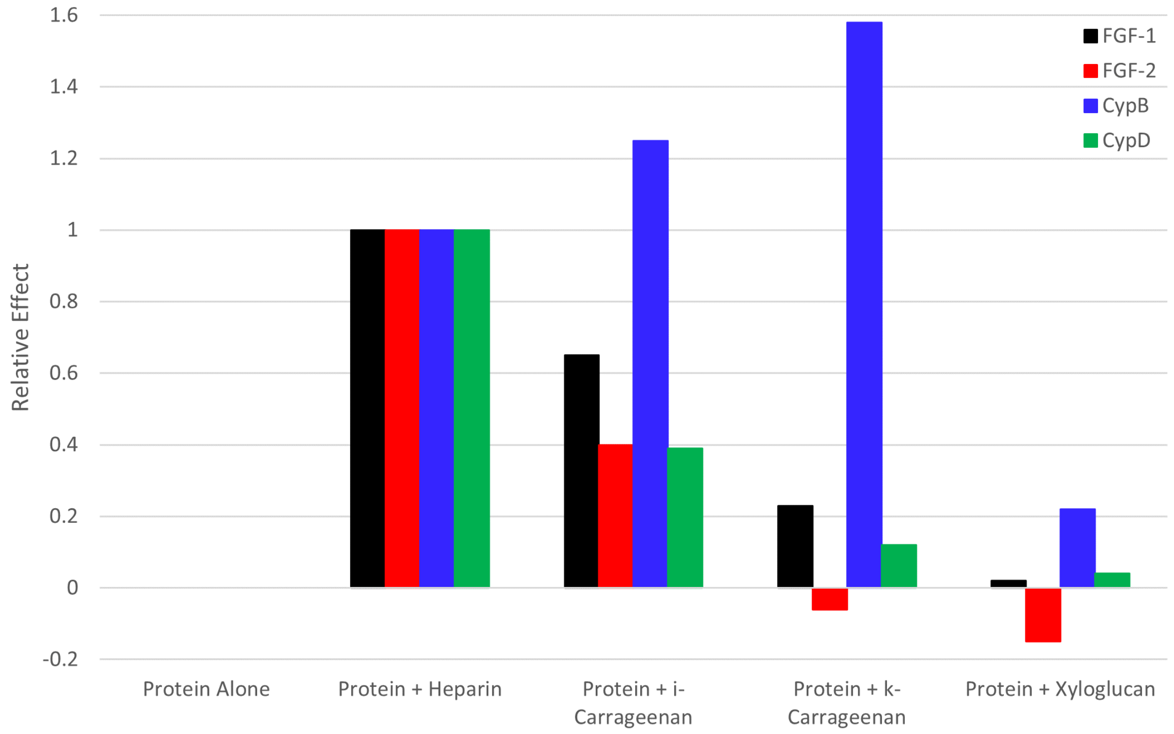


Figure 80: The effect on 5µM FGF-1 (Black), FGF-2 (Red), CypB (Blue), and Cyp D (Green) melting temperature, relative to heparin, when alone, and with 0.5mg/mL i-carrageenan, k-carrageenan and xyloglucan, relative to that of heparin, which was assigned the value 1.

Given the melting temperature differences observed, and the stabilising effects relative to heparin, it could be suggested that interactions formed between FGFs 1 and 2 with heparin and ι-carrageenan. FGF-1 was also observed to interact with κ-carrageenan, where this interaction had a reduced effect on protein stabilisation. Similarly, the melting temperature differences recorded suggested that Cyp B interacted with both carrageenan polymers as well as that of heparin, while CypD only formed interactions with ι-carrageenan and heparin. Whilst DSF identified interactions between several of the carbohydrates and proteins tested, further biophysical analysis techniques such as NMR, Circular Dichroism, or Isothermal Titration Calorimetry could be used to either support the identification of these interactions or provide further information about them. However, care would be required during the selection of the further biophysical analysis techniques to be used, including during the analysis of any data collected, so as not to draw inappropriate conclusions. One notable example of this from previously published research, during the analysis of antithrombin-heparin interactions, was the lack of secondary structural changes observed by circular dichroism, indicative of a lack of interactions having been formed, in spite of the data from DSF having confirmed an interaction had occurred (Lima et al., 2013).

The differences in melting temperature resulting from the addition of unmodified polysaccharide samples indicated the importance of degree of sulfation on FGF and cyclophilin melting temperature, and therefore protein stability and interaction. Both of the FGFs, 1 and 2, tested were seen to exhibit a greater increase in melting temperature with increasing degree of sulfation while CypD exhibited a greater decrease in melting temperature with increasing degree of sulfation. In contrast to these three proteins, CypB melting temperature decreases were more extreme with decreasing degree of sulfation, provided at least one sulfate group is present. Consequently, the effect that chemical sulfation of these unmodified polysaccharides had on melting temperature, protein stability, and carbohydrate-protein interactions, was examined.

6.2.2.3 – Effect of Carbohydrate Sulfation on FGF and Cyclophilin Stability:

The observation that increasing degree of carbohydrate sulfation correlated with increased melting temperature variation resulted in combinations of the four proteins that were tested with a selection of the modified samples produced in Chapter Five. These combinations were then used to determine the effect that chemical sulfation had on melting temperature, protein stability, and interaction formation. The chemically sulfated heparin samples, produced using the traditional pyridine sulfur trioxide in DMF protocol, that were screened for interactions with FGFs 1 and 2 as well as cyclophilins B and D are shown in Table 5 below. The samples produced by the pyridine sulfur trioxide sulfation protocols in DMF were used instead of those produced via the novel reactions, run in BMImCl, as they exhibited greater variability between themselves, whilst the samples individually were more homogeneous.

Table 5: A table detailing the modified and unmodified heparin sulfation patterns used for interaction screening with FGFs 1 and 2 as well as Cyclophilins B and D.

Modified Sulfation Pattern #	Sulfation Pattern (% - A = glucosamine, I = iduronic acid)			
	$A_2 - N(H)SO_3$	$A_3 - OSO_3$	$A_6 - OSO_3$	$I_3 - OSO_3$
Unmodified Polymer	99.2	0.1	74.3	3.7
1	31.7	0.0	88.5	9.8
2	32.7	11.9	93.0	24.5
3	60.0	80.8	97.7	75.0
4	90.8	95.4	98.4	50.7*

Each of the four modified heparin samples tested was observed to generate normalised fluorescence curves similar to that of the unmodified polysaccharide, Figure 81. When combined with FGF-1 the four modified sulfation patterns generated curves which clustered together at a slightly lower melting temperature than the unmodified polysaccharide, Figure 81A. In contrast to this, with FGF-2 three of the four modified heparin samples were seen to cluster at the same temperature as the unmodified polymer, the fourth, present at a slightly lower melting temperature, being modified sulfation pattern four, which was most highly sulfated, Figure 81B. Whilst this corroborated the fact that FGFs 1 and 2 had different tolerances for sulfation pattern differences, in this instance FGF-1 was stabilised less by samples with a lower degree of sulfation, which was not the case with the starting materials.

As with the FGFs, a similar clustering of multiple samples was observed for the cyclophilins. All four modified samples appeared to give unfolding curves present close to, if not at slightly lower temperatures than, the unmodified polysaccharide when combined with CypB, Figure 81C. However, for CypD while the unfolding curves for modified sulfation patterns one and two were present at the same melting temperature as the unmodified polysaccharide, modified sulfation patterns three and four were shifted to a higher melting temperature, Figure 81D. However, each of the CypB and D protein samples supplemented by either the unmodified polysaccharide or modified sulfation pattern samples gave unfolding curves that were shifted to lower temperatures than was seen for either of the proteins alone, consistent with the changes seen in Figure 79C and Figure 79D. As carbohydrate addition to cyclophilin samples was shown to give unfolding curves present at lower temperatures, this suggested that modified sulfation patterns three and four have less of an effect on protein stability. This, coupled with the fact that CypD showed no variation in melting temperature when combined with the single sulfate group containing κ-carrageenan sample suggested that CypD may have had an optimal degree of sulfation of two or three groups for carbohydrate interaction. The clustering of all five modified and unmodified heparin sulfation pattern samples for CypB could have conversely suggested a wider range of degrees of sulfation were capable of forming interactions.

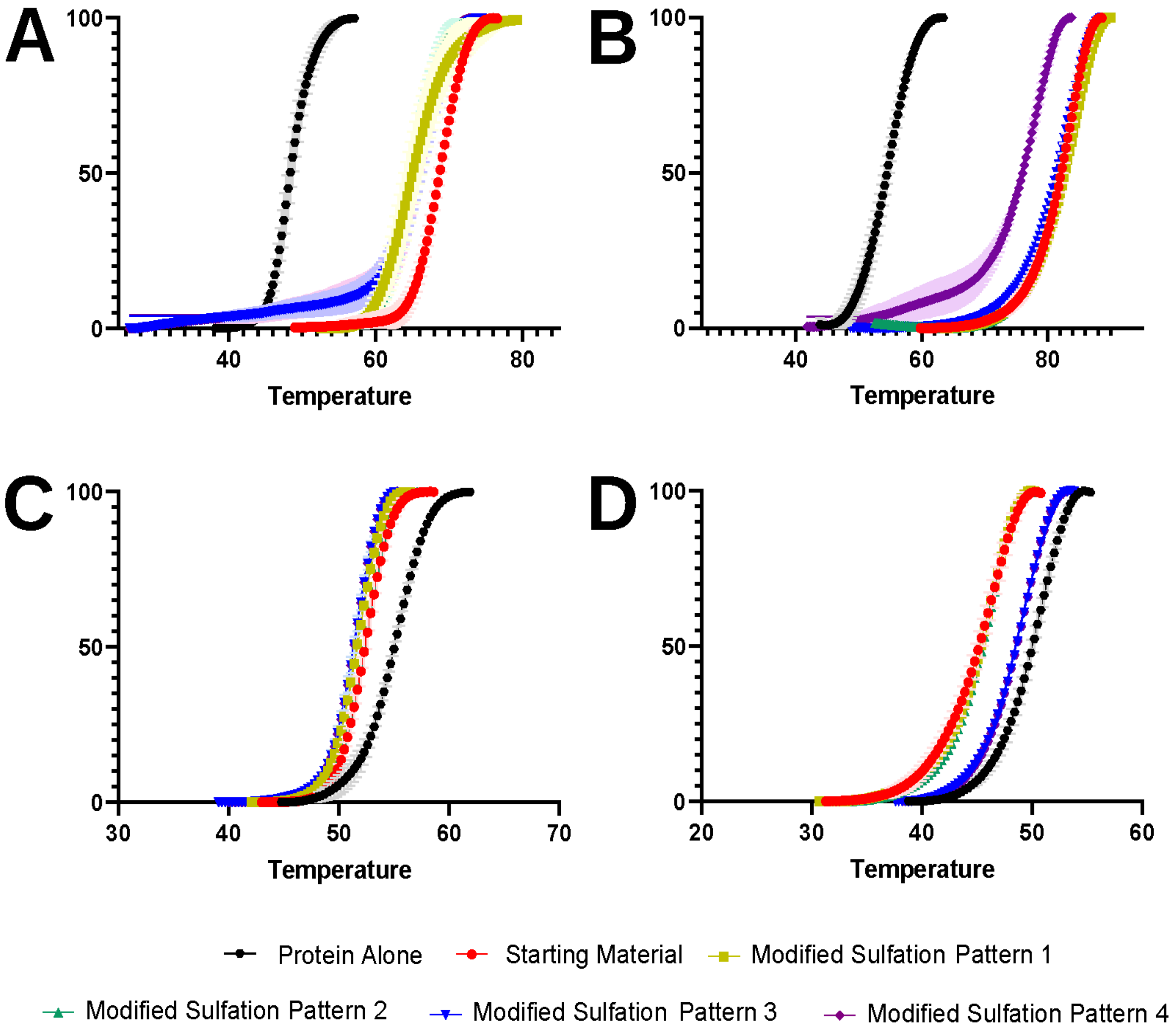


Figure 81: Stabilisation, or destabilisation, effect of Modified Heparin Sulfation Patterns 1 (yellow), 2 (green), 3 (blue), and 4 (purple), as well as the unmodified polysaccharide (red) and the protein alone (black) on the melting of 5μM FGF-1 (A), FGF-2 (B), CypB (C), and Cyp D (D). As above, the shading present around the points measured corresponds to the error bars for each of those measurements.

The melting temperatures and consequent stabilising, or destabilising, effect on the proteins, relative to heparin, determined from these normalised fluorescence curves and those of the protein alone highlighted the effect that sulfation pattern modification had on protein stability, Figure 82. Compared to the unmodified polysaccharide, which caused a 20.4°C melting temperature increase, that engendered by the modified sulfation patterns was less extreme, between 16.5°C and 18°C, when combined with FGF-1. This equated to the four modified sulfation pattern samples having had between 81% and 88% of the stabilising effect of the unmodified polysaccharide. However, whilst lower than the unmodified polysaccharide, each of these melting temperature variations, and relative stabilising effects, suggested a considerable increase in protein stability on carbohydrate addition and hence the formation of an interaction having occurred. Within the four modified sulfation pattern samples there was a general trend for increased melting temperature variation, from that of the protein alone, with increased degree of sulfation, however, as mentioned, the less sulfated unmodified polysaccharide derived the greatest effect, and consequently did not fit with this general trend, Figure 82. As with FGF-1 each of the modified heparin sulfation pattern samples increased FGF-2 melting temperature, in a manner which stabilised the protein and was suggestive of an interaction having formed. The differences in melting temperature that resulted from the addition of these samples ranged from an increase of 23.8°C, i.e. 83% of the stabilising effect of the unmodified polysaccharide, for modified sulfation pattern four to 29°C, a stabilising effect that was 102% of that observed from unmodified heparin, for modified sulfation pattern one, thus having generated a marginally more stable protein, Figure 82. The FGF-2 melting temperature increases recorded appear to follow a trend of increasing degree of sulfation resulting in a lesser melting temperature increase, with temperatures of 83.6°C, 82.5°C, 82.3°C, and 78.4°C recorded for modified sulfation patterns one, two, three, and four, respectively, compared to 54.6°C for FGF-2 alone.

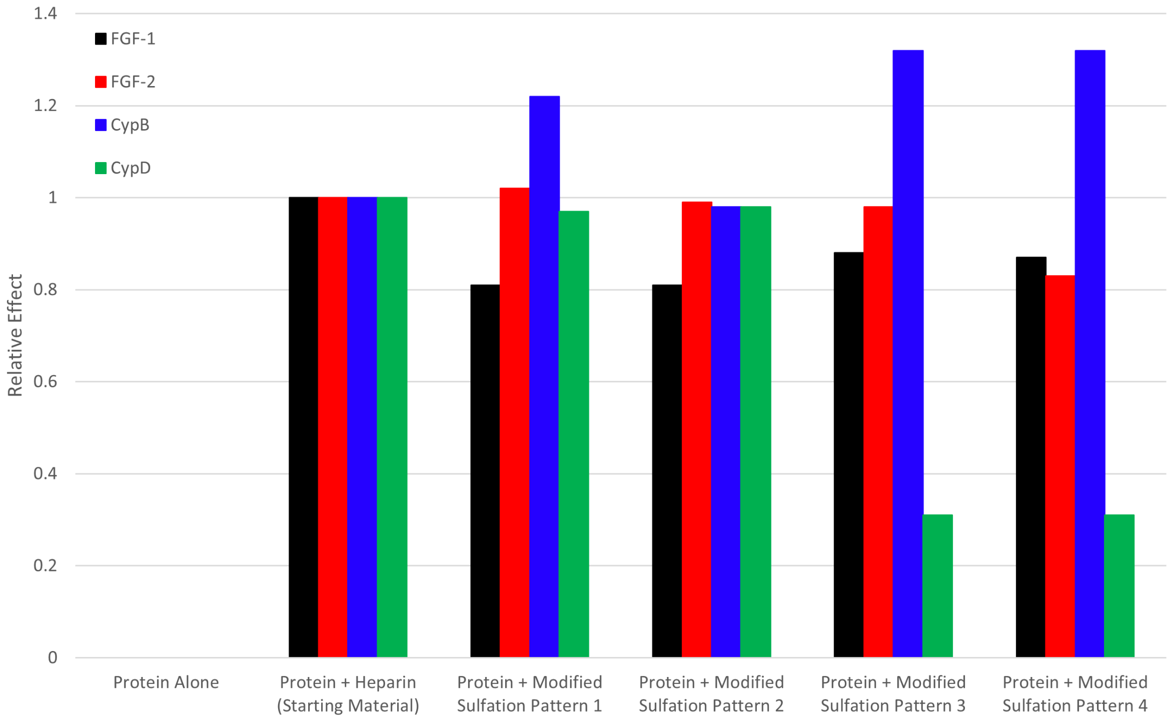


Figure 82: The effect on 5μM FGF-1 (Black), FGF-2 (Red), CypB (Blue), and Cyp D (Green) melting temperature, relative to heparin, when alone, and with 0.5mg/mL starting material heparin, assigned the value 1, and modified sulfation patterns one, two, three, and four.

Whilst there were no obvious trends in CypB melting temperature difference with increasing sulfation the samples were seen to form two groups, where modified sulfation patterns one, three, and four, had a greater destabilising effect on the protein than the unmodified polysaccharide and modified sulfation pattern two, Figure 82. In terms of the destabilising effect of these four modified sulfation patterns relative to heparin, patterns one, three, and four had 122%, 132%, and 132% of that of heparin respectively, while modified sulfation pattern two had 98% of the effect observed for heparin. This agrees with the oversulfated heparin repeating unit sulfation pattern previously shown to interact with CypB, that was 2-O- and 3-O-sulfated on the iduronic acid residue and 3-O-, 6-O-, and N-sulfated on the glucosamine, i.e. that present in modified sulfation patterns three and four (Hanoulle et al., 2007; Vanpouille et al., 2007). However, the greater CypB stabilising effect observed for modified sulfation pattern one relative to that of unmodified heparin was unexpected. This was owing to the modified sulfation pattern one sample being sparsely sulfated on both the 3-O-positions of the glucosamine and iduronic acid residues and considerably desulfated on the amine group of the glucosamine residue. CypD when combined with the modified heparin sulfation patterns showed the same increase in melting temperature with increasing degree of sulfation as was seen for FGF-1. However, as CypD was destabilised by carbohydrate addition this meant that increasing sulfation pattern had a reduced effect on protein stability as identified from melting temperature variations. These modified sulfation pattern samples had 97%, 98%, 31%, and 31% of the destabilising effect seen for heparin, for modified sulfation patterns one, two, three, and four respectively, Figure 82. As with FGF-1, but unlike CypB, the unmodified carbohydrate had the greatest effect on CypD melting temperature.

Having tested the effect that modified heparin sulfation pattern samples had on interactions with the four proteins tested, the effect engendered by modification of κ-carrageenan was examined, again using a selection of the samples produced in Chapter Five, shown in Table 6.

Table 6: A table detailing the unmodified polysaccharide and modified κ-carrageenan sulfation patterns used for interaction screening with FGFs 1 and 2 as well as Cyclophilins B and D.

Modified Sulfation Pattern #	Sulfation Pattern (% - G4S = galactose-4-sulfate, DA = 3,6-anhydrogalactose)		
	$G4S_2 - OSO_3$	$G4S_6 - OSO_3$	$DA_2 - OSO_3$
Unmodified Polymer	2.6	0.0	0.0
1	9.8	9.05	5.5
2	25.7	19.2	6.5
3	95.5	98.9	66.3

As would be expected given the greater increase in degree of sulfation when κ-carrageenan was modified, compared to heparin, a greater shift was seen in the normalised fluorescence curves. When combined with FGFs 1 and 2 the increased degree of sulfation samples were all seen to generate curves at very similar positions, Figure 83A and Figure 83B. Whilst it would have been expected that FGF-1 was stabilised by each of the modified sulfation patterns, given it was by unmodified κ-carrageenan, the dramatic increase in melting temperature exhibited by FGF-2 when combined with modified sulfation pattern one especially was unexpected, Figure 83B. This modified sulfation pattern only showed between 5% and 10% sulfation at each on the unsubstituted hydroxyl groups, thus having a total degree of sulfation per repeating unit closer to κ-carrageenan, which did not stabilise FGF-2, than ι-carrageenan, which did. However, the location of the sulfated hydroxyl groups across the polysaccharide chain remained unknown, consequently a greater concentration of potential binding sites could have been formed by these modifications than was implied by the raw percentage sulfation data.

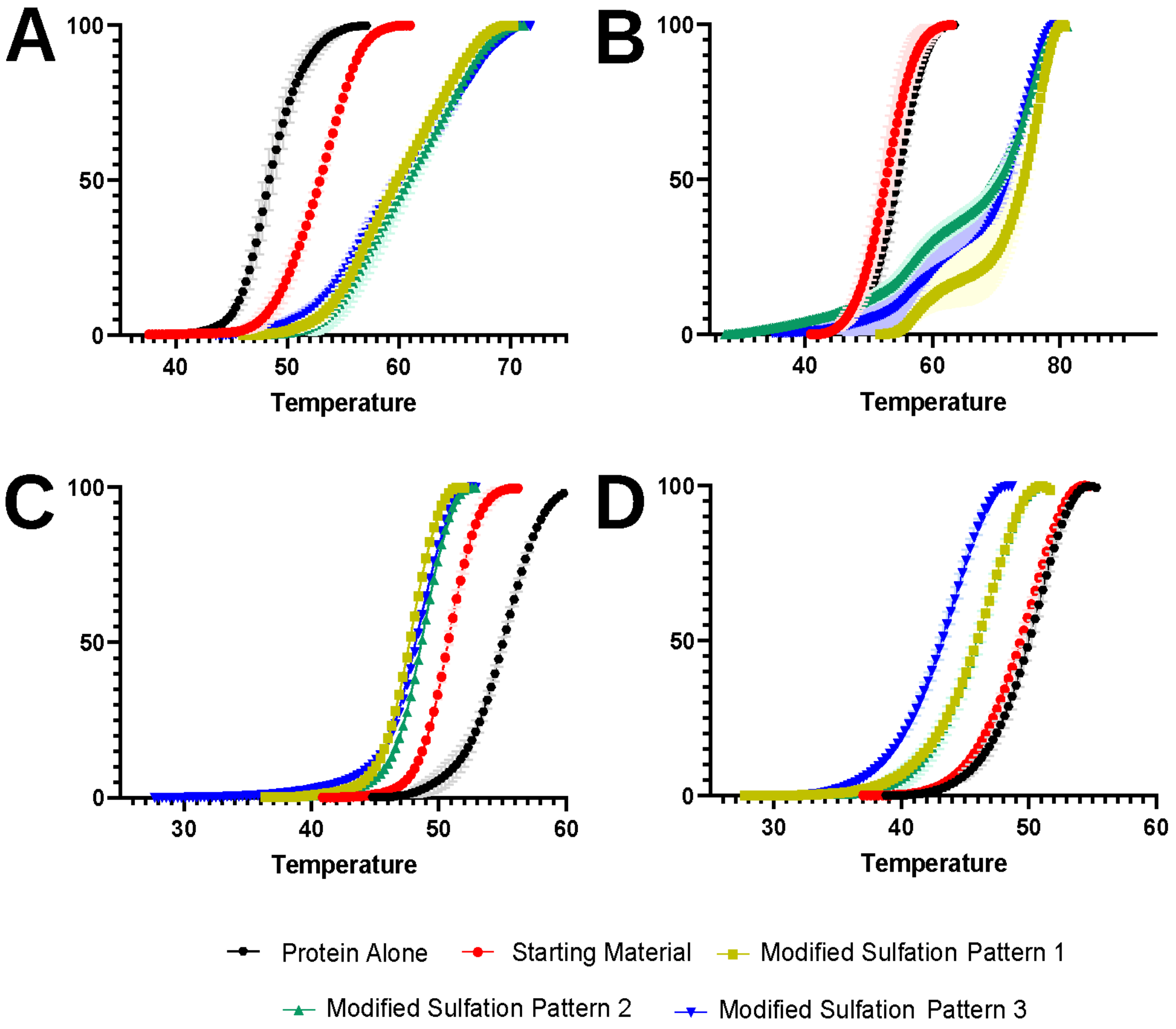


Figure 83: Stabilisation, or destabilisation, effect of Modified κ-carrageenan Sulfation Patterns 1 (yellow), 2 (green), and 3 (blue), as well as the unmodified polysaccharide (red) and the protein alone (black) on the melting of 5μM FGF-1 (A), FGF-2 (B), CypB (C), and Cyp D (D). As above, the shading present around the points measured corresponds to the error bars for each of those measurements.

As with the FGFs, CypB was effected, albeit destabilised rather than stabilised, to the same extent by the three modified κ-carrageenan sulfation patterns, Figure 83C. However, for CypD, modified sulfation pattern three was seen to have a far greater destabilising effect on the protein than either modified sulfation patterns one or two, which in turn had a greater effect than the unmodified polysaccharide, Figure 83D. In contrast to the other samples measured by DSF the melting curves for FGF-2 with the modified κ-carrageenan samples have two transitions indicative of there being two unfolding events. This most likely results from there being a two component system with more than one population of bound protein present, alternatively this effect could be caused by inhomogeneity within the κ-carrageenan samples.

The extent to which these variations occurred was quantified by the melting temperatures, and consequent stabilisation or destabilisation effects, relative to unmodified heparin, calculated for the combinations of modified and unmodified κ-carrageenan with the four proteins tested. It was seen from these temperatures that although only two of the proteins, FGF-1 and CypB, were affected by unmodified κ-carrageenan addition they all exhibited an altered melting temperature on the addition of the modified sulfation pattern samples, Figure 84. FGF-1 samples exhibited a melting temperature increase of between 12.0°C and 12.9°C on modified sulfation pattern sample addition corresponding to a stabilising effect of between 58% and 63% of that derived from the addition of unmodified heparin. This is considerably more than the 23% relative stabilising effect derived from addition of the κ-carrageenan starting material. For FGF-2, previously shown not the bind the unmodified polysaccharide, the effect of modified sulfation pattern addition was similar, with melting temperatures having increased by no lower than 16.1°C for the three samples. This equated to relative stabilising effects that were no less than 56% of that of the unmodified heparin polymer, Figure 84. Despite considerable increases in melting temperature variation compared to that engendered by the unmodified polysaccharide, no general trend regarding the effect of increasing degree of sulfation was observed on the extent of melting temperature increase between the three modified samples.

Protein stability decreases were also readily apparent from the cyclophilin B and D melting temperatures, where the decreases observed when a modified sulfation pattern sample was added were no less than 6.3°C and 4°C, respectively, much higher than those of the unmodified polysaccharides, which were 4.3°C and 0.6°C, respectively. Comparison of the destabilising effect observed when modified κ-carrageenan samples were added to CypB and CypD versus when unmodified heparin was added indicated that these carrageenan samples had at least 229% and 75% of the effect engendered by unmodified heparin, respectively, Figure 84. While FGF-1, FGF-2, and CypB interactions with the modified sulfation pattern samples did not show a trend for the effect of increasing degree of sulfation on melting temperature beyond the considerable increase from that derived from the unmodified polysaccharide, the CypD samples did. In these samples it was observed that increasing degree of sulfation caused a greater melting temperature decrease, and hence more extremely destabilised protein. As was seen from the normalised fluorescence curves the melting temperature decrease on addition of the unmodified sample was less (0.6°C – 10% of the effect of heparin) than that derived from modified sulfation patterns one and two (4°C – 75% of the effect of heparin) which were themselves less than for modified sulfation pattern three (6.7°C – 125% of the effect of heparin), Figure 84. This followed a theme which has appeared repeatedly throughout the DSF screening experiments, performed here and reported thus far. This indicated that a comparison between the unmodified polysaccharides and most extremely modified samples for each carbohydrate would be an appropriate supporting test of these screening experiments.

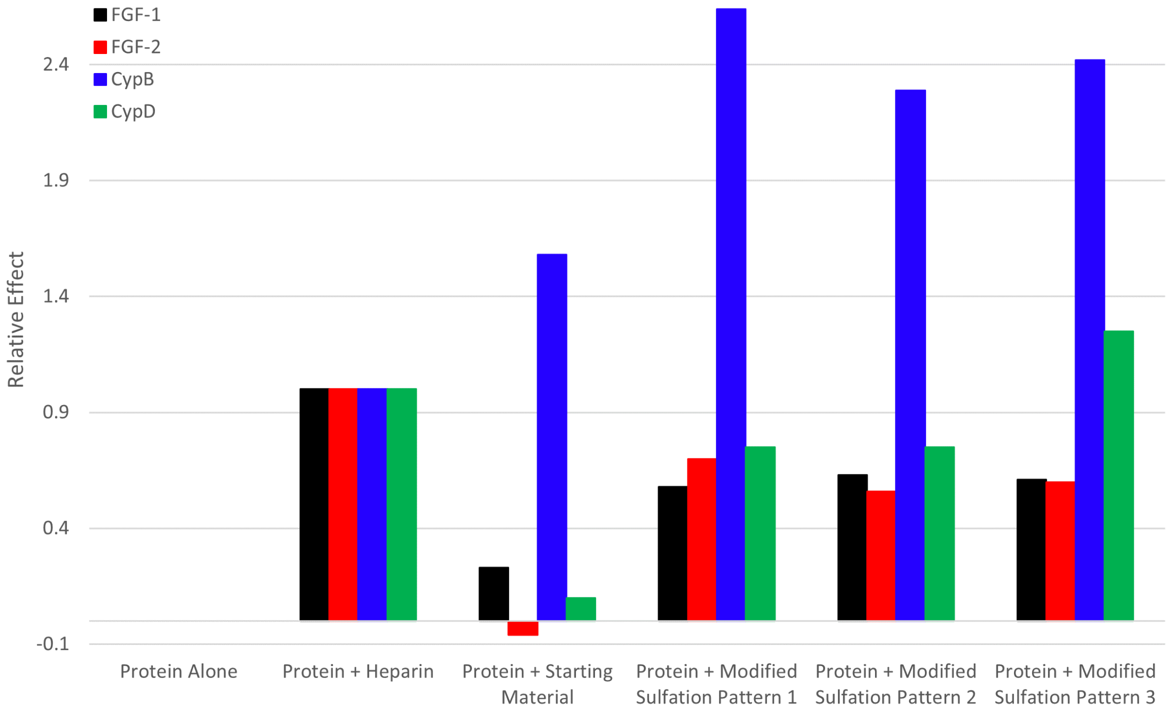


Figure 84: The effect on 5µM FGF-1 (Black), FGF-2 (Red), CypB (Blue), and Cyp D (Green) melting temperature, relative to heparin, when alone, and with 0.5mg/mL starting material heparin, assigned the value 1, starting material κ-carrageenan and modified sulfation patterns one, two, and three.

Having shown that increasing the degree of sulfation of two naturally sulfated unmodified polysaccharides affected the melting temperature of the proteins with which they had been combined, sulfated cellulose was tested with the FGFs and cyclophilins to identify potential interactions. Three sulfation pattern variants, produced in Chapter 5, were therefore used in these tests, Table 7.

Table 7: A table detailing the modified cellulose sulfation patterns used for interaction screening with FGFs 1 and 2 as well as Cyclophilins B and D.

Modified Sulfation Pattern #	Sulfation Pattern (% - Glc = glucose)		
	$Glc_2 – OSO_3$	$Glc_3 – OSO_3$	$Glc_6 – OSO_3$
1	1.1	1.8	65.3
2	83.9	85.8	99.9
3	95.2	97.3	100

The normalised fluorescence curves for the three modified cellulose samples, when added to FGF-1, FGF-2, CypB, and CypD, indicated a range of different effects on protein stability, relative to one another. When FGF-1 and modified sulfation pattern one were combined the melting temperature was shifted higher than it was for modified sulfation pattern two, while modified sulfation pattern three was present at a higher temperature still, Figure 85A. When combined with FGF-2, modified sulfation pattern two remained that which generated the lowest temperature, however, here there melting temperature caused by modified sulfation pattern one addition was higher than that for three, Figure 85B.

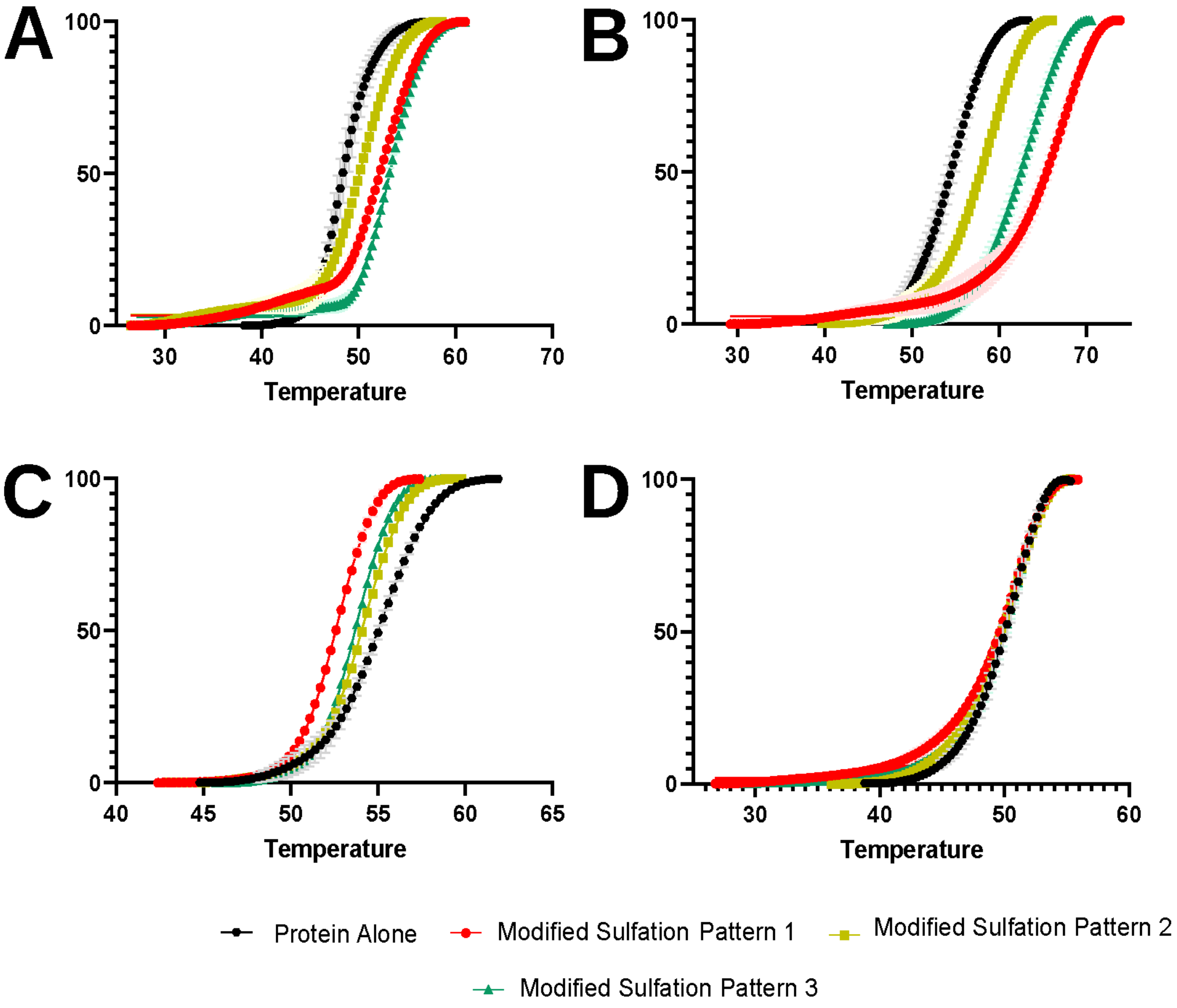


Figure 85: Stabilisation, or destabilisation, effect of Modified Cellulose Sulfation Patterns 1 (red), 2 (yellow), and 3 (green) and the protein alone (black) on the melting of 5µM FGF-1 (A), FGF-2 (B), CypB (C), and Cyp D (D).

The exact opposite pattern to that seen for FGF-2 was present for CypB. The location of the transition of the normalised fluorescence curve for this protein on the addition of modified sulfation pattern one was present at the lowest temperature. The location of this transition for CypB on the addition of modified sulfation pattern three was slightly higher than that for pattern one, while that for modified sulfation pattern two was higher still. However, given that the CypB samples were destabilised by carbohydrate addition modified sulfation pattern one exhibited the greatest effect, Figure 85C. Meanwhile all three of the modified cellulose sulfation patterns gave melting curves for

CypD that were present at the same position, which was also the position of the protein alone suggestive of no interaction having occurred, Figure 85D.

When comparing the effect engendered by the addition of modified cellulose sulfation pattern samples to FGF-1 it was seen that the increase in melting temperature compared to the protein alone (48.4°C) was far more profound for modified sulfation patterns one (4.2°C) and three (5°C) than it was for modified sulfation pattern two (2°C). Comparing these with the effect derived by unmodified heparin indicated that addition of modified sulfation patterns one and three had 21% and 24% or the effect of heparin, respectively, while modified sulfation pattern two had only 10% of the effect of heparin, Figure 86. The melting temperature increases seen with FGF-2 were in general greater than those for FGF-1. The lowest melting temperature increase compared to that of the protein alone (54.6°C) was 3.8°C, for modified sulfation pattern two, whilst the highest was 12.5°C, for modified sulfation pattern one, which equated to 14% and 44% of the effect derived from the addition of unmodified heparin respectively, Figure 86. That the greatest increase in FGF-2 melting temperature was present for the least sulfated modified sample corresponds to that which was previously observed for the modified κ-carrageenan and heparin samples despite this protein having been seen not to interact with the least sulfated unmodified polysaccharides.

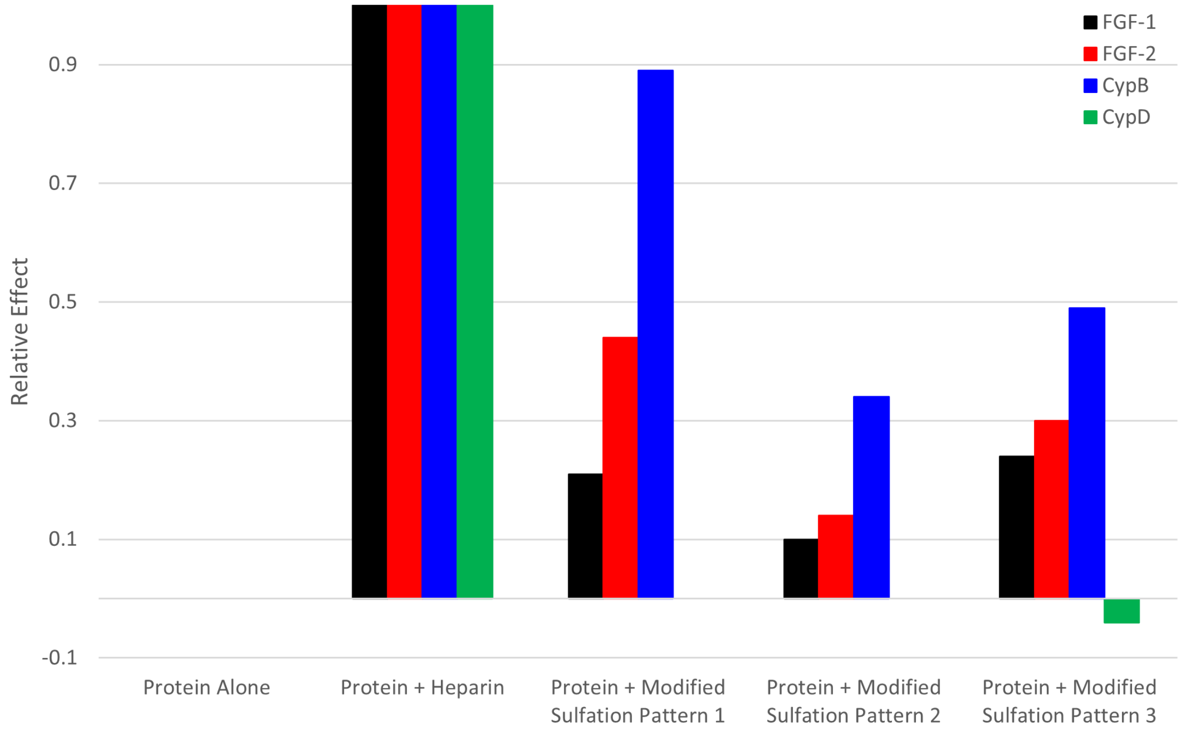


Figure 86: The effect on 5μM FGF-1 (Black), FGF-2 (Red), CypB (Blue), and Cyp D (Green) melting temperature, relative to heparin, when alone, and with 0.5mg/mL starting material heparin, assigned the value 1, and modified cellulose sulfation patterns one, two, and three.

As with the FGFs, each of the three modified cellulose samples affected the melting temperature of CypB, with the decreases observed ranging from 0.9°C, 34% of the destabilising effect of unmodified heparin, for modified sulfation pattern two, to 2.5°C, 89% of that of unmodified heparin, for modified sulfation pattern one, Figure 86. In contrast the this, none of the modified cellulose sulfation pattern samples affected the melting temperature of CypD, thus it was concluded that no interaction was present between the carbohydrates and CypD protein within these samples, Figure 86.

Screening the modified heparin, κ-carrageenan, and cellulose sulfation pattern samples for interactions with FGF-1, FGF-2, CypB, and CypD indicated that all combinations were observed to interact, with the exception of the modified cellulose samples when combined with CypD. Unlike that observed for the unmodified polysaccharides most groups of modified samples did not show a general trend for increased, or decreased, melting temperature variations with increasing degree of carbohydrate sulfation. However, there were a few specific examples where such trends were observed. These included FGF-1 and Cyclophilin B, where increased heparin sulfation has a greater effect on melting temperature variation. Similarly, increased κ-carrageenan degree of sulfation exacerbated the melting temperature decrease seen for cyclophilin D. In contrast to this, increased heparin degree of sulfation reduced the melting temperature increases observed for FGF-2, as detailed above. In order to support the interactions identified by DSF screening and generate thermodynamic data for a limited selection of carbohydrate-protein interactions, the unmodified heparin and κ-carrageenan polysaccharides as well as the modified heparin, κ-carrageenan, and cellulose samples with the highest degree of sulfation were combined with FGF-1 and CypB for ITC experimentation.

6.2.3 – Isothermal Titration Calorimetry Analyses of Select Modified Carbohydrate – Protein Interactions:

Having screened for carbohydrate protein interactions via DSF and identified a limited number of combinations that were of particular interest, further experiments were performed using isothermal titration calorimetry (ITC). These experiments allowed for a more detailed analysis of the selected screening interactions as they provided thermodynamic data, including stoichiometry, Gibbs free energy, enthalpy, entropy, and the equilibrium dissociation constant. The ITC experiments performed used two of the four proteins screened via DSF, namely FGF-1 and CypB, alongside the starting materials for heparin and κ-Carrageenan as well as oversulfated variants of both of these carbohydrates and fully sulfated cellulose.

Experiments performed using FGF-1 contained the protein and carbohydrates at the concentrations reported in Table 8. The results produced by these combinations indicated interactions were present between FGF-1 and heparin, oversulfated heparin, and fully sulfated κ-carrageenan in a manner which generated thermodynamic data. These samples had previously been shown to bind via DSF with considerable melting temperature increases compared to FGF-1 alone, of 20.4^{o}C, 17.8^{o}C, and 12.5^{o}C respectively, on carbohydrate addition. Samples containing FGF-1 alone, as well as when combined with κ-carrageenan and fully sulfated cellulose gave no thermodynamic data. Whilst this could have been indicative of there being no interaction between κ-carrageenan or fully sulfated cellulose and FGF-1, DSF results highlighted melting temperature increases of 4.5^{o}C and 4.9^{o}C respectively, consistent with an interaction having occurred. Therefore, another possibility was that the interactions which formed did so in a manner that didn't cause a measurable temperature change, the parameter used during ITC experiments to determine the thermodynamic data. Whilst the DSF melting temperature increases observed for κ-carrageenan or fully sulfated cellulose with FGF-1 were considerable, those for the three carbohydrates whose interactions generated thermodynamic data were far greater. The entries for FGF-1 alone, and when present with either κ-carrageenan or fully sulfated cellulose, where thermodynamic data was not generated, are displayed by three horizontal lines in Table 8.

Table 8: Protein and carbohydrate concentrations as well as the resulting thermodynamic data for interactions with FGF-1.

	FGF-1 Alone	FGF-1 + heparin	FGF-1 + 3-O-S heparin	FGF-1 + κ-carrageenan	FGF-1 + sulfated κ-carrageenan	FGF-1 + sulfated cellulose
FGF-1 Concentration (μM)	40	40	40	40	40	40
Repeating Unit Concentration (μM)	---	200	400	200	200	200
Stoichiometry (n)	---	0.342	0.764	---	0.497	---
Interacting Repeating Disaccharides (n)	---	2.92	1.31	---	2.01	---
Gibbs Free Energy (ΔG -/kJ/mol)	---	-37.7	-35.9	---	-33.3	---
Enthalpy (ΔH / kJ/mol)	---	-41.2 ± 1.4	-24.3 ± 1.28	---	-50.6 ± 2.61	---
Entropy (-TΔS / kJ/mol)	---	3.58	-11.6	---	17.3	---
Equilibrium Dissociation Constant (K_D / μM)	---	0.256 ± 0.07	0.514 ± 0.218	---	1.48 ± 0.305	---

The stoichiometries of the three interactions reported here, Table 8 and Figure 87, represent the number of FGF-1 proteins that formed interactions, per carbohydrate repeating disaccharide.

These stoichiometries were then used to calculate the number of repeating disaccharides that interacted with each protein, by dividing the one FGF-1 protein by the respective stoichiometry value reported. The result of these calculations was that 2.92, 1.31, and 2.01 heparin, oversulfated heparin, and oversulfated κ-carrageenan repeating disaccharides interacted with FGF-1 respectively. This method of reporting was implemented as the total length of the polysaccharide was unknown, therefore it was not possible to accurately record how many proteins bound each carbohydrate chain. The fact that the polysaccharides used were composed of long chains of repeating units reduced the effect that edges, incapable of binding the protein, had on the thermodynamic properties reported. The number of repeating disaccharides, reported above, that bound each FGF-1 protein's carbohydrate binding site(s) differed profoundly between the three polysaccharides. Doubling the number of repeating disaccharides present values calculated gave the length of carbohydrate chain that bound FGF-1 for heparin, oversulfated heparin, and oversulfated κ-carrageenan respectively, which were 5.8, 2.6, and 4.0 residues. The 5.8 heparin residues that were calculated to have interacted with FGF-1 closely correlated to previous research where labelling of both lysine and arginine residues, as well as NMR spectra collection, highlighted interactions between hexasaccharide length heparin chains and the canonical FGF-1 heparin binding site, although longer chain lengths have also been observed to bind (Bui, 2019; Ogura et al., 1999; Canales et al., 2006; DiGabriele et al., 1998). These values were also closely correlated with the results of previous ITC studies into FGF-1-heparin interactions, which indicated between 4.7 and 5.0 residues of a hexasaccharide heparin interacted with the protein (Brown et al., 2013).

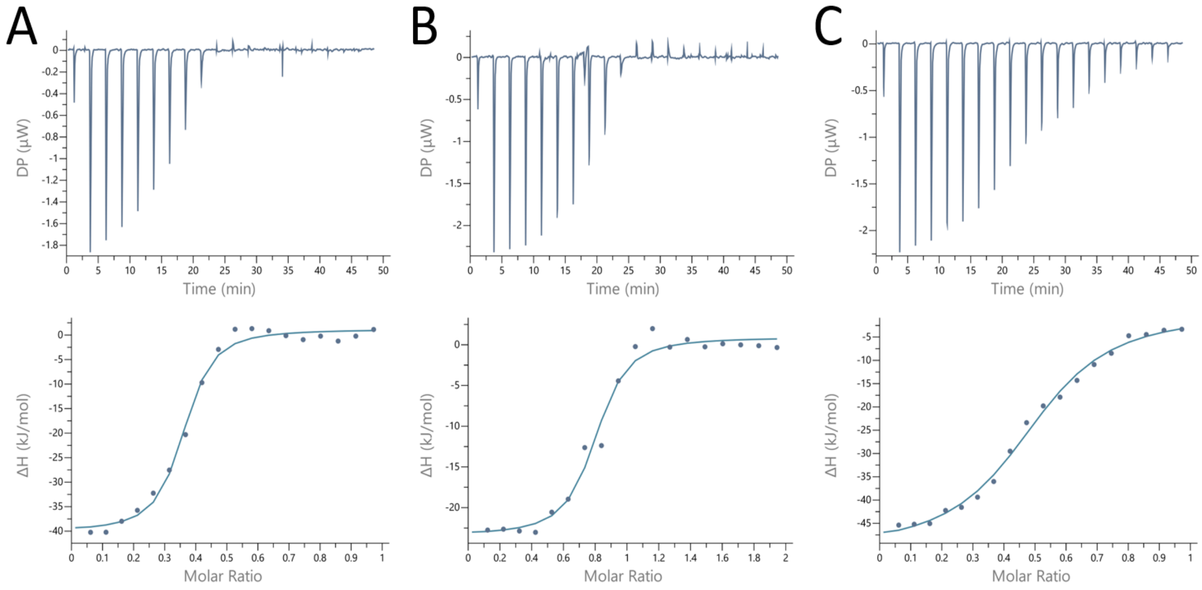


Figure 87: ITC traces, shown with the baseline subtracted, for the three carbohydrate and protein mixtures seen to interact, i.e. FGF-1 with heparin (A), Oversulfated heparin (B), and fully sulfated κ-carrageenan (C).

Coupled with the number of interacting residues is the affinity with which the carbohydrate-protein interaction occurs. The equilibrium dissociation constants (K_D) for each of the interactions present for FGF-1 with heparin, oversulfated heparin, and κ-carrageenan were all of moderate to high affinity, being within the nanomolar or low micromolar range, specifically 256nM, 514nM, and 1.46μM respectively. The observed interaction between fully sulfated κ-carrageenan and FGF-1 was novel, although not unexpected, and could be further characterised by mapping the interaction sites of the protein. The differences in the K_D values reported corresponded closely with the FGF-1 melting temperatures recorded by DSF which increased by 20.4°C, 17.8°C, and 12.5°C for heparin, oversulfated heparin and persulfated κ-carrageenan respectively.

The Gibbs free energies (ΔG) reported indicated that the three interactions observed occurred spontaneously. Similarly, the negative enthalpy (ΔH) values alluded to the energy donating nature of these interactions, Table 8, which suggested they were exothermic, and that non-covalent bond formation was favourable. This was as expected for protein and carbohydrate samples which were interacting with each other. Whilst the ΔG and ΔH parameters showed the same directionality across the three interacting combinations, the entropy (-TΔS) observed did not. Both of the systems containing interactions between FGF-1 and heparin, or fully sulfated κ-carrageenan reported positive -TΔS entropies and consequently suggested an increase in order within the system, this was, however, not seen for the interaction between FGF-1 and oversulfated heparin. Whilst it is difficult to say what this increased or decreased order resulted from, it could derive from the carbohydrate and protein molecules themselves or the water molecules that surround these components.

Comparing the ΔH values for FGF-1 interactions with heparin and oversulfated heparin it can be seen that despite the similar K_D and ΔG values reported, there is a vastly greater decrease in absolute enthalpy for heparin, -41.2kJ/mol, than for oversulfated heparin, -24.3kJ/mol. This difference is, however, compensated for by the greater value of absolute entropy for the heparin starting material, 3.58kJ/mol, than the oversulfated variant, -11.6kJ/mol. Comparison of the values reported for both the enthalpy and entropy of these two interactions suggested oversulfated heparin was more dynamic during its interactions with FGF-1. However, the presence of stronger interactions between oversulfated heparin and water, compared to those with the starting material heparin, was also a possibility. A similar comparison of the enthalpy and entropy data for heparin and oversulfated κ-carrageenan interactions highlighted their similar enthalpy values, -41.2kJ/mol and -50.6kJ/mol, respectively. However, the difference in the reported entropy values was almost as extreme that between heparin and oversulfated heparin with -TΔS for oversulfated κ-carrageenan being 17.3kJ/mol.

Examining the interactions that generated thermodynamic data it could not be pinpointed whether enthalpy or entropy were driving the changes in K_D observed, as, whilst generating similar Gibbs free energies, both varied dramatically. However, the data presented is suggestive of the interaction between FGF-1 and heparin as well as that with oversulfated κ-carrageenan being enthalpically driven, while that with oversulfated heparin was entropically driven. One potential explanation for the enthalpically driven interactions is that increased sulfation, between the heparin and oversulfated κ-carrageenan samples, resulted in increased disorder of the complex.

ITC experiments performed for CypB alone and in combination with the five carbohydrates tested were composed of the samples present at the concentrations indicated in Table 9. As with FGF-1, three of the carbohydrates, heparin, oversulfated heparin, and fully sulfated κ-carrageenan showed evidence of an interaction having occurred in a manner that generated thermodynamic data. Equally, as for FGF-1, experiments containing CypB alone and when in combination with κ-carrageenan and fully sulfated cellulose did not generate thermodynamic data, again likely suggestive of interactions which didn't result in a temperature change, from which ITC data could not be recorded. Given the DSF data reported for these five samples suggested all would interact with CypB, the same carbohydrates which gave different results for DSF and ITC when tested with FGF-1 also did so with CypB. The melting temperature decreases recorded for the heparin, oversulfated heparin, and fully sulfated κ-carrageenan samples shown to generate thermodynamic data were 2.7°C, 3.7°C, and 6.7°C respectively. By comparison, the unmodified κ-carrageenan sample generated a melting temperature decrease of 4.3°C and the fully sulfated cellulose sample one of 1.3°C.

Table 9: Protein and carbohydrate concentrations as well as the resulting thermodynamic data for interactions with CypB.

	CypB Alone	CypB + heparin	CypB + 3-O-S heparin	CypB + κ-carrageenan	CypB + sulfated κ-carrageenan	CypB + sulfated cellulose
CypB Concentration (μM)	60	60	60	60	60	60
Repeating Unit Concentration (μM)	---	300	600	300	600	300
Stoichiometry (n)	---	0.457	1.40	---	0.800	---
Interacting Repeating Disaccharides	---	2.19	0.710	---	1.25	---
Gibbs Free Energy (ΔG / kJ/mol)	---	-30.4	-32.2	---	-32.3	---
Enthalpy (ΔH / kJ/mol)	---	-12.1 ± 6.94	-10.7 ± 4.17	---	-18.4 ± 1.46	---
Entropy ($-T\Delta S$ / kJ/mol)	---	-18.3	-21.6	---	-13.8	---
Equilibrium Dissociation Constant (K_D / μM)	---	4.68 ± 8.2	2.28 ± 3.1	---	2.25 ± 0.961	---

As with the combinations of interacting carbohydrates with FGF-1, the stoichiometries of carbohydrate-CypB interactions were reported as the number of CypB proteins that bind to one carbohydrate repeating disaccharide. These values were then converted to the number of repeating disaccharides interacting with each instance of the CypB protein, and from this the number of interacting residues were determined. These calculations showed considerable variation between the three samples with 4.4, 1.4, and 2.5 residues shown to bind the protein for heparin, oversulfated heparin and fully sulfated κ-carrageenan respectively, Table 9 and Figure 88. Comparing the interacting residues data for FGF-1 and CypB, it can be seen that heparin had the greatest number of residues interact with the proteins, which were 5.8 and 4.4 respectively. Similarly, oversulfated heparin had the least residues interact, with 2.6 forming interactions with FGF-1 and 1.4 with CypB. There was also a general trend where the carbohydrates' interactions with FGF-1 did so through a greater number of residues than their corresponding interactions with CypB. The K_D values recorded for heparin, oversulfated heparin, and fully sulfated κ-carrageenan each indicated a lower affinity for CypB than FGF-1, where those for CypB were all at the lower end of the micromolar range, being 4.68μM, 2.28μM, and 2.25μM respectively, representing moderate to high affinity interactions in each case, Table 9. The relative order of affinity between the three carbohydrates and FGF-1, where heparin bound with the greatest affinity and fully sulfated κ-carrageenan the lowest, was reversed for interactions with CypB where both oversulfated carbohydrates bound with similar affinity while that of the unmodified heparin polysaccharide was lower. As with FGF-1, these three interactions appeared to follow the trend that greater melting temperature resulted in a lower K_D. However, the K_D value for the interaction between CypB and oversulfated heparin was almost identical to that with fully sulfated κ-carrageenan while the difference between the protein's melting temperatures for these two carbohydrates was more significant. Therefore, the differences in melting temperature relative to each other are not proportionate with the differences in K_D values relative to each other.

The ΔG and ΔH values recorded for CypB with all three of the carbohydrates generating thermodynamic data gave negative values indicative of spontaneous reactions occurring in an exothermic manner, with the formation of non-covalent bonds as would be expected in an interaction between carbohydrates and proteins. Conversely, all three of the carbohydrate-CypB samples gave negative -TΔS values suggesting that interaction caused an increase in disorder within the system, and as before whilst estimations could be made as to what caused this, it is a complex factor to determine conclusively. These findings would, however, correlate with the melting temperature shifts recorded from DSF where the addition of heparin, oversulfated heparin, and oversulfated κ-carrageenan appeared to destabilise the protein, which would be identified in ITC as an increase in disorder, such as has been reported. In comparison with the data recorded for FGF-1 samples, the baselines of the ITC experiments performed on CypB showed greater variation in the noise present, Figure 88. This is a factor of the ITC experiments performed which would have been optimised further, with three replicates having been measured, had time allowed.

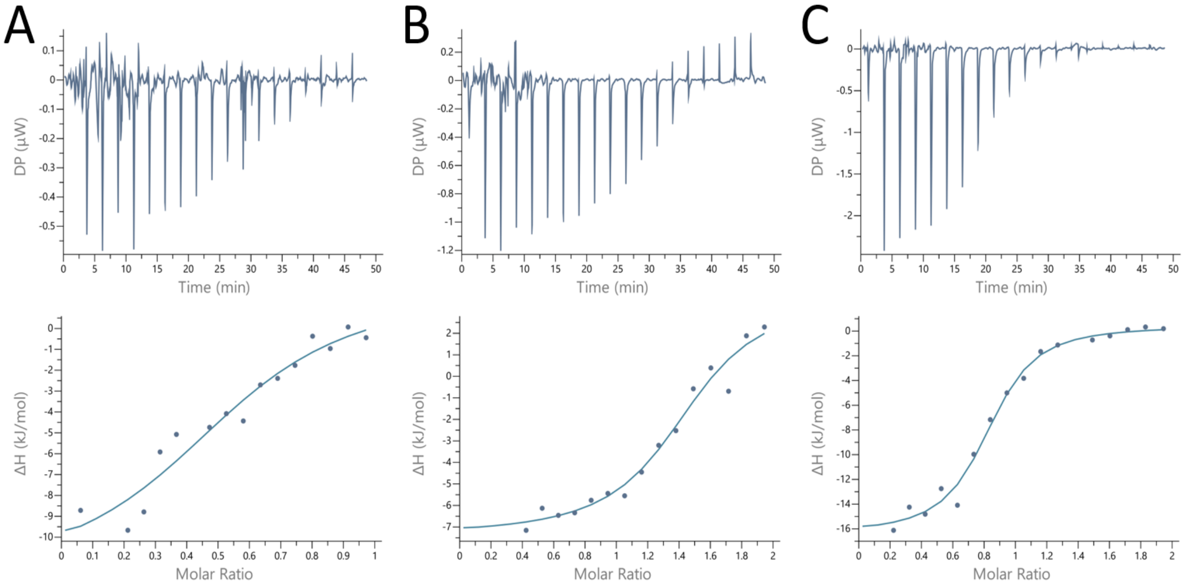


Figure 88: ITC traces, shown with the baseline subtracted, for the three carbohydrate and protein mixtures seen to interact, i.e. CypB with heparin (A), Oversulfated heparin (B), and fully sulfated κ-carrageenan (C).

ITC experiments were performed to support the results seen during DSF screening whilst also providing further thermodynamic data regarding the interactions which were taking place. A partial correlation was seen between the samples shown to interact via DSF and ITC with discrepancies for the combinations containing the two proteins and unmodified κ-carrageenan or fully sulfated cellulose which each generated a melting temperature shift in DSF experiments but did not exhibit an enthalpy transition when measured using ITC. However, while one possible explanation is that no interaction occurred, another possibility is that the interactions which did occur between these carbohydrates and proteins did so in such a manner that there was no observable temperature change, and as such no reportable data generated. Each of the carbohydrate-protein combinations that resulted in thermodynamic data being reported did so at moderate to high affinities, being no lower than 5μM for any of the combinations tested, although further work would be required to run replicates of these interactions for greater reliability.

6.3 - Discussion:

This aim for this, final, results chapter was to investigate the interactions which occur between sulfated carbohydrate samples and proteins, with four specific examples having been chosen for analysis. The proteins chosen were Fibroblast Growth Factors (FGFs) 1 and 2, owing to extensive prior research having been carried out on their interactions with heparin and other carbohydrates. Alongside this there were Cyclophilins (Cyps) B and D, of which CypB has also previously been shown to bind heparin while CypD, a protein with closely related structure, has not. Characterisation of the interactions between recombinant FGFs 1 and 2 and CypsB and D required the expression and purification of protein in sufficient quantity for the biophysical analysis techniques which were used. These protein production methods proved successful with high yield samples of pure protein generated in line with that seen in previous studies that could then be applied to the protein interaction analysis methods (Sun et al., 2015; Hanoulle et al., 2007).

Interaction screening using DSF to monitor changes in melting temperature, indicative of protein stability, for combinations of CypsB and D with variably sulfated carbohydrates and further sulfated modified samples, highlighted a number of potential interactions which had not previously been determined. The cyclophilins were both seen to interact with heparin, ι-carrageenan, and κ-carrageenan, when analysed using DSF with decreased melting temperatures upon addition. Whilst previous studies have highlighted the interaction between CypB and oversulfated heparin, they have not characterised the resulting protein stability change via DSF; hence no comparison was made with the results of the research performed here (Vanpouille et al., 2007; Hanoulle et al., 2007). Similarly, CypD-carbohydrate interactions have not previously been studied meaning that the potential interactions identified during the course of this research cannot be corroborated or compared.

In contrast to this, the interactions between FGFs 1 and 2 and sulfated carbohydrates, commonly heparin and other members of the glycosaminoglycan family, have been studied extensively. From these prior studies it was possible to compare the DSF data generated during this research project which for FGF-2 showed increased melting temperatures from 54.5°C to 83.0°C upon the addition of heparin. The FGF-2 melting temperature previously recorded was similar to that found in this research, 55.1°C, and the temperature recorded for a 1:1 carbohydrate:protein molar ratio was 77.6°C (Uniewicz et al., 2010). Comparing the effect that further heparin sulfation has on its ability to stabilise both FGF-1 and FGF-2, the research completed here gave melting temperatures similar to that of the unmodified polysaccharide. The prior research carried out indicated a relative stabilising effect of 0.8 compared to the unmodified polysaccharide, which is the same as what was found in the data collected for FGF-1 in this project, however, the relative stabilising effect with FGF-2 was closer to 1, only dropping to 0.8 for the final modified sulfation pattern sample (Uniewicz et al., 2010). The ITC data recorded on samples for a select number of interactions in this results chapter could, in most cases, not be compared to any prior results, however, FGF-1 interactions with polymeric heparin, which gave an equilibrium dissociation constant (K_D) value of 256nM, could. The K_D value reported from this research was far higher than when these interactions were examined in previous studies, which reported K_D values of 60nM and 83nM for an octasaccharide fragment and polysaccharide heparin, respectively (Xu et al., 2012; Zhao et al., 2018). The similarity of K_D values across different chain lengths has also been reported for interactions with hexasaccharide, octasaccharide and hexadecasaccharide chain lengths, where each value reported was between 1.1μM and 3.0μM, respectively (Brown et al., 2013). Whilst it is apparent from the previously reported data that chain length doesn't have a major effect on K_D, there is considerable variation in the values recorded across the range of FGF-1-heparin interactions measured. This was seen in the three studies mentioned above, with the K_D values reported appearing to form a continuum, and consequently other factors must be involved. One factor shown to induce variation in the K_D values of FGF-1 interactions with polysaccharide heparin is temperature, with 83nM affinities reported when measured at 10°C and 25°C, rising to 110nM at 30°C (Zhao et al., 2018).

A similar continuum of K_D values has previously been found for interactions of FGF-1 with the biologically relevant polysaccharide heparan sulfate. In those prior studies, heparan sulfate-FGF-1 interactions were seen to give K_D values ranging from 400nM to 8.6μM depending on the type of cell or culture medium in which they were immobilised (Rahmoune et al., 1998). The affinities of these interactions therefore range from being similar to those produced by polysaccharide heparin chains, used in this results chapter, to being of considerably lower affinity. Of the other carbohydrates potentially shown to interact with the FGFs here, only κ-carrageenan has been previously studied where an interaction was suggested but DSF was not carried out, as such no comparisons could be drawn on protein stability and melting temperature (Hoffmann and Sykes, 1993).

In conclusion, DSF screening assays for interactions between sulfated carbohydrates and FGFs 1 and 2 as well as those with Cyps B and D highlighted a number of potential combinations where carbohydrate:protein interactions have occurred. However, all of these will require further analysis by alternative interaction detecting techniques to provide additional information, such as which amino acids within the proteins are interacting with the carbohydrates. These additional interaction analysis techniques would be particularly important for the combinations that didn't generate thermodynamic data from the ITC experiments performed. There are two notable examples from the ITC data collected, fully sulfated κ-carrageenan with FGF-1 and CypB, which, having not been identified previously, these could also be further analysed via improved ITC as well as NMR. NMR analyses, using ^{15}N-labelled protein, would be able to determine the sites of interaction which could then be paired with the thermodynamics determined via ITC assays.

Chapter 7 – General Discussion

7.1 – Discussion:

The diversity of carbohydrates' natural roles coupled with the myriad applications of both their natural structures and modified variants was the motivation behind this project. The research carried out included NMR method development for improved carbohydrate structural analysis as well as for modification monitoring. This was supplemented by the development of a sulfation protocol using solely the recyclable ionic liquid, BMImCl, in the solvent system replacing DMF, or other organic solvents (Chopin et al., 2015). The sulfated carbohydrate samples produced were tested for their ability to bind four proteins, FGFs 1 and 2 and CypsB and D, using two separate biophysical methods.

The NMR method was capable of generating spectra, measured on a *T.indica* xyloglucan test case sample, with peaks present approaching, if not at, their natural line width, resulting in at least thirteen distinguishable doublets in a region of the spectra where previously it had been impossible to separate more than four or five. Comparing the method to some previously published assignments for xyloglucan from *T.indica*, highlights it superiority. One protocol used solely 1D ^{1}H NMR, whilst in another there were only two assignments evident within the 2D ^{1}H,^{13}C-HSQC region corresponding to that where the thirteen discernible peaks were present in the spectra measured for this project, when using the same carbohydrate. Therefore the results of the other research were similar to the starting spectra recorded during this project and shown in Figure 24 (Tuomivaara et al., 2015; Ding et al., 2016). The NMR spectrum resolution previously reported for xyloglucan samples appears to be commonplace for the assignment of carbohydrates, where structural characterisations of non-xyloglucan samples contained spectra composed of equally low resolution peaks (Komatsu and Kikuchi, 2013; Arruda et al., 2015). Alongside the increased resolution of the spectra recorded, a wider variety of experiments were employed than are commonly used for carbohydrate samples. In these structural characterisation experiments, 2D ^{1}H,^{13}C-HSQC-TOCSYs proved to be particularly informative for assignment, as did 2D ^{1}H,^{13}C-HMBC spectra for assigning linkages, however, these have been little mentioned in the carbohydrate structure characterisation methods previously reported.

Alongside analysis of the test case xyloglucan sample detailed in chapter four, this method was applied to a series of eight different xyloglucan samples which had been sourced from separate commercial suppliers that were then characterised structurally. The method was also applied to a series of sulfated samples, where problems during the sample preparation phase meant that only two carbohydrates could be measured at a resolution approaching the natural linewidth. However, before these high resolution structural analyses could take place, the sulfation protocol to be used required further optimisation to reduce the amount of organic solvent present (Chopin et al., 2015; Gericke et al., 2009; Wang et al., 2007). The protocol developed, when applied to either of the two carrageenan samples, or cellulose, indicated persulfation had been successful in a manner similar to that observed for the methods detailed in publications containing their prior assignments (Kowsaka et al., 1991; de Araújo et al., 2013). However, in each of these cases, potentially protracted or loss generating steps were required with carrageenan samples having to be reduced before the process of sulfation could take place (Kowsaka et al., 1991; de Araújo et al., 2013). The method used in this project could, therefore, be seen as advantageous for the production of sulfated carbohydrates, which were either gel forming or water insoluble, because they did not require either a pre-sulfation reduction step or a lengthy dispersion step, which would only have added time and, potentially, caused partial loss of sample. Whilst highly successful for the conversion of carrageenan and cellulose samples to their persulfated variants, this method was also shown to sulfate the far more complicated xyloglucan repeating units. However, while observable sulfation had occurred on the polysaccharide xyloglucan samples, which previously hadn't been reported, this was to a far lesser extent than that of cellulose or the carrageenans, with the complete sulfation seen for those samples not attained. The final carbohydrate tested, heparin, was capable of further sulfation, however, at longer experiment times the proportion of iduronic acid and glucosamine position three hydroxyl groups that were sulfated had dropped again to near zero having peaked after two hours.

With these sulfated samples produced and recombinant proteins expressed using conventional methods, they were combined to generate samples for carbohydrate-protein interaction screening using a DSF protocol. From these experiments several, previously unknown interactions were identified, including those of the FGFs with κ- and ι-carrageenan starting materials and persulfated variants. It was possible to identify a limited series of samples which should be studied with further biophysical analysis techniques. Of these, the interaction between FGF-1 and heparin gave a 256nM K_D, indicating a considerably lower interaction affinity than that of 60nM which was previously observed (Xu et al., 2012). However, these K_D values were obtained from interactions of different chain length heparin samples, one being a polysaccharide and the other an octasaccharide, which may have affected the interaction, and may be, at least partially, responsible for the difference in observed measurements. Regardless, the K_D value reported from this work is within the range of those reported by other studies (Xu et al., 2012; Zhao et al., 2018; Brown et al., 2013). Both the persulfated heparin and carrageenan samples gave higher K_D values, 514nM and 1.48μM respectively, than heparin for interactions with FGF-1 however, each of the K_D values reported indicated a moderate to high affinity interaction with FGF-1. The interaction between FGF-1 and persulfated carrageenan was not one that had previously been reported, although it could have been predicted from the interactions previously observed between FGFs 1 and 2 with λ-carrageenan samples (Sun et al., 2019). Whilst the three samples discussed above showed evidence of interactions with FGF-1 in ITC analyses which corroborated the DSF results, those containing κ-carrageenan or persulfated cellulose, and either FGF-1 or CypB, did not give thermodynamic data, which could either be suggestive of their having been no interaction when studied via ITC. However, the clear change in melting temperature when measured using DSF, as seen in Figure 84 and Figure 86, respectively, is more suggestive that an interaction formed but did not result in a temperature change that could be monitored by ITC. However, as with FGF-1, the interactions between CypB and heparin, persulfated heparin or persulfated carrageenan could be observed when examined using both DSF and ITC. In these analyses each of the three samples had affinities in the low micromolar range, 4.68μM, 2.28μM, and 2.25μM respectively, potentially indicating that carbohydrate-CypB interactions are driven mostly by electrostatic forces between positively charged patches on the protein surface and the highly negatively charged carbohydrates. Of these interactions, only that of CypB and persulfated heparin had been previously reported (Vanpouille et al,. 2007), potentially offering a target for future work characterising the nature of the CypB-fully sulfated carrageenan interaction.

7.2 – Future Work:

As has been highlighted in the discussion sections within each results chapter, the research completed in this project has indicated areas which could be built upon by further experimentation and method development. The first these is the continued refinement of the NMR based structure characterisation method that was developed in chapter four. Whilst the spin fractionation and chromatographic methods used were capable of separating some repeating units from one another, alternatives are available which could improve separation to provide individual repeating units in distinct fractions. One such method, which has been used previously, is high performance anion-exchange chromatography, which when coupled with a pulsed amperometric detector, would enable easy detection (Tuomivaara et al., 2015). This detection system would also negate the need 1D ^{1}H NMR spectra as the location of the different fractions would be known, hence it would be possible to skip to the high resolution NMR spectrum collection phase, representing considerable time saving.

Alternatively, the implementation of diffusion NMR spectrscopy, could enable the differentiation of signals from a sample mixture without having to perform any physical separation techniques. These experiments measure diffusion within a sample between a pair of gradient pulses designed to encode and decode diffusion information respectively with varying pulse powers used depending on the sample being detected (Groves, 2017). 2D diffusion NMR spectra, plotting the ^{1}H peaks of the compounds present within a mixture on one axis and the diffusion coefficient on the other has been used to separate many carbohydrates from one another. Measurement of these samples using optimised NMR experiment parameters for a series of compounds of known molecular weight can also serve to form a calibration curve, which, alongside the diffusion coefficient of the target sample, could provide the target compound's molecular weight and, in certain cases, has even been able to determine the anomer present (Viel et al., 2003; Politi et al., 2006; Yamanoi et al., 2017). The 2D diffusion NMR experiments previously used for the analysis of carbohydrate samples have recently been improved to generate, 3D spectra, where the ^{1}H and ^{13}C axes for the carbohydrate 2D ^{1}H,^{13}C-HSQC spectra (as recorded in this project) were supplemented with one for the diffusion coefficient (Vitorge and Jeanneat, 2006; Brand et al., 2008; Dal Poggetto et al., 2018).

Alongside the future work which could be performed for chapter four, there are a number of elements for chapter five which would also provide highly informative data. The first of these is the hydrolysis of a xyloglucan sample to its repeating unit oligosaccharides followed by sulfation using the method developed in chapter five. Subsequent analysis of the resulting sample using the structure characterisation method developed in chapter four would then provide a comprehensive series of assignments. From these, and if a sulfation time course were performed, it would be possible to define which of the hydroxyl groups present within the xyloglucan repeating unit was preferentially sulfated and determine whether this follows the typical pattern for the relative reactivities of hydroxyl groups. Having completed these experiments on the xyloglucan sample it would be important to do the same for the heparin and cellulose samples which required further structural analysis on the basis of the initial screening NMR experiments and ATR-FITR data.

Despite the optimisation of a sulfation method using pyridine sulfur trioxide with the only solvent being BMImCl generating persulfated samples for most of the carbohydrates tested, there remain issues to improve. Therefore, further development of the method would be important. This is due to the high concentrations of pyridine sulfur trioxide complex which were required for the experiments performed in their present set up compared to those using the traditional organic solvents. Such optimisation would need to focus on reducing the amount of this compound present, however, another focus for the development of this reaction could be to incorporate an additional biorenewable solvent, present with the ionic liquid, to determine what effect that would have on the reactant concentrations required. The addition of such a solvent could also lead to additional challenges however, as it would need to be anhydrous for the sulfation reactions to proceed successfully, rather than generating sulfuric acid from the breakdown of the sulfur trioxide complex.

There are also a number of elements within chapter six which could be turned into a focus for future work. One is the collection of NMR data to map the interaction sites of the carbohydrate on the protein for the interacting pairs identified by both ITC and DSF, i.e. CypB and FGF-1 with heparin, persulfated heparin, and persulfated κ-carrageenan. This would require the generation of ^{15}N-labelled protein. Alternative techniques to identify interacting residues could also be included in any further work to be carried out. One such exemplar technique is protect and label, which has previously been used to identify the heparin interaction sites on FGF family members (Xu et al., 2012). Experiments targeted to the identification of interacting residues would need to be supplemented by further ITC data to make any conclusions drawn more reliable. Additionally, the use of other further biophysical analysis techniques, such as surface plasmon resonance or bio-layer interferometry, to monitor competition between carbohydrates during their interactions with proteins would be an important progression of the research detailed in this thesis. It would also be worthwhile to expand the proteins used to incorporate FGF-2 and CypD as well as the carbohydrates to attempt to identify and characterise as many new interacting pairs as possible from the screening data generated by DSF. In addition to these there are a wealth of factors which could be studied to progress this research from that point, such as altering the cation present or the modifying group on the carbohydrate, however, they will not be discussed in any detail here.

Chapter 8 - References

Ahmad NH, Mustafa S, and Che Man YB (2015). Microbial Polysaccharides and Their Modification Approaches: A Review. International Journal of Food Properties 18, pp.332-347.

Ahmed-Belkacem A, Colliandre L, Ahnou N, Nevers Q, Gelin M, Bessin Y, Brillet R, Cala O, Douguet D, Bourguet W, Krimm I, Pawlotsky JM, and Guichou JF (2016). Fragment-based discovery of a new family of non-peptidic small molecule cyclophilin inhibitors with potent antiviral activities. Nature Communications 7, pp.12777-12777.

Allain F, Vanpouille C, Carpentier M, Slomianny M-C, Durieux S, and Spik G (2002). Interaction with glycosaminoglycans is required for cyclophilin B to trigger integrin-mediated adhesion of peripheral blood T lymphocytes to extracellular matrix. Proceedings of the National Academy of Sciences of the USA, 99, pp.2714-2719.

Alonso JL, Lozoya MA, Peña I, López JC, Cabezas C, Mata S, and Blanco S (2014). The conformational behaviour of free D-glucose – at last. Chemical Science 5, pp.515-522.

Andersen B and Grønlund F (1979). High Pressure Kinetics of Glucose Mutarotation Studied by Optical Rotation. acta Chemica Scandinavia A 33, pp.275-280.

Arruda IRS, Albuquerque PBS, Santos GRC, Silva AG, Mourão PAS, Correia MTS, Vincente AA, and Carneiro-da-Cunha MG (2015). Structure and rheological properties of a xyloglucan extracted from *Hymenaea courbaril* var. *courbaril* seeds. International Journal of Biological Macromolecules 73, pp.31-38.

Ashline DJ, Lapadula AJ, Liu Y-H, Lin M, Grace M, Pramanik B, and Reinhold VN (2007). Carbohydrate Structural Isomers Analyzed by Sequential Mass Spectrometry. Analytical Chemistry 79, pp.3830-3842.

Astronomo R.D., and Burton D.R. (2010). Carbohydrate vaccines: developing sweet solution to sticky situation? Nature Reviews Drug Discovery 9, pp.308-324.

Avci FY, and Kasper DL (2020). How Bacterial Carbohydrates Influence the Adaptive Immune System. Annual Review of Immunology 28, pp.107-130.

Banfalvi G (2021). Prebiotic Pathway from Ribose to RNA Formation. International Journal of Molecular Sciences 22, 3857.

Barbar E (2004). NMR Characterization of Partially Folded and Unfolded Conformational Ensembles of Proteins. Biopolymers (Peptise Science) 51, pp.191-207.

Baum A, Dominiak M, Vidal-Melgosa S, Willats WGT, Søndergaard KM, Hansen PW, Meyer AS, and Mikkelsen JD (2017). Prediction of Pectin Yied and Quality by FTIR and Carbohydrate Microarray Analysis. Food Bioprocess Technology 10, pp.143-154.

BeMiller JN (2019). Polysaccharides: Occurrence, Structures, and Chemistry, in Carbohydrate Chemistry for Food Scientists. Duxford, UK: Woodhead Publishing, pp.75-101.

Beni S, Limtiaco JFK, and Larive CK (2011). Analysis and characterization of heparin impurities. Analytical and Bioanalytical Chemistry 399, pp.527-539.

Berthomieu C and Hienerwadel R (2009). Fourier transform infrared (FTIR) spectroscopy. Photosynthesis Research 101, pp.157-170.

Bertoft E (2013). On the Building Block and Backbone Concepts of Amylopectin Structure. Cereal Chemistry 90, pp.294-311.

Bianchera A, Salomi E, Pezzanera M, Ruwet E, Bettini R, and Elviri L (2014). Chitosan Hydrogels for Chitosan Sulphate Controlled Release: An Analytical Characterization. Journal of Analytical Methods in Chemistry 2014, 808703.

Binley JM, Ban Y-EA, Crooks ET, Eggink D, Osawa K, Schief WR, and Sanders RW (2010). Role of Complex Carbohydrates in Human Immunodeficiency Virus Type 1 Infection and Resistance to Antibody Neutralization. Journal of Virology 84, pp.5637-5655.

Blaum BS, Deakin JA, Johansson CM, Herbert AP, Barlow PN, Lyon M, and Uhrín D (2010). Lysine and Arginine Side Chains in Glycosaminoglycan-Protein Complexes Investigated by NMR, Cross-Linking, and Mass Spectrometry: A Case Study of the Factor H – Heparin Interaction. Journal of the American Chemical Society 132, pp.6374-6381.

Bonvin AMJJ, Boelens R, and Kaptein R (2005). NMR analysis of protein interactions. Current Opinion in Structural Biology 9, pp.501-508.

Bousfield TW, Pearce KPR, Nyamini SB, Angelis-Dimakis A, and Camp JE (2019). Synthesis of amides from acid chlorides and amines in the bio-based solvent Cyrene™. Green Chemistry 21, pp.3675-3681.

Brand T, Cabrita EJ, Morris GA, Günther R, Hofmann H-J, and Berger S (2008). Residue specific NH exchange rates studied by NMR diffusion experiments. Journal of Magnetic Resonance 187, pp.97-104.

Brodaczewska N, Košťálová Z, and Uhrín D (2018). (3, 2)D ^{1}H, ^{13}C BIRD^{r,X}-HSQC-TOCSY for NMR structure elucidation of mixtures: application to complex carbohydrates. Journal of Biomolecular NMR 70, pp.115-122.

Brown A, Robinson CJ, Gallagher JT, and Blundell TL (2013). Cooperative Heparin-Mediated Oligomerization of Fibroblast Growth Factor-1 (FGF1) Precedes Recruitment of FGFr2 to Ternary Complexes. Biophysical Journal 104, pp.1720-1730.

Brunchi C-E, Avadanei M, Bercea M, and Morariu S (2019). Chain conformation of xanthan in solution as influenced by temperature and salt in addition. Journal of Molecular Ligands 287, 111008.

Bucior I and Burger MM (2004). Carbohydrate-carbohydrate interactions in cell recognition. Current Opinion in Structural Biology 14, pp.631-637.

Bui TP (2019). Selective labelling of arginine residues in protein sulfated glycosaminoglycan binding sites. PhD Thesis, University of Liverpool.

Bukrinsky MI (2002). Cyclophilins: unexpected messengers in intercellular communications. Trends in Immunology 23, pp.323-325.

Byrne DP, Li Y, Ramakrishnan K, Barsukov IL, Yates EA, Eyers CE, Papy-Garcia D, Chantepie S, Pagadaa V, Liu J, Wells C, Drewry DH, Zuercher WJ, Berry NG, Fernig DG, and Eyers PA (2018). New tools for carbohydrate sulfation analysis: heparin sulfate 2-O-sulfotransferase (HS2ST) is a target for small-molecule protein kinase inhibitors. Biochemical Journal 475, pp.2417-2433.

Camp JE (2018). Bioavailable Solvent Cyrene: Synthesis, Derivatization, and Applications. ChemSusChem 11, pp.3048-3055.

Campo VL, Kawano F, Braz da Silva Jr. D, and Carvalho I (2009). Carrageenans: Biological properties, chemical modifications and structural analysis – A review. Carbohydrate Polymers 77, pp.167-180.

Canales A, Lozano R, Lopez-Mendez B, Angulo J, Ojeda R, Nieto PM, Martin-Lomas M, Gimenez-Gallego G, and Jimenez-Barbero J (2006). Solution NMR structure of a human FGF-1 monomer, activated by a hexasaccharide heparin- analogue. FEBS Journal 273, pp.4716-4727.

Capila I and Linhardt RJ (2002). Heparin-Protein Interactions. Angewandte Chemie International Edition 41, pp.390-412.

Carlsson P, Presto J, Spillmann D, Lindahl U, and Kjellen L (2008). Heparin/Heparan Sulfate Biosynthesis Processive Formation of N-sulfated Domains. The Journal of Biological Chemistry 283, pp.20008-20014.

Carmona JA, Calero N, Ramírez P, and Muñoz J (2017). Rheology and Structural Recovery Kinetics of an Advanced Xanthan Gum with Industrial Application. Applied Rheology 27, 25555.

Cassinelli G. and Naggi A. (2016). Old and new applications of non-anticoagulant heparin. International Journal of Cardiology 212S1, S14-S21.

Castillo-Valdivia M.E., López-Montes A., Espejo T., Vílchez J.L., and Blanc R. (2014). Identification of starch and determination of its botanical source in ancient manuscripts by MEKC-DAD and LDA. Microchemical Journal 112, pp.75-81.

Cataldi TRI, Campa C, and De Benedetto GE (2000). Carbohydrate analysis by high performance anion-exchange chromatography with pulsed amperometric detection: The potential is still growing. Fresenius Journal of Analytical Chemistry 368. pp.739-758.

Chandel NS (2021). Glycolysis. Cold Spring Harbour Perspectives in Biology 13, a040535.

Chappell EP and Liu J (2013). Use of biosynthetic enzymes in heparin and heparan sulfate synthesis. Bioorganic and Medicinal Chemistry 21, pp.4786-4792.

Charrier B., Rabillé H., and Billoud B. (2019). Gazing at Cell Wall Expansion under a Golden Light. Trends in Plant Science 24, pp.130-141.

Chen Y, Lin L, Agyekum I, Zhang X, St Ange K, Yu Y, Zhang F, Liu J, Amster IJ, and Linhardt RJ (2017). Structural Analysis of Heparin-Derived 3-O-Sulfated Tetrasaccharides: Antithrombin Binding Site Variants. Journal of Pharmaceutical Sciences 106, pp.973-981.

Chiu L-Y, Hung K-W, Tjong S-C, Chiang Y-W, and Sue S-C (2014). NMR characterization of the electrostatic interaction of the basic residues in HDGF and FGF2 during heparin binding. Biochimica et Biophysica Acta 1844, 1851-1859.

Choay J, Petitou M, Lormeau JC, Sinay P, Casu B, and Gatti G (1983). Structure -Activity Relationship in Heparin: A Synthetic Pentasaccharide with High Affinity for Antithrombin III and Eliciting High Anti Factor Xa Activity. Biochemical and Biophysical Research Communications 116, pp.492-499.

Chopin N, Sinquin C, Ratiskol J, Zykwinska A, Weiss P, Cérantola S, Le Bideau J, and Colliec-Jouault S (2015). A Direct Sulfation Process of a Marine Polysaccharide in Ionic Liquid. BioMed Redearch International 2015, 508656.

Cilliers L., and Retief F.P. (2008). Bees, Honey and Health in Antiquity. Akroterion 53, pp.7-9.

Ciulli A (2013). Biophysical Screening for the Discovery of Small-Molecule Ligands. Methods in Molecular Biology 1008, pp.357-388.

Claridge TDW (2016). *High-Resolution NMR Techniques in Organic Chemistry (3rd Edition)*. Elsevier: Amsterdam.

Cobb BA, and Kasper DL (2005). Coming of age: carbohydrates and immunity. European Journal of Immunilogy 35, pp.352-356.

Corsaro C, Mallamace D, Łojewska J, Mallamace F, Pietronero L, and Missori M (2013). Molecular degradation of ancient documents revealed by ^{1}H HR-MAS NMR spectroscopy. Scientific Reports 3, 2896; DOI:10.1038/srep02896.

Cosgrove DJ (2005). Growth of the Plant Cell Wall, Nature Reviews Molecular Cell Biology 6, pp.851-861.

Coskun O (2016). Separation Techniques: Chromatography. North Clinics of Istanbul 3, pp.156-160.

Costa S, Almeida A, Castro A, and Domingues L (2014). Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. Frontiers in Microbiology 5, doi: 10.3389/fmicb.2014.00063

Cumpstey I (2013). Chemical Modification of Polysaccharides. ISRN Organic Chemistry 2013, 417672.

Dadhich P, Das B, and Dhara S (2015). Microwave assisted rapid synthesis of N-methylene phosphonic chitosan via Mannich-type reaction. Carbohydrate Polymers 133, pp.345-352.

Dal Poggetto G, Castañar L, Foroozandeh M, Kiraly P, Adams RW, Morris GA, and Nilsson M (2018). Unexploited Dimension: New Software for Mixture Analysis by 3D Diffusion-Ordered NMR Spectroscopy. Analytical Chemistry 90, 13695-13701.

Dam TK and Brewer CF (2002). Thermodynamic studies of Lectin-Carbohydrate Interactions by Isothermal Titration Calorimetry. Chemical Reviews 102, pp.387-429.

de Araújo CA, Noseda MD, Cipriani TR, Gonçalves AG, Duarte MER, and Ducatti DRB (2013). Selective sulfation of carrageenans and the influence of sulfate regiochemistry on anticoagulant properties. Carbohydrate Polymers 91, pp.483-491.

de la Concepción JG, Martínez RF, Cintas, P, and Babiano R (2020). Mutarotation of aldoses: Getting a deeper knowledge of a classic equilibrium enabled by computational analyses. Carbohydrate Research 490, 107964.

de la Fuente JM and Penadés S (2004). Understanding carbohydrate-carbohydrate interactions by means of glyconanotechnology. Glycoconjugate Journal 21, pp.149-163.

Deepa SS, Umehara Y, Higashiyama S, Itoh N, and Sugahara K (2002). Specific Molecular Interactions of Oversulfated Chondroitin Sulfate E with Various Heparin-binding Growth Factors. Journal of Biological Chemistry 277, pp.43707-43716.

del Carmen Fernández-Alonso M, Díaz D, Berbis MÁ, Marcelo F, Cañada J, and Jiménez-Barbero J (2012). Protein-Carbohydrate Interactions Studied by NMR: From Molecular Recognition to Drug Design. Current Protein and Peptide Science 13, pp.816-830.

Devaraj S, Jagdhane RC, and Shashidhar MS (2009). Relative reactivity of hydroxyl groups in inositol derivatives: role of metal ion chelation. Carbohydrate Research 344, pp.1159-1166.

Devlin A, Mauri L, Guerrini M, Yates EA, and Skidmore MA (2019b). bioRxiv, doi:10.1101/744532.

Devlin A, Mycroft-West CJ, Turnbul JE, Guerrini M, Yates EA, and Skidmore MA (2019a). bioRxiv, doi: 10.1101/538074.

DiGabriele AD, Lax I, Chen DI, Svahn CM, Jaye M, Schlessinger J, and Hendrickson WA (1998). Structure of a heparin-linked biologically active dimer of fibroblast growth factor. Nature 393, pp.812-817.

Dimakos V and Taylor MS (2018). Site-selective Functionalisation of Hydroxyl Groups in Carbohydrate Derivatives. Chemical Reviews 118, pp.11457-11517.

Ding HH, Cui SW, Goff HD, Chen J, Guo Q, and Wang Q (2016). Xyloglucans from flaxseed kernel cell wall: Structural and conformational characterisation. Carbohydrate Polymers 151, pp.538-545.

Dispenza C, Todaro S, Bulone D, Sabatino MA, Ghersi G, San Biagio PL, and Lo Presto C (2017). Physico-chemical and mechanical characterization of in-situ forming xyloglucan gels incorporating a growth factor to promote cartilage reconstruction. Material Science and Engineering C 70, pp.745-752.

Doyle ML (1997).Characterization of binding interactions by isothermal titration calorimetry. Current Opinion in Biotechnology 8, pp.31-35.

Drohan D, Tziboula A, McNUlty D, and Horne DS (1997). Milk protein-carrageenan interactions. Food Hydrocolloids 11, pp.101-107.

Duus JØ, Gotfredsen CH, and Bock K (2000). Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations. Chemical Reviews 100, pp.4589-4614.

Dvořáčková E, Šnóblová M, and Hrdlička P (2014). Carbohydrate analysis: From sample preparation to HPLC on different stationary phases coupled with evaporative light-scattering detection. Journal of Separation Science 37, pp.323-337.

Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, and Wishart DS (2019). NMR Spectroscopy for Metabolomics Research. Metabolites 9, 123.

Esko JD and Lindahl U (2001). Molecular diversity of heparan sulfate. The Journal of Clinical Investigation 108, pp.169-173.

Evershed R.P., Berstan R., Grew F., Copley M.S., Charmant A.J.H., Barham E., Mottram H.R., and Brown G. (2004). Nature 432, pp.35-36.

Falconer RJ (2016). Applications of isothermal titration calorimetry – the research and technical developments from 2011 to 2015. Journal of Molecular Recognition 29, pp.504-515.

Fenn LS and McLean JA (2010). Structural resolution of carbohydrate positional structural isomers based on gas-phase ion mobility-mass spectrometry. Physical Chemistry Chemical Physics 13, pp.2196-2205.

Fernández C, Hattan CM, and Kerns RJ (2006). Semi-synthetic heparin derivatives: chemical modifications of heparin beyond chain length, sulfate substitution pattern and N-sulfo/N-acetyl groups. Carbohydrate Research 341, pp.1253-1265.

Ferreira VP, Pangburn MK, and Cortés C (2010). Complement control protein factor H: the good, the bad, and the inadequate. Molecular Immunology 47, pp.2187-2197.

Fischer H.E. (1989). Origin of the 'Weisse Schlesische Rübe' (white Silesian beet) and resynthesis of sugar beet. Euphytica 41, pp.75-80.

Fox SC, Li B, Xu D, and Edgar KJ (2011). Regioselective Esterification and Etherification of Cellulose: A Review. Biomacromolecules 12, pp.1956-1972.

Freire E, Mayorga OL, and Straume M (1990). Isothermal Titration. Analytical Chemistry 62, pp.950-959.

Fu Q, Liang T, Li Z, Xu X, Ke Y, Jin Y, and Liang X (2013). Separation of carbohydrates using hydrophilic interaction liquid chromatography. Carbohydrate Research 379, pp.13-17.

Fuchs B, Ellencweig A, Tartakovsky E, and Aped P (1986). Solvent Polarity and the Anomeric Effect. Angewandte Chemie International Edition 25, pp.287-298.

Gabius H-J, André S, Jiménez-Barbero J, Romero A, and Solís D (2011). From lectin structure to functional glycomics: principles of the sugar code. Trends in Biochemical Sciences 36, pp.298-313.

Galonić DP and Gin DY (2007). Chemical glycosylation in the synthesis of glycoconjugate antitumour vaccines. Nature 446, pp.1000-1007.

Gao K, Oerlemans R, and Groves MR (2020). Theory and applications of differential scanning fluorimetry in early stage drug discovery. Biophysical Reviews 12, pp.85-104.

García B, Fernández-Vega I, García-Suárez O, Castañón S, and Quirós LM (2014). The Role of Heparan Sulfate Proteoglycans in Bacterial Infections. Journal of Medical Microbiology and Diagnosis 3, 157.

Geng X and Wang L (2008). Liquid chromatography of recombinant proteins and protein drugs. Journal of Chromatography B 866, pp.133-153.

Gericke M, Liebert T, and Heinze T (2009). Interaction of Ionic Liquids with Polysaccharides, 8 – Synthesis of Cellulose Sulfates Suitable for Polyelectrolyte Complex Formation. Macromolecular Bioscience 9, pp.343-353.

Gerwert K and Kötting C (2010). Fourier Transform Infrared (FTIR) Spectroscopy. In: *Encyclopaedia of Life Sciences (ELS)* (pp.1-8). John Wiley and Sons: Chichester.

Glish GL and Vachet RW (2003). The basics of mass spectrometry in the twenty-first century. Nature Reviews Drug Discovery 2, pp.140-150.

Gómez-Ordóñez E and Rupérez P (2011). FTIR-ATR spectroscopy as a tool for polysaccharide identification in edible brown an red seaweeds. Food Hydrocolloids 25, pp.1514-1520.

Gray CJ, Schindler B, Migas LG, Pičmanová M, Allouche AR, Green AP, Mandal S, Motawia MS, Sánchez-Pérez R, Bjarnholt N, Møller BL, Rijs AM, Barran PE, Compagnon I, Eyers CE, and Flitsch SL (2017). Bottom-UP Elucidation of Glycosidic Bond Stereochemistry. Analytical Chemistry 89, pp.4540-4549.

Grembecka M (2015). Sugar alcohols – their role on the modern world of sweeteners: a review. European Food Research and Technology 241, pp.1-14.

Groves P (2017). Diffusion ordered spectroscopy (DOSY) as applied to polymers. Polymer Chemistry 8, pp.6700-6708.

Guimond SE, Rudd TR, Skidmore MA, Ori A, Gaudesi D, Cosentino C, Guerrini M, Edge R, Collison D, McInnes E, Torri G, Turnbull JE, Fernig DG, and Yates EA (2009). Cations Modulate Polysaccharide Structure To Determine FGF-FGFR Signaling: A comparison of Signaling and Inhibitory Polysaccharide Interactions with FGF-1 in Solution. Biochemistry 48, pp.4772-4779.

Ha S, Gao J, Tidor B, Brady JW, and Karplus M (1991). Solvent Effect on the Anomeric Equilibrium in D-Glucose: A Free Energy Simulation Analysis. Journal of the American Chemical Society 113, pp.1553-1557.

Hagan WR (2006). EPR Spectroscopy as a probe of metal centres in biological systems. Dalton Transactions 2006, pp.4415-4434.

Hanafusa K., Hotta T., and Iwabuchi K. (2020). Glycolipids: Linchpins in the Organization and function of Membrane Microdomains. Frontiers in Cell and Developmental Biology 8, 589799.

Hanoulle X, Melchior A, Sibille N, Parent B, Denys A, Wieruszeski J-M, Horvath D, Allain F, Lippens G, and Landrieu I (2007). Structural and Functional Characterization of the Interaction between Cyclophilin B and a Heparin-derived Oligosaccharide. Journal of Biological Chemistry 282, pp.34148-34158.

Harris RK, Becker ED, De Menezes SMC, Goodfellow R, and Granger P (2001). NMR Nomenclature. Nuclear Spin Properties and Conventions for Chemical Shifts. Pure and Applied Chemistry 73, pp.1795-1818.

Hoffman R (1993). Carrageenans inhibit growth-factor binding. Biochemistry Journal 289, pp.331-334.

Hoffmann R and Sykes D (1993). Inhibition of binding of basic fibroblast growth factor to low and high affinity receptors by carrageenans. Biochemical Pharmacology 45, pp.2348-2351.

Hofmann J, Hahm HS, Seeberger PH, and Pagel K (2015). Identification of carbohydrate anomers using ion mobility-mass spectrometry. Nature 526, pp.241-244.

Hotchkiss AT, Nuñez A, Strahan GD, Chau HK, White AK, Marais JPJ, Hom K, Vakkalanka MS, Di R, Yam KL, and Khoo C (2015). Cranberyy Xyloglucan Structure and Inhibition of *Escherichia coli* Adhesion to Epithelial Cells. Journal of Agricultural and Food Chemistry 63, pp.5622-5633.

Houser J, Kozmon S, Mishra D, Hammerová Z, Wimmerová M, and Koča J (2020). The CH-π Interaction in Protein-Carbohydrate Binding: Bioinformatics and In Vitro Quantification. Chemistry – A European Journal 26, pp.10769-10780.

Hricovíni M, Guerrini M, Bisio A, Torri G, Petitou M, and Casu B (2001). Conformation of heparin pentasaccharide bound to antithrombin III. Biochemical Journal 359, pp.265-272.

Hsieh P-H, Thieker DF, Guerrini M, Woods RJ, and Liu J (2016). Uncovering the Relationship between Sulphation Patterns and Conformation of Iduronic Acid in Heparan Sulphate. Scientific Reports 6, 29602.

Hsieh YSY and Harris PJ (2009). Xyloglucans of Monocotyledons Have Diverse Structures. Molecular Plant 2, pp.943-965.

Huang L, Shen M, Morris GA, and Xie J (2019). Sulfated polysaccharides: Immunomodulation and signaling mechanimsms. Trends in Food Science & Technology 92, pp.1-11.

Huang X, Jiang B, Chen J, Zhang T, and Zheng L (2021). Enzymatic Preparation of Non-Reducing Oligosaccharides from Maltodextrins and Nigerooligosaccharides. Starch 73, 2100028.

Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, and Woolfson DN (2015). Carbohydrate – Aromatic Interactions in Proteins. Journal of the American Chemical Society 137, pp.15152-15160.

Hughes A, Meneghetti M, Huang T-Y, Hung S-C, Elli S, Guerrini M, Rudd T, Lima M, and Yates EA (2017). Investigating the relationship between temperature, conformation and calcium binding in heparin model oligosaccharides. Carbohydrate Research 438, pp.58-64.

Iijima K, and Hashizume M. (2015). Application of Polysaccharides as Structural Materials. Trends in Glycoscience and Glycotechnology 27, pp.67-79.

Imperiali B (2019). Bacterial carbohydrate diversity – a Brave New World. Current Opinion in Chemical Biology 53, 1-8.

Irwin DJ and Hurtig HI (2018). The Contribution of Tau, Amyloid-Beta and Alpha-Synuclein Pathology to Dementia in Lewy Body Disorders. Journal of Alzheimer's Disease and Parkinsonism 8, doi: 10.4172/2161-0460.1000444.

Isik M, Sardon H, and Mecerreyes D (2014). Ionic Liquids and Cellulose: Dissolution, Chemical Modification and Preparation of New Cellulosic Materials. International Journal of Molecular Sciences 15, pp.11922-11940.

Jakobek L (2015). Interactions of polyphenols with carbohydrates, lipids and proteins. Food Chemistry 175, pp.556-567.

Jia Z, Cash M, Darvill AG, and Yok WS (2005). NMR characterization of endogenously O-acetylated oligosaccharides isolated from tomato (Lycopersicon esculentum) xyloglucan. Carbohydrate Research 340, pp.1818-1825.

Jiao G, Yu G, Zhang J, and Ewart HS (2011). Chemical Structures and Bioactivities of Sulfated Polysaccharides from Marine Algae. Marine Drugs 9, pp.196-223.

Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN, and Carrell RW (1997). The anticoagulant activation of antithrombin by heparin. Proceedings of the National Academy of Sciences of the USA 94, pp.14683-14688.

Kailema MJ, Ruhaak LR, Lebrilla CB, and Amster IJ (2014). Oligosaccharide Analysis by Mass Spectrometry: A review of Recent Developments. Analytical Chemistry 86, pp.196-212.

Kamhi E, Joo EJ, Dordick JS, and Linhardt RJ (2013). Glycosaminoglycans in infectious disease. Biological Reviews 88, pp.928-943.

Kanungo S, Wells K, Tribett T, and El-Gharbawy A (2018). Glycogen metabolism and glycogen storage disorders. Annals of Translational Medicione 6, 474.

Kapaev RR and Toukach PV (2018). GRASS: semi-automated NMR-based structure elucidation of saccharides. Bioinformatics 34, pp.957-963.

Karpozilos A, and Pavlidis N (2004). The treatment of cancer in Greek antiquity. European Journal of Cancer 40, pp.2033-2040.

Kazarian SG and Chan KLA (2013). ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems. Analyst 138, pp.1940-1951.

Keeler J (2010). Understanding NMR Spectroscopy (2nd Edition). John Wiley and Sons: Chichester.

Kimple M, Brill AL, and Pasker RL (2015). Overview of Affinity Tags for Protein Purification. Current Protocols in Protein Science 73, doi:0.1002/0471140864.ps0909s73.

Kljun A, Benians TAS, Goubet F, Meulewaeter F, Knox JP, and Blackburn RS (2011). Comparative Analysis of Crystallinity Changes in Cellulose I Polymers Using ATR-FTIR, X-ray Diffraction, and Carbohydrate-Binding Module Probes. Biomacromolecules 12, pp.4121-4126.

Kogelberg H, Solís D, and Jiménez-Barbero J (2003). New structural insights into carbohydrate-protein interactions from NMR spectroscopy. Current Opinions in Structural Biology 13, pp.646-653.

Komatsu T and Kikuchi J (2013). Comprehensive Signal Assignment of 13C-Labeled Lignocellulose Using Multidimensional Solution NMR and 13C Chemical Shift Comparison with Solid-State NMR. Analytical Chemistry 85, pp.8857-8865.

Kowsaka K, Okajima K, and Kamide K (1991). Determination of the Distribution of Substituent Groups in Sodium Cellulose Sulfate: Assignment of 1H and 13C NMR Peaks by Two-Dimensional COSY and CH-COSY Methods. Polymer Journal 23, pp.823-836.

Kozlov G, Bastos-Aristizabal S, Maattanen P, Rosenauer A, Zheng F, Killikelly A, Trempe JF, Thomas DY, and Gehring K (2010). Structural Basis of Cyclophilin B Binding by the Calnexin/Calreticulin P-domain. Journal of Biological Chemistry 285, pp.35551-35557.

Krasnova L, and Wong C-H (2016). Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis. Annual Review of Biochemistry 85, pp.599-630.

Kunz H (2002). Emil Fischer – Unequalled Classicist, Master of Organic Chemistry Research, and Inspired Trailblazer of Biological Chemistry. Angewandte Chemie 41, pp.4439-4451.

Kupče Ē, Yong JRJ, Widmalm G, and Claridge TDW (2021). Parallel NMR Supersequences: Ten Spectra in a Single Measurement. Journal of the American Chemical Society Au 1, pp.1892-1897.

Kuszmann J (2006). Introduction to Carbohydrates. In: Levy DE and Fügedi P (Eds.) The Organic Chemistry of Sugars (pp.25-52). Taylor and Francis Group: Boca Raton.

Lairon D, Play B, and Jourdheuil-Rahmani D (2007). Journal of Nutritional Biochemistry 18, pp.217-227.

Laurienzo P (2010). Marine Polysaccharides in Pharmaceutical Applications: An Overview. Marine Drugs 8, 2435-2465.

Leavitt S and Freire E (2001). Direct measurement of protein binding energetics by isothermal titration calorimetry. Current Opinion in Biotechnology 11, pp.560-566.

Lentze MJ (2018). The History of Maltose-active Disaccharides. Journal of Paediatric Gastroenterology and Nutrition 66, pp.S4-S6.

Leopold LF, Leopold N, Diehl H-A, and Socaciu C (2011). Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. Spectroscopy 26, pp.93-104.

Li Y, Sun C, Yates EA, Jiang C, Wilkinson MC, and Fernig DG (2016). Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification. Open Biology 6, 150275.

Li YC, Ho IH, Ku CC, Zhong YQ, Hu YP, Chen ZG, Chen CY, Lin WC, Zulueta MM, Hung SC, Lin MG, Wang CC, and Hsiao CD (2014). Interactions that influence the binding of synthetix heparin sulfate based disaccharides to fibroblast growth factor-2. ACS Chemical Biology 9, pp.1712-1717.

Liang Y (2008). Applications of isothermal titration calorimetry in protein science. Acta Biochimica et Biophysica Sinica 40, pp.565-576.

Lichtenthaler F.W. (1992). Emil Fischer's Proof of the Configuration of Sugars: A Centennial Tribute. Angewandte Chemie 31, pp.1541-1556.

Lima MA, Hughes AJ, Veraldi N, Rudd TR, Hussain R, Brito AS, Chavante SF, Tersariol II, Siligardi G, Nader HB, and Yates EA (2013). Antithrombin stabilisation by sulfated carbohydrates correlates with anticoagulant activity. Medicinal Chemistry Communications 4, 870-873.

Lin C-E, Yu C-J, Chen C-L, Chou L-D, and Chou C (2010). Kinetics of Glucose Mutarotation Assessed by an Equal-Amplitude Paired Polarized Heterodyne Polarimeter. Journal of Physical Chemistry A 114, pp.1665-1669.

Lindhorst TK (2003). Essentials of Carbohydrate Chemistry and Biochemistry (2nd Edition). Wiley-VCH: Weinheim.

Liu D, Zhou P, and Nicolai T (2020). Effect of Kappa carrageenan on acid-induced gelation of whey protein aggregates. Part I: Potentiometric titration, rheology and turbidity. Food Hydrocolloids 102, 105589.

Liu H, Zhang Z, and Linhardt RJ (2009). Lessons learnt from the contamination of heparin. Natural Product Reports 26, pp.313-321.

Liu J, Pervin A, Gallo CM, Desai UR, van Gorp CL, and Linhardt RJ (1994). New Approaches for the Preparation of Hydrophobic Heparin Derivatives. Journal of the Pharmaceutical Sciences 83, pp.1034-1039.

Liu L, Bai Y, Sun N, Xia L, Lowary TL, and Klassen JS (2012). Carbohydrate-Lipid Interactions: Affinities of Methylmannose Polysaccharides for Lipids in Aqueous Solution. Chemistry A European Journal 18, pp.12059-12067.

Liu L, Siuda I, Richards MR, Renaud J, Kitova EN, Mayer PM, Tieleman DP, Lowary TL, and Kalssen JS (2016). Structure and Stability of Carbohydrate-Lipid Interactions Methylmannose Polysaccharide-Fatty Acid Complexes. ChemBioChem 17, pp.1571-1578.

Lunt SY, and Vander-Heiden MG (2011). Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. Annual Review of Cell and Developmental Biology 27, pp.441-464.

Madine J, Clayton JC, Yates EA, and Middleton DA (2009). Exploiting a 13C-labelled heparin analogue for in situ solid-state NMR investigations of peptide-glycan interactions within amyloid fibrils. Organic and Biomolecular Chemistry 7, pp.2414-2420.

Marino C and Bordoni AV (2022). Deoxy sugars. General methods for carbohydrate deoxygenation and glycosidation. Organic and Biomolecular Chemistry 20, pp.934-962.

Marsh SE and Blurton-Jones M (2012). Examining the mechanisms that link b-amyloid and a-synuclein pathologies. Alzheimer’s Research and Therapy 4, 11.

Maxwell EJ and Chen DDY (2008). Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. Analytica Chimica Acta 627, pp.25-33.

McClure SM, Ahl PL, and Blue JT (2018). High throughout Differential Scanning Fluorimetry (DSF) Formulation Screening with Complementary Dyes to Assess Protein Unfolding and Aggregation in Presence of Surfactants. Pharmaceutical Research 35, 81.

Mehtiö T, Toivari M, Wiebe MG, Harlin A, Penttilä M, and Koivula A (2016). Production and applications of carbohydrate-derived sugar acids as generic biobased chemicals. Critical Reviews in Biotechnology 36, pp. 904-916.

Mellado-Mojica E, Seeram NP, and López MG (2016). Comparative analysis of maple syrups and natural sweeteners: Carbohydrate composition and classification (differentiation) by HPAEC-PAD and FTIR spectroscopy-chemomtrics. Journal of Food Composition and Analysis 52, pp.1-8.

Mielke SP and Krishnan VV (2009). Characterization of protein secondary structure from NMR chemical shifts. Progress in Nuclear Magnetic Resonance Spectroscopy 54, pp.141-165.

Migliorini E, Thakar D, Kühnle J, Sadir R, Dyer DP, Li Y, Sun C, Volkman BF, Handel TM, Coche-Guerente L, Fernig DG, Lortat-Jacob H, and Richter RP (2015). Cytokines and growth factors cross-link heparin sulfate. Open Biology 5, 150046.

Milland J, and Sandrin MS (2006). ABO blood group and related antigens, natural antibodies and transplantation. Tissue Antigens 68, pp.459-466.

Moniruzzaman M, Nakashima K, Kamiya N, and Goto M (2010). Recent advances of enzymatic reactions in ionic liquids. Biochemical Engineering Journal 48, pp.295-314.

Namias N (2003). Honey in the Management on Infections. Surgical Infection 4, pp.219-226.

Niesen FH, Berglund H, and Vedadi M (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols 2, pp.2212-2221.

Nishiyama Y, Langan P, and Chanzy H (2002). Crystal Structure and Hydrogen-Bonding System in Cellulose Ib from Synchrotron X-ray and Neutron Fiber Diffraction. Journal of the American Chemical Society 124, pp.9074-9082.

Nishiyama Y, Sugiyama J, Chanzy H, and Langan P (2003). Crystal Structure and Hydrogen Bonding System in Cellulose Ia from Synchrotron X-ray and Neutron Fiber Diffraction. Journal of the American Chemical Society 125, pp.14300-14306.

Niu T-T, Zhang D-S, Chen H-M, and Yan X-J (2015). Modulation of the binding of basic fibroblast growth factor and heparanase activity by purified l-carrageenan oligosaccharides. Carbohydrate Polymers 125, pp.76-84.

Novikova EV, Smychenko VM, and Iozep AA (2007). Sulfation in Dextran with Chlorosulfonic Acid in Organic Sovlents. Macromolecular Chemistry and Polymeric Materials 80, pp.1151-1153.

Nunes QM, Su D, Brownridge PJ, Simpson DM, Sun C, Li Y, Bui TP, Zhang X, Huang W, Rigden DJ, Benyon RJ, Sutton R, and Fernig DG (2019). The heparin-binding proteome in normal pancreas and murine experimental acute pancreatitis. PLoS ONE 14, e0217633.

Oduah EI, Linhardt RJ and Sharfstein ST (2016). Heparin: Past, Present, and Future. Pharmaceuticals 9, 38.

Ogura K, Nagata K, Hatanaka H, Habuchi H, Kimata K, Tate S, Ravera MW, Jaye M, Schlessinger J, and Inagaki F (1999). Solution structure of human acidic fibroblast growth factor and interaction with heparin-derived hexasaccharide. Journal of Biomolecular NMR 13, pp.11-24.

Ohmenhaeuser M, Monakhova YB, Kuballa T, and Lachenmeier DW (2013). Qualitative and Quantitative Control of Honeys Using NMR Spectroscopy and Chemometrics. ISRN Analytical Chemistry, 825318.

Oliva A, Llabrés M, and Fariña JB (2001). Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. Journal of Pharmaceutical and Biomedical Analysis 25, pp.833-841.

Olivier-Bourbigou H, Magna L, and Morvan D (2010). Ionic liquids and catalysis: Recent progress from knowledge to applications. Applied Catalysis A: General 373, pp.1-56.

Olsen SN (2006). Applications of isothermal titration calorimetry to measure enzyme kinetics and activity in complex solutions. Thermochimica Acta 448, pp.12-18.

Onishi A, St Ange K, Dordick JS, and Linhardt RJ (2016). Heparin and anticoagulation. Frontiers in Bioscience, Landmark 21, pp.1372-1392.

Ori A, Wilkinson MC, and Fernig DG (2008). The heparanome and regulation of cell function: structures, functions and challenges. Frontiers in Bioscience 13, pp.4309-4338.

Ornitz DM and Itoh N (2015). The Fibroblast Growth Factor signaling pathway. WIREs Developmental Biology 4, pp.215-266.

Ozawa K, Niki R, and Arima S (1984). Interaction of b-Casein and k-Carrageenan. I. Viscosity and Turidity under Non-gelling Conditions. Agricultural and Biological Chemistry 48, pp.627-632.

Palhares LCGF, Brito AS, Lima MA, Nader HB, London JA, Barsukov IL, Andrade GPV, Yates EA, and Chavante SF (2019). A further unique chondroitin sulfate from the shrimp Litopenaeus vannamei with antithrombin activity that modulates acute inflammation. Carbohydrate Polymers 222, 115031.

Palhares LCGF, London JA, Kozlowski AM, Esposito E, Chavante SF, Ni M, and Yates EA (2021). Chemical Modifications of Glycosaminoglycan Polysaccharides. Molecules 26, 5211.

Park S, Baker JO, Himmel ME, Parilla PA, and Johnson DK (2010). Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. Biotechnology for Fuels 3, 10.

Park SS (2019). Post-Glycosylation Modification of Sialic Acid and Its Role in Virus Pathogenesis. Vaccines 7, 171.

Park YB and Cosgrove DJ (2015). Xyloglucan and its Interactions with Other Component of the Growing Cell Wall. Plant and Cell Physiology 56, pp.180-194.

Pauly M and Keegstra K (2016). Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan*. The Annual Review of Plant Biology 67, pp.235-259.

Pauly M and Keegstra K (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. The Plant Journal 54, pp.559-568.

Pazourek J (2010). Monitoring of mutarotation of monosaccharides by hydrophilic interaction chromatography. Journal of Separation Science 33, pp.974-981.

Pellegrini L, Burke DF, von Delft F, Mulloy B, and Blundell TL (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 407, pp.1029-1034.

Pereira L, Amado AM, Critchley AT, van de Velde F, and Ribeiro-Claro PJA (2009). Identification of selected seaweed polysaccharides (phycocolloids) by vibrational spectroscopy (FTIR-ATR and FT-Raman). Food Hydrocolloids 23, pp.1903-1909.

Petri DFS (2015). Xanthan Gum: A versatile polymer for biomedical and technological applications. Journal of Applied Polymer Science 132, 42035.

Peysselon F and Ricard-Blum S (2014). Heparin-protein interactions: From affinity and kinetics to biological roles. Application to an interaction network regulating angiogenesis. Matrix Biology 35, pp.73-81.

Pierce MM, Raman CS, and Nall BT (1999). Isothermal Titration Calorimetry of Protein-Protein Interactions. Methods 19, pp.213-221.

Pigman W (1957). Introduction: Structure and Stereochemistry of the Monosaccharides. In W. Pigman (Ed.), The Carbohydrates (1st Edition, pp.1-76). New York: Academic Press.

Politi M, Groves P, Chávez MI, Cañada FJ, and Jiménez-Barbero J (2006). Useful applications of DOSY experiments for the study of mushroom polysaccharides. Carbohydrate Research 341, pp. 84-89.

Pomin V (2014). NMR Chemical Shifts in Structural Biology of Glycosaminoglycans. Analytical Chemistry 86, pp.65-94.

Poupard N, Badarou P, Fasani F, Groult H, Bridiau N, Sannier F, Bordenave-Juchereau S, Kieda C, Piot J-M, Grillon C, Fruitier-Arnaudin I, and Maugard T (2017). Assessment of Heparanase-Mediates Angiogenesis Using Macrovascular Endothelial Cells: Identification of l-Carrageenan Derivative as a Potent Anti-Angiogenic Agent. Marine Drugs 15, 134.

Purslow JA, Khatiwada B, Bayro MJ, and Venditti V (2020). NMR Methods for Structural Characterisation of Protein-Protein Complexes. Frontiers in Molecular Biosciences 7, 9.

Rabenstein DL (2002). Heparin and heparan sulfate: structure and function. Natural Product Reports 19, pp.312-331.

Raghothama S (2010). NMR of Peptides. Journal of the Indian Institute of Science 90, 145-161.

Rahmoune H, Chen H-L, Gallagher JT, and Rudland PS (1998). Interaction of Heparan Sulfate from Memory Cells with Acidic Fibroblast Growth Factor (FGF) and Basic FGF. The Journal of Biological Chemistry 273, pp.7303-7310.

Raman R, Tharakaraman K, Sasisekharan V, and Sasisekharan R (2016). Glycan-protein interactions in viral pathogenesis. Current Opinion in Structural Biology 40, pp.153-162.

Reily C, Stewart TJ, Renfrow MB, and Novak J (2019). Glycosylation in health and disease. Nature Reviews Nephrology 15, pp.346-366.

Ricard-Blum S, Féraud O, Lortat-Jacob H, Rencurosi A, Fukai N, Dkhissi F, Vittet D, Imberty A, Olsen BR, and van der Rest M (2004). Characterization of Endostatin Binding to Heparin and Heparan Sulfate by Surface Plasmon Resonance and Molecular Modeling: Role of Divalent Cations. The Journal of Biological Chemistry 279, pp.2927-2936.

Righini G (1998). An X-Ray Photoelectron Spectroscopic Study of Ancient Paper and Its Deterioration. Naturwissenschaften 85, pp.171-175.

Robyt JF (1998). Essentials of Carbohydrate Chemistry. Springer Verlag: New York.

Rosalam S, and England R (2006). Review of xanthan gum production from unmodified starches by Xanthamonas comprestris sp. Enzyme and Microbial Technology 39, pp.197-207.

Roy R, Murphy PV, and Gabius H-J (2016). Multivalent Carbohydrate-Lectin Interactions: How Synthetic Chemistry Enables Insights into Nanometric Recognition. Molecules 21, 626.

Rudd TR and Yates EA (2012). A highly efficient tree structure for the biosynthesis of heparan sulfate accounts for the commonly observed disaccharides and suggests a mechanism for domain synthesis. Molecular BioSystems 8, pp.1499-1506.

Rudd TR, Guimond SE, Skidmore MA, Duchesne L, Guerrini M, Torri G, Cosentino C, Brown A, Clarke DT, Turnbull JE, Fernig DG, and Yates EA (2007). Influence of substitution pattern and cation binding on conformation and activity in heparin derivatives. Glycobiology 17, pp.983-993.

Rudd TR, Uniewicz KA, Ori A, Guimond SE, Skidmore MA, Gaudesi D, Xu R, Turnbull JE, Guerrini M, Torri G, Siligardi G, Wilkinson MC, Fernig DG and Yates EA (2010). Comparable stabilisation, structural changes and activities can be induced in FGF by a variety of HS and non-GAG analogues: implications for sequence-activity relationships. Organic and Biomolecular Chemistry 8, pp.5390-5397.

Saha D and Bhattacharya S (2010). Hydrocolloids as thickening and gelling agents in food: a critical review. Journal of Food Science and Technology 47, pp.587-597.

Sassaki GL, Guerrini M, Serrato RV, Filho APS, Carlotto J, Simas-Tosin F, Cipriani TR, Iacomini M, Torri G, and Gorin PAJ (2014). Monosaccharide composition of glycans based on Q-HSQC NMR. Carbohydrate Polymers 104, pp.34-41.

Sattelle BM, Hansen SU, Gardiner J, Almond A (2010). Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. Journal of the American Chemical Society 132, pp.13132-13134.

Schauer R (2000). Achievements and challenges if sialic acid research. Glycoconjugate Journal 17, 485-499.

Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova AV, Yeh BK, Yayon A, Linhardt RJ, and Mohammadi M (2000). Crystal Structure of a Ternary FGF-FGFR-Heparin Complex Reveals a Dul Role for Heparin in FGFR Binding and Dimerization. Molecular Cell 6, pp.743-750.

Schmid J and Sieber V (2015). Enzymatic Transformation Involved in the Biosynthesis of Microbial Exo-polysaccharides Based on the Assembly of Repeat Units. ChemBioChem 16, pp.1141-1147.

Scholefield Z, Yates EA, Wayne G, Amour A, McDowell W, and Turnbull JE (2003). Heparan sulfate regulates amyloid precursor protein processing by BACE1, the Alzheimer’s b-secretase. The Journal of Cell Biology 163, pp.97-107.

Schultink A, Liu L, Zhu L, and Pauly M (2014). Structural Diversity and Function of Xyloglucan Sidechain Subunits. Plants 3, pp.526-542.

Seyrek E and Dubin PL (2010). Glycosaminoglycans as polyelectrolytes. Advances in Colloid and Interface Science 158, pp.119-129.

Seyrek E, Dubin PL, and Henriksen J (2007). Nonspecific Electrostatic Binding Characteristics of the Heparin-Antithrombin Interaction. Biopolymers 86, pp.249-259.

Shao H, Oltion K, Wu T, and Gestwicki JE (2020). Differential scanning fluorimetry (DSF) screen to identify inhibitors of Hsp60 protein-protein interactions. Organic and Biomolecular Chemistry 18, pp.4157-4163.

Sherwood J, De Bruyn M, Constantinou A, Moity L, McElroy CR, Farmer TJ, Duncan T, Raverty W, Hunt AJ, and Clark JH (2014). Dihydrolevoglucosenone (Cyrene) as a bio-based alternative for dipolar aprotic solvents. Chemical Communications 50, pp.9650-9652.

Silva AM, da Silva EC, and da Silva CO (2006). A theoretical study of glucose mutarotation in aqueous solution. Carbohydrate Research 341, pp.1029-1040.

Silva FRF, Dore CMPG, Marques CT, Nascimento MS, Benevides NMB, Rocha HAO, Chavante SF, and Leite EL (2010). Anticoagulant activity, paw edema and pleurisy induced carrageenan: Action of major types of commercial carrageenans. Carbohydrate Polymers 79, pp.26-33.

Simi CK and Abraham TE (2010). Pysico chemical properties of aminated tamarind xyloglucan. Colloids and Surfaces B: Biointerfaces 81, pp.513-520.

Simon-Davis DA and Parish CR (2013). Heparin sulfate: a ubiquitous glycosaminoglycan with multiple roles in immunity. Frontiers in Immunology 4, 470.

Simsek S, Whitney K, and Ohm J-B (2013). Analysis of Cereal Starched by High-Performance Size Exclusion Chromatography. Food Analytical Methods 6, pp.181-190.

Sindrewicz P, Yates EA, Turnbull JE, Lian L-Y, and Yu L-G (2020). Interaction with the heparin-derived binding inhibitors destabilizes galectin-3 protein structure. Biochemical and Biophysical Research Communications 523, pp.336-341.

Sinnaeve D (2014). Simultaneous solvent and J-modulation suppression in PGSTE-based diffusion experiments. Journal of Magnetic Resonance 245, pp.24-30.

Sinnott M (2013). Carbohydrate Chemistry and Biochemistry: Structure and Mechanism (2nd Edition). RSC Publishing: Cambridge.

Situ AJ, Schmidt T, Mazumder P, and Ulmer TS (2014). Characterization of Membrane Protein Interactions by Isothermal Titration Calorimetry. Journal of Molecular Biology 426, pp.3670-3680.

Skarbek K and Milewska MJ (2016). Biosynthetic and synthetic access to amino sugars. Carbohydrate Research 434, pp.44-71.

Soares da Costa D, Reis RL, and Pashkuleva I (2017). Sulfation of Glycosaminoglycans and Its Implications on Human Health and Disorders. Annual Review of Biomedical Engineering 19, pp.1-26.

Some D, Amartely H, Tsadok A, and Lebendiker M (2019). Characterization of Proteins by Size Eclusion Chromatography Couped to Multi-Angle Light Scattering. Journal of Visualised Experiments 148, e59615.

Souillac PO, Middaugh CR, and Rytting JH (2002). Inverstigation of protein/carbohydrate interactions in the dried state. 2. Diffuse refplectance FTIR studies. International Journal of Pharmaceutics 235, pp.207-218.

Sousa F., Guebitz G.M., Kokol V. (2009). Antimicrobial and antioxidant properties of chitosan enzymatically functionalized with flavonoids. Process Biochemistry 44, pp.749-756.

Spiwok V (2017). CH/p Interactions in Carbohydrate Recognition. Molecules 22, 1038.

Stick RV and Williams SJ (2009). Carbohydrates The Essential Molecules of Life (2nd Edition). Elsevier: Oxford.

Stone AK and Nickerson MT (2012). Formation and functionality of whey protein isolate - (kappa-, iota-, and lambda-type) carrageenan electrostatic complexes. Food Hydrocolloids 27, pp.271-277.

Støve SI, Flydal MI, Hausvik E, Underhaug J, and Martinez A (2020). Differential scanning fluorimetry in the screening and validation of pharmacological chaperones for soluble and membrane proteins. In: Pey AL (Ed.) Protein Homeostasis Diseases. Academic Press:London.

Sun C, Li Y, Taylor SE, Mao X, Wilkinson MC, and Fernig DG (2015). HaloTag in an effective expression and solubilisation fusion partner for a range of fibroblast growth factors. PeerJ, 3. DOI:10.7717/peerj.1060.

Sun C, Liu M, Sun P, Yang M, Yates EA, Guo Z, and Fernig DG (2019). Sulfated polysaccharides interact with fibroblast growth factors and protect from denaturation. FEBS Open Bio 9, pp.1477-1487.

Swatloski RP, Spear SK, Holbrey JD, and Rogers RD (2002). Dissolution of Cellulose with Ionic Liquids. Journal of the American Chemical Society 124, pp.4974-4975.

Thinakaran G, Slunt HH, and Sisodia SS (1995). Novel Regulation of Chondroitin Sulfate Glycosaminoglycan Modification of Amyloid Precursor Protein and Its Homologue APLP2. Journal of Biological Chemistry 270, pp.16522-16525.

Thomsen M (2020). Determination of the Molecular Mass of Membrane Proteins Using Size-Exclusion Chromatography with Multiangle Laser Light Scattering (SEC-MALLS). In: Postis VLG, and Goldman A (Eds.) Biophysics of Membrane Proteins: Methods and Protocols. Humana Press: New York.

Tiernan H, Byrne B, and Kazarian SG (2020). ATR-FTIR spectroscopy and spectroscopic imaging for the analysis of biopharmaceuticals. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 24, 118636.

Tobola F, Lelimousin M, Varrot A, Gillon E, Darnhofer B, Blixt O, Birner-Gruenberger R, Imberty A, and Wiltschi B (2018). Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. ACS Chemical Biology 13, pp.2211-2219.

Toida T, Hileman RE, Smith AE, Vlahova PI, and Linhardt RJ (1996). Enzymatic Preparation of Oligosaccharides Containing Antithrombin III Binding Sites. The Journal of Biological Chemistry 271, pp.32040-32047.

Travers A. and Muskhelishvili G. (2015). DNA structure and function. FEBS Journal 282, pp.2279-2295.

Tuomivaara ST, Yaoi K, O'Neill MA, and York WS (2015). Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharide, including an update on xyloglucan nomenclature. Carbohydrate Research 402, pp.56-66.

Uniewicz KA, Ori A, Xu R, Ahmed Y, Wilkinson MC, Fernig DG, and Yates EA (2010). Differential Scanning Fluorimetry Measurement of Protein Stability Changes upon Binding to Glycosaminoglycans: A Screening Test for Binding Specificity. Analytical Chemistry 82, pp.3796-3802.

Usov AI (1998). Structural analysis of red seaweed galactans of agar and carrageenan groups. Food Hydrocolloids 12, pp.301-308.

van de Velde F, Pereira L, and Rollema HS (2004). The revised NMR chemical shift data of carrageenans. Carbohydrate Research 339, pp.2309-2313.

Vanpouille C, Deligny A, Delehedde M, Denys A, Melchior A, Liénard X, Lyon M, Mazurier J, Fernig DG, and Allain F (2007). The Heparin/Hepara Sulfate Sequence That Interacts with Cyclophilin B Contains a 3-O-Sulfated N-Unsubstituted Glucosamine Residue. Journal of Biological Chemistry 282, pp.24416-24429.

Varki A (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446, pp.1023-1029.

Vekariya RL (2017). A review of ionic liquids: Applications towards catalytic organic transformations. Journal of Molecular Liquids 227, pp.44-60.

Velazquez-Compoy A, Leavitt SA, and Freire E (2004). Characterization of Protein-Protein Interactions by Isothermal Titration Calorimetry. In: Hu F (Ed) Methods in Molecular Biology vol. 261: Protein-Protein Interactions: Methods and Protocols. Humana Press: Totowa.

Vessella G, Vázquez JA, Valcárel J, Lagartera L, Monterrey DT, Bastida A, García-Junceda E, Bedini E, Fernández-Mayoralas A, and Revuelta J (2021). Deciphering Structural Determinant in Chondroitin Sulfate Binding to FGF-2: Paving the Way to Enhanced Predictability of Their Biological Functions. Polymers 13, 313.

Viel S, Capitani D, Mannina L, and Segre A (2003). Diffusion-Ordered NMR Spectroscopy: A Versatile Tool for the Molecular Weight Determination of Uncharged Polysaccharides. Biomacromolecules 4, pp.1843-1847.

Vitorge B and Jeanneat D (2006). NMR Diffusion Measurement in Complex Mixtures Using Constant-Time-HSQC-IDOSY and Computer-Optimized Spectral Aliasing for High Resolution in the Carbon Dimension. Analytical Chemistry 78, pp.5601-5606.

Volbeda AG, van der Marel GA, and Codée JDC (2019). Protecting Group Strategies in Carbohydrate Chemistry. In: Vidal S (Ed.) Protecting Groups: Strategies and Applications in Carbohydrate Chemistry. Wiley-VCH: Weinheim.

Walport LJ, Low JKK, Matthews JM, and Mackay JP (2021). The characterization of protein interactions – what, how and how much? Chemical Society Reviews 50, 12292.

Wang ZM, Li L, Zheng BS, Normakhamatov N, and Guo SY (2007). Preparation and anticoagulation activity of sodium cellulose sulfate. International Journal of Biological Macromolecules 41, pp.376-382.

Wang Z-M, Xiao K-J, Li L, and Wu J-Y (2010). Molecular weight-dependent anticoagulation activity of sulfated cellulose derivatives. Cellulose 17, pp.953-961.

Wang J, Yang W, Yang T, Zhang X, Zuo Y, Tian J, Yao J, Zhang J, and Lei Z (2015). Catalytic synthesis of sulfated polysaccharides I: Characterisation of chemical structure. International Journal of Biological Macromolecules 74, pp.61-67.

Watson JD, and Crick FHC (1953). Molecular Structure of Nucleic Acids. Nature 171, pp.737-738.

Watson JT and Sparkman OD (2007). Introduction. In: Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation (4th Edition, pp.1-52) John Wiley and Sons: Chichester.

Weiler JM, Edens RE, Linhardt RJ, and Kapelanski DP (1992). Heparin and modified heparin inhibit complement activation in vivo. The Journal of Immunology 148, pp.3210-3215.

Weis WI and Drickamer K (1996). Structural Basis of Lectin-Carbohydrate Recognition. Annual Review of Biochemistry 65, pp.441-473.

Wessel HP, Labler L, and Tschopp TB (1989). Synthesis of an N-Acetylated Heparin Pentasaccharide and its Anticoagulant Activity in Comparison with the Heparin Pentasaccharide with High Anti-Factor-Xa Activity. Helvetica Chimica Acta 72, pp.1268-1277

Whitfield C and Paiment A (2003). Biosynthesis and assembly of Group 1 capsular polysaccharides in Escherichia coli and related extracellular polysaccharides in other bacteria. Carbohydrate Research 338, pp.2491-2502.

Wiberg KB, Bailey WF, Lambert KM, and Stempel ZD (2018). The Anomeric Effect: It's Complicated. The Journal of Organic Chemistry 83, pp.5242-5255.

Wiedemann HG and Bayer G (1983). Papyrus: The Paper of Ancient Egypt. Analytical Chemistry 55, pp.1220-1230.

Wintjens R, Bifani AM, and Bifani P (2020). Impact of glycan cloud on the B-cell epitope prediction of SARS-CoV-2 Spike protein. NPJ Vaccines 5:81.

Wittmann V (2013). Structural investigation of multivalent carbohydrate-protein interactions using synthetic biomolecules. Current Opinion in Structural Biology 17, pp.982-989.

Wormald MR, Petrescu AJ, Pao Y-L, Glithero A, Elliott T, and Dwek RA (2002). Conformational Studies of Oligosaccharide and Glycopeptides: Complementarity of NMR, X-ray Crystallography, and Molecular Modelling. Chemical Reviews 102, 371-386.

Xu P, Laval S, Guo Z, and Yu B (2016). Microwave-assisted simultaneous O,N-sulfonation in the synthesis of heparin like oligosaccharides. Organic Chemistry Frontiers 3, pp.103-109.

Xu R, Ori A, Rudd TR, Uniewicz KA, Ahmed YA, Guimond SE, Skidmore MA, Siligardi G, Yates EA, and Fernig DG (2012). Diversification of the Structural Determinants of Fibroblast Growth Factor-Heparin Interactions: Implications for binding specificity. The Journal of Biological Chemistry 287, p.40061-40073.

Yamanoi T, Oda Y, and Katsuraya K (2017). Separation of the α- and β-Anomers of carbohydrates by Diffusion-Ordered NMR Spectroscopy. Magnetochemistry 3, 38.

Yates EA and Rudd TR (2016). Recent innovation in the structural analysis of heparin. International Journal of Cardiology 212S1, pp.S5-S9.

Yates EA, Santini F, De Cristofano B, Payre N, Cosentino C, Guerrini M, Naggi A, Torri G, and Hricovini M (2000). Effect of substitution pattern on 1H,13C NMR chemical shifts and 1JCH coupling constants in heparin derivatives. Carbohydrate Research 329, pp.239-247.

Yates EA, Santini F, Guerrini M, Naggi A, Torri G, and Casu B (1996). 1H and 13C NMR spectral assignments of the major sequences of twelve systematically modified heparin derivatives. Carbohydrate Research 294, pp.15-27.

Young CL, Britton ZT, and Robinson AS (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnology Journal 7, pp.620-634.

Yu H and Chen X (2007). Carbohydrate post-glycosylational modifications. Organic and Biomolecular Chemistry 5, pp.865-872.

Yuguchi Y, Thuy TTT, Urakawa H, and Kajiwara K (2002). Structural characteristics of carrageenan gels: temperature and concentration dependence. Food Hydrocolloids 16, pp.515-522.

Zabotina OA (2012). Xyloglucan and its biosynthesis. Frontiers in Plant Science 3, 134.

Zabotina OA, Zhang N, and Weerts R (2021). Polysaccharide Biosynthesis: Glycosyltransferases and Their Complexes.

Zacharski DM , Brandt S, Esch S, König S, Mormann M, Ulrich-Merzenich G, and Hensel A (2015). Xyloglucan from Tropaeolum majus Seeds Induces Cellular Differentiation of Human Keratinocytes by Inhibition of EGFR Phosphorylation and Decreased Activity of Transcription Factor CREB. Biomacromolecules 16, pp.2157-2167.

Zaia J (2004). Mass Spectrometry of Oligosaccharides. Mass Spectrometry Reviews 23, pp.161-227.

Zeng X, Andrade CAS, Oliveira MDL, and Sun X-L (2012). Carbohydrate-protein interactions and their biosencing applications. Analytical and Bioanalytical Chemistry 402, pp.3161-3176.

Zhang F, Wu Y, Ma Q, Hoppensteadt D, Fareed J, and Linhardt RJ (2004). Studies on the Effect of Calcium Interactions Between Heparin and Heparin Cofactor II Using Surface Plasmon Resonance. Clinical and Applied Thrombosis/Haemostasis 10, pp.249-257.

Zhang R, Liu S, and Edgar KJ (2017). Efficient synthesis of secondary amines by reductive amination of curdlan Staudinger ylides. Carbohydrate Polymers 171, pp.1-8.

Zhao H, Baker GA, Song Z, Olubajo O, Crittle T, and Peters D (2008). Designing enzyme-compatible ionic liquids that can dissolve carbohydrates. Green Chemistry 10, pp.696-705.

Zhao J, Kong Y, Zhang F, and Linhardt RJ (2018). Impact of Temperature on Heparin and Protein Interactions. Biochemistry and Physiology 7. DOI:10.4172/2168.1000241.

Zsila F, Juhász T, Kohut G, and Baka-Somfai T (2018). Heparin and Heparan Sulfate Binding if the Antiparasitic Drug Imidocarb: Circular Dichroism Spectroscopy, Isothermal Titration Calorimetry, and Computational Studies. Journal of Physical Chemistry B 122, pp.1781-1791.

Appendix 1 – Chapter 4 Xyloglucan Test-Case Sample Assignments and Tile Plots

A1.1 – XLLG Repeating Unit 'A':

Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.521	3.328	3.501	3.5085	3.687		3.924	3.767
		^{13}C	105.70	75.69	78.21	72.14	76.98		68.59	68.59
	Xyl	^{1}H	4.931	3.532	3.726	3.600	3.703	3.533		
		^{13}C	100.95	74.17	75.71	72.15	63.91	63.91		
2	Glc	^{1}H	4.527	3.402	3.671	3.658	3.894		3.959	3.910
		^{13}C	105.25	75.36	77.76	82.45	76.26		69.33	69.33
	Xyl	^{1}H	5.166	3.659	3.912	3.654	3.567	3.717		
		^{13}C	101.40	82.92	74.56	72.04	63.85	63.86		
	Gal	^{1}H	4.542	3.606	3.651	3.912	3.653		3.787	3.753
		^{13}C	107.18	73.76	75.28	71.26	76.78		63.71	63.71
3	Glc	^{1}H	4.557	3.403	3.672	3.663	3.907		3.905	3.951
		^{13}C	104.99	75.40	77.78	82.48	76.14		69.49	69.49
	Xyl	^{1}H	5.156	3.664	3.898	3.653	3.569	3.719		
		^{13}C	101.38	82.78	74.60	72.05	63.86	63.85		
	Gal	^{1}H	4.547	3.611	3.651	3.912	3.682		3.787	3.752
		^{13}C	107.12	73.72	75.29	71.26	76.85		63.70	63.71
4	Glc-α	^{1}H	5.215	3.585	3.828	3.669	3.940		3.863	3.854
		^{13}C	94.47	73.88	74.11	81.90	72.69		62.64	62.64
	Glc-β	^{1}H	4.654	3.294	3.632	3.681	3.597		3.940	3.801
		^{13}C	98.39	76.56	77.06	81.72	77.39		62.78	62.78

A1.2 – XLXG Repeating Unit ‘B’:

Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.521	3.328	3.500	3.507	3.686		3.767	3.925
		^{13}C	105.70	75.69	78.21	72.14	76.98		68.59	68.58
	Xyl	^{1}H	4.930	3.532	3.725	3.600	3.702	3.534		
		^{13}C	100.95	74.17	75.71	72.14	63.90	63.90		
2	Glc	^{1}H	4.526	3.404	3.669	3.657	3.893		3.957	3.906
		^{13}C	105.24	75.36	77.75	82.46	76.25		69.35	69.36
	Xyl	^{1}H	5.165	3.659	3.912	3.652	3.566	3.716		
		^{13}C	101.39	82.92	74.56	72.05	63.85	63.85		
	Gal	^{1}H	4.542	3.607	3.651	3.912	3.657		3.753	3.788
		^{13}C	107.17	73.76	75.28	71.26	76.80		63.71	63.71
3	Glc	^{1}H	4.546	3.403	3.666	3.724	3.810		3.996	3.888
		^{13}C	105.02	75.40	76.70	81.98	76.00		68.79	68.78
	Xyl	^{1}H	4.946	3.532	3.725	3.601	3.561	3.721		
		^{13}C	101.56	74.18	75.71	72.14	64.21	64.21		
4	Glc-α	^{1}H	5.215	3.585	3.826	3.665	3.940		3.864	3.854
		^{13}C	94.47	73.88	74.11	81.88	72.68		62.64	62.64
	Glc-β	^{1}H	4.654	3.293	3.631	3.679	3.596		3.940	3.799
		^{13}C	98.38	76.56	77.05	81.69	77.39		62.78	62.78

A1.3 – XXLG Repeating Unit ‘C’:

Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.537	3.329	3.503	3.508	3.689		3.767	3.924
		^{13}C	105.55	75.64	78.22	72.19	76.97		68.61	68.60
	Xyl	^{1}H	4.9281	3.5332	3.7217	3.6081	3.7000	3.5342		
		^{13}C	100.93	74.16	75.70	72.15	63.90	63.90		
2	Glc	^{1}H	4.544	3.401	3.659	3.724	3.811		3.890	3.994
		^{13}C	105.28	75.36	76.67	81.97	75.99		68.79	68.79
	Xyl	^{1}H	4.944	3.534	3.722	3.607	3.560	3.718		
		^{13}C	101.47	74.15	75.70	72.14	64.22	64.21		
3	Glc	^{1}H	4.559	3.404	3.672	3.664	3.905		3.907	3.954
		^{13}C	104.99	75.40	77.78	82.48	76.14		69.52	69.52
	Xyl	^{1}H	5.156	3.664	3.899	3.655	3.568	3.720		
		^{13}C	101.39	82.81	74.60	72.05	63.86	63.85		
	Gal	^{1}H	4.547	3.611	3.651	3.912	3.680		3.752	3.787
		^{13}C	107.12	73.72	75.30	71.26	76.86		63.71	63.71
4	Glc-α	^{1}H	5.214	3.585	3.828	3.669	3.939		3.861	3.852
		^{13}C	94.47	73.88	74.11	81.90	72.68		62.64	62.64
	Glc-β	^{1}H	4.654	3.293	3.633	3.682	3.594		3.939	3.800
		^{13}C	98.38	76.56	77.05	81.72	77.38		62.77	62.77

A1.4 – XXXG Repeating Unit ‘D’:

Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.537	3.330	3.503	3.508	3.689		3.767	3.925
		^{13}C	105.54	75.64	78.22	72.19	76.97		68.60	68.60
	Xyl	^{1}H	4.929	3.534	3.722	3.607	3.702	3.534		
		^{13}C	100.93	74.16	75.70	72.15	63.90	63.91		
2	Glc	^{1}H	4.543	3.402	3.660	3.723	3.815		3.888	3.996
		^{13}C	105.27	75.36	76.67	81.96	76.00		68.79	68.79
	Xyl	^{1}H	4.946	3.533	3.723	3.607	3.559	3.719		
		^{13}C	101.52	74.16	75.70	72.15	64.22	64.21		
3	Glc	^{1}H	4.545	3.404	3.665	3.723	3.813		3.993	3.888
		^{13}C	105.01	75.40	76.69	81.98	75.99		68.79	68.78
	Xyl	^{1}H	4.947	3.534	3.722	3.606	3.561	3.721		
		^{13}C	101.56	74.16	75.70	72.15	64.21	64.21		
4	Glc-α	^{1}H	5.214	3.585	3.826	3.666	3.939		3.862	3.852
		^{13}C	94.47	73.88	74.11	81.88	72.68		62.64	62.64
	Glc-β	^{1}H	4.654	3.292	3.631	3.678	3.594		3.938	3.801
		^{13}C	98.37	76.56	77.05	81.70	77.39		62.78	62.78

A1.5 – XLG Fragment ‘E’:

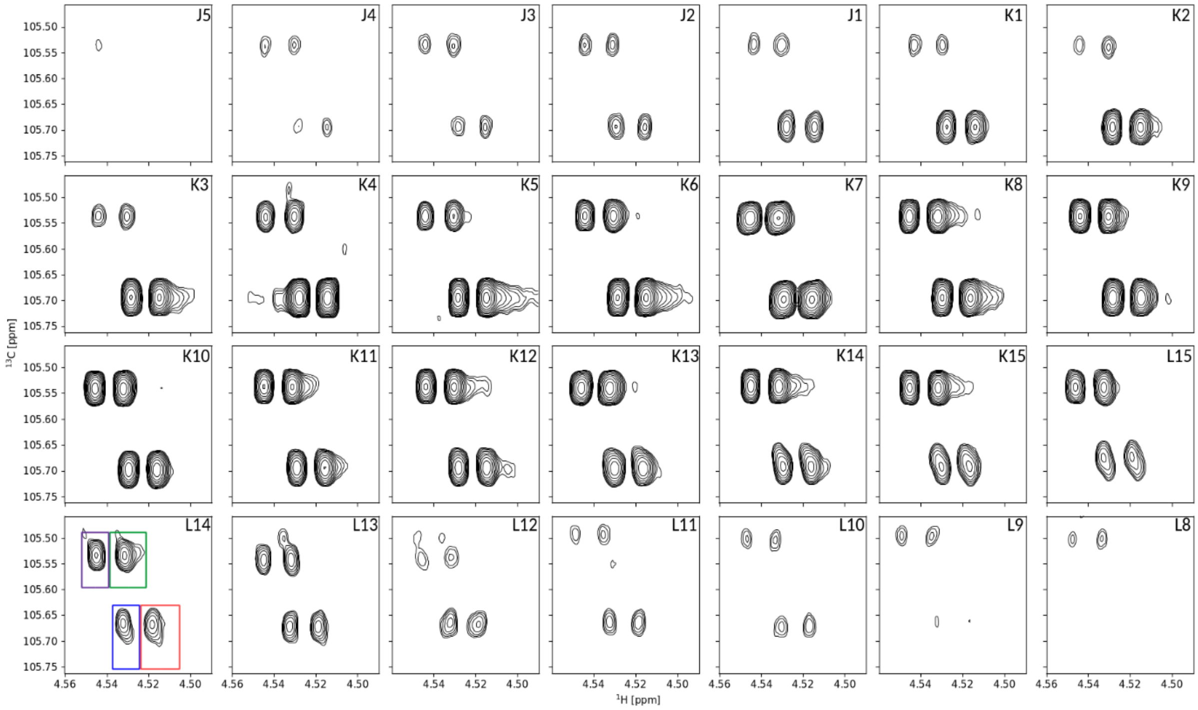
Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.521	3.330	3.500	3.507	3.688		3.924	3.766
		^{13}C	105.69	75.69	78.21	72.14	76.97		68.59	68.59
	Xyl	^{1}H	4.932	3.532	3.723	3.599	3.697	3.535		
		^{13}C	100.95	74.18	75.67	72.15	63.89	63.88		
2	Glc	^{1}H	4.560	3.388	3.671	3.648	3.895		3.959	3.907
		^{13}C	105.13	75.35	77.77	82.31	76.18		69.39	69.39
	Xyl	^{1}H	5.167	3.664	3.905	3.652	3.563	3.715		
		^{13}C	101.37	82.81	74.58	72.04	63.83	63.84		
	Gal	^{1}H	4.546	3.609	3.651	3.912	3.655		3.752	3.787
		^{13}C	107.11	73.74	75.29	71.25	76.80		63.71	63.71
3	Glc-α	^{1}H	5.213	3.572	3.819	3.631	3.940		3.862	3.851
		^{13}C	94.47	73.85	74.05	81.81	72.66		62.59	62.59
	Glc-β	^{1}H	4.653	3.278	3.620	3.641	3.593		3.938	3.795
		^{13}C	98.37	76.51	76.99	81.69	77.34		62.72	62.72

A1.6 – XXG Fragment ‘F’:

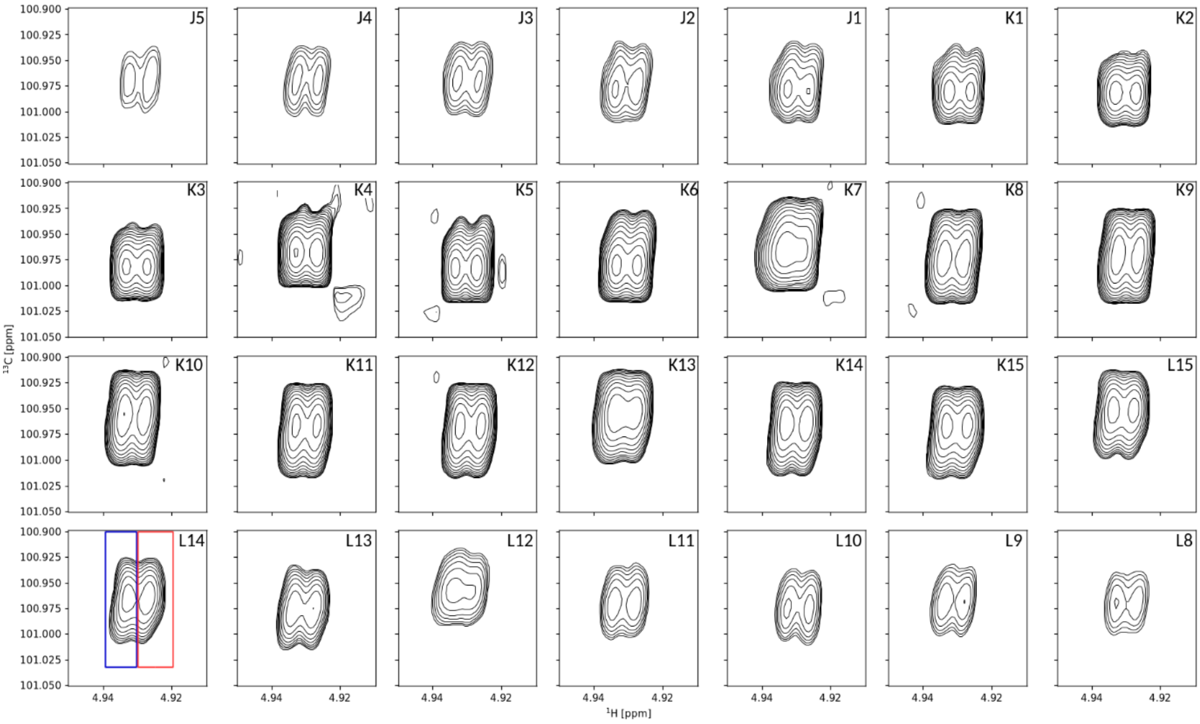
Sidechain	Residue	Atom	Atom Position							
			1	2	3	4	5		6	
1	Glc	^{1}H	4.536	3.327	3.503	3.508	3.688		3.925	3.768
		^{13}C	105.54	75.64	78.22	72.19	76.97		68.60	68.61
	Xyl	^{1}H	4.930	3.534	3.722	3.605	3.702	3.535		
		^{13}C	100.92	74.16	75.67	72.15	63.91	63.90		
2	Glc	^{1}H	4.547	3.388	3.669	3.649	3.896		3.993	3.881
		^{13}C	105.16	75.40	77.76	82.34	76.18		68.71	68.72
	Xyl	^{1}H	4.945	3.534	3.722	3.606	3.720	3.560		
		^{13}C	101.50	74.16	75.68	72.14	64.21	64.22		
3	Glc-α	^{1}H	5.215	3.571	3.817	3.631	3.939		3.861	3.850
		^{13}C	94.47	73.85	74.04	81.81	72.66		62.59	62.59
	Glc-β	^{1}H	4.653	3.279	3.621	3.644	3.594		3.938	3.797
		^{13}C	98.37	76.52	76.99	81.71	77.34		62.73	62.73

Appendix 2 –Chapter 4 Integral Intensity Graphical and Tabulated Data

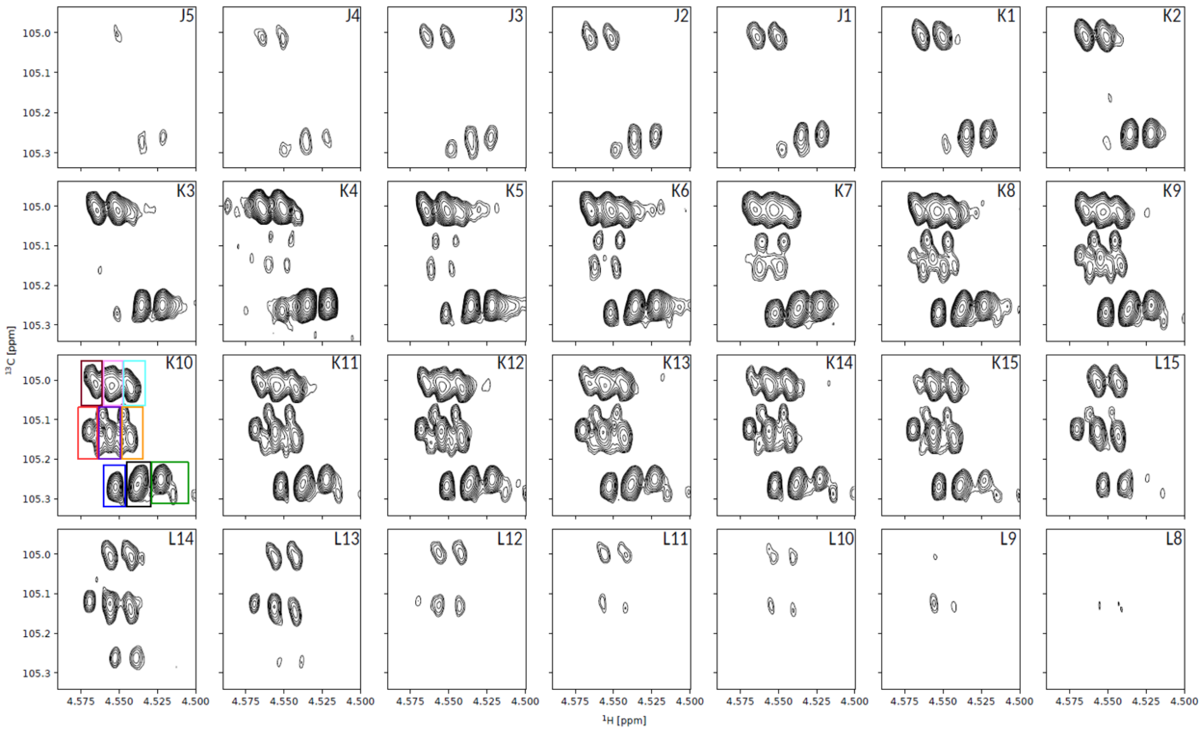
A2.1 – Integral Intensities Measurement Regions Tile Plots:



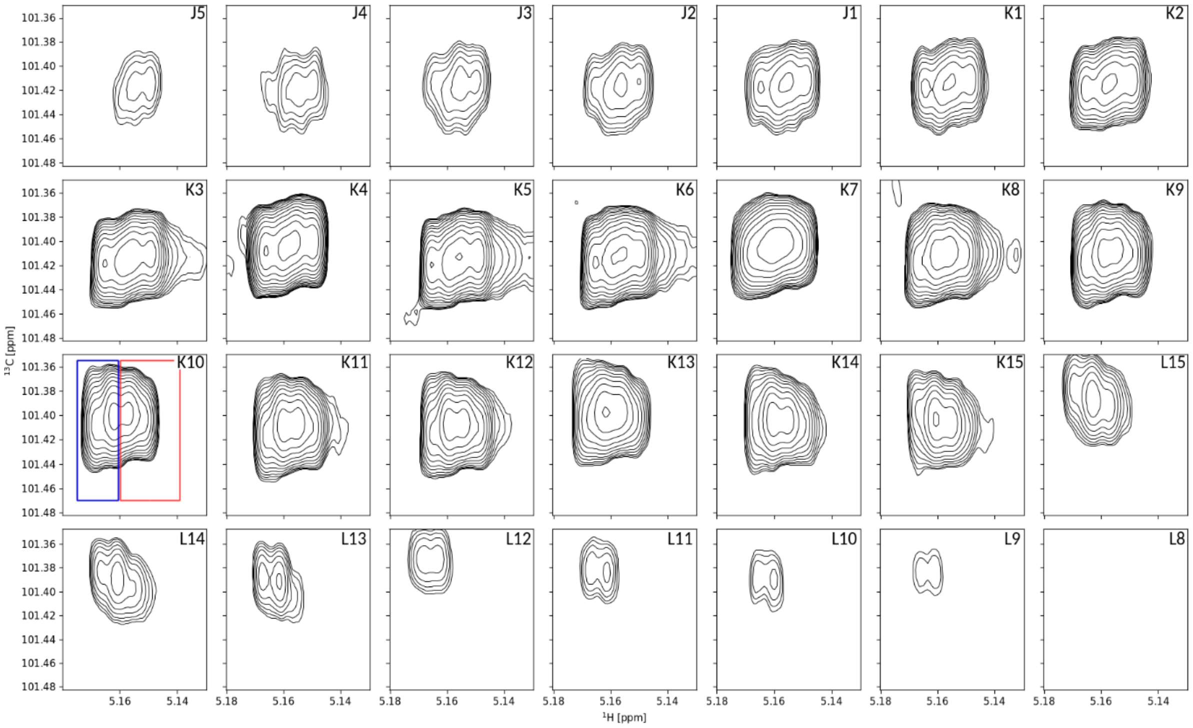
Appendix 2 Figure 1 – Tile plots showing the xyloglucan repeating units’ non-reducing end backbone glucose residue peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in L14 by the blue box for 1G(2L)-L, red box for 1G(2L)-R, purple box for 1G(2X)-L, and green box for 1G(2X)-R.



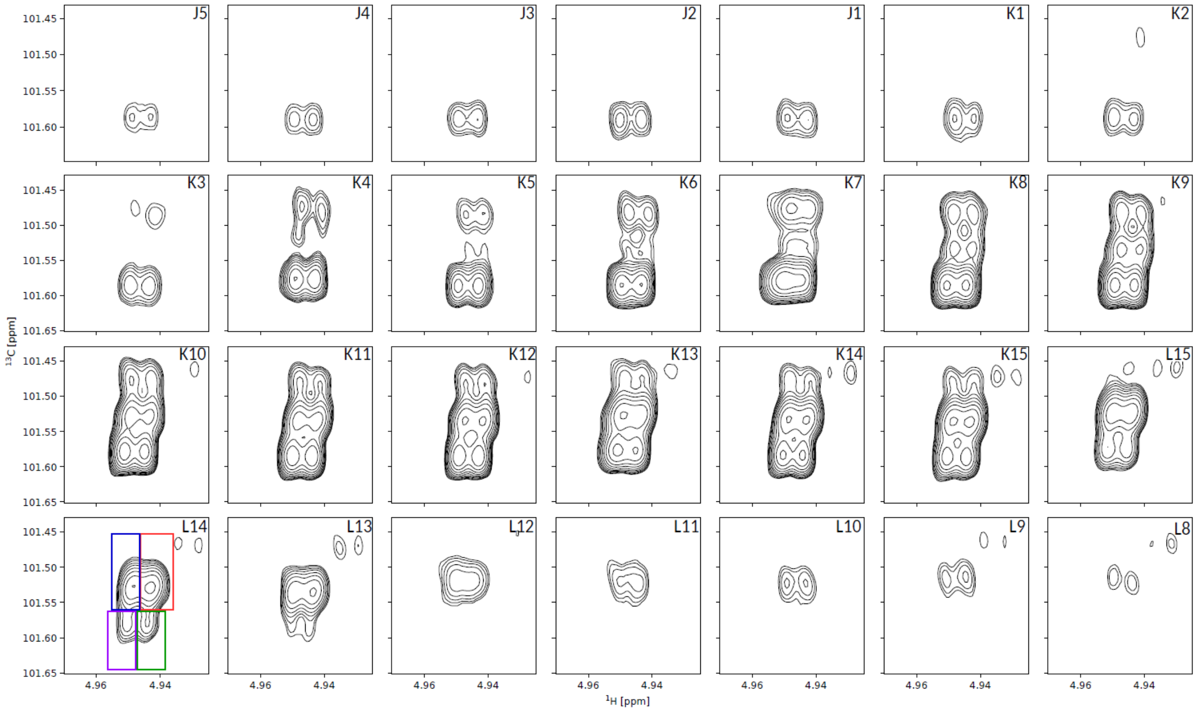
Appendix 2 Figure 2 – Tile plots showing the xyloglucan repeating units' non-reducing end side chain terminal xylose residue peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in L14 by the blue box for 1X-L, and red box for 1X-R.



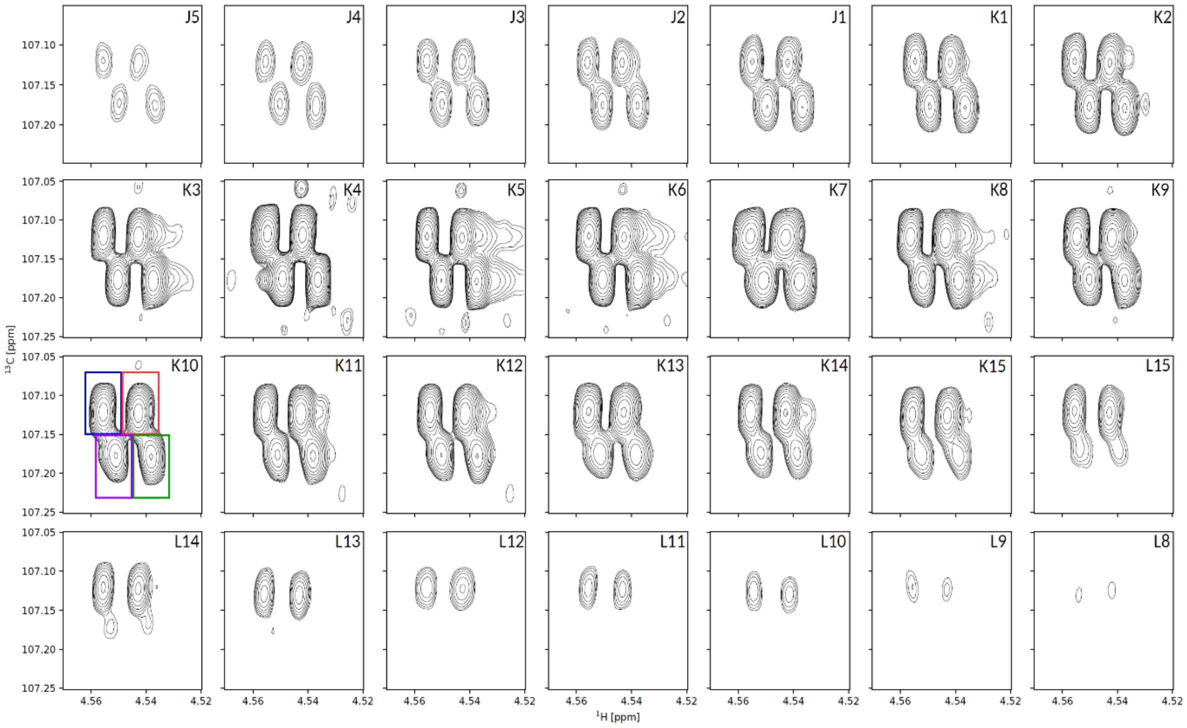
Appendix 2 Figure 3 – Tile plots showing the xyloglucan repeating units' internal backbone glucose residues peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in K10 by the blue box for 2G(X)-L, black box for 2G(X)-R & 2G(L)-L, green box for 2G(L)-R, red box for 2G(XLG)-L, purple box for 2G(XLG)-R, & 2G(XXG)-L, orange box for 2G(XXG)-R, brown box for 3G(L)-L, pink box for 3G(L)-R & 3G(X)-L, and cyan box for 3G(X)-R.



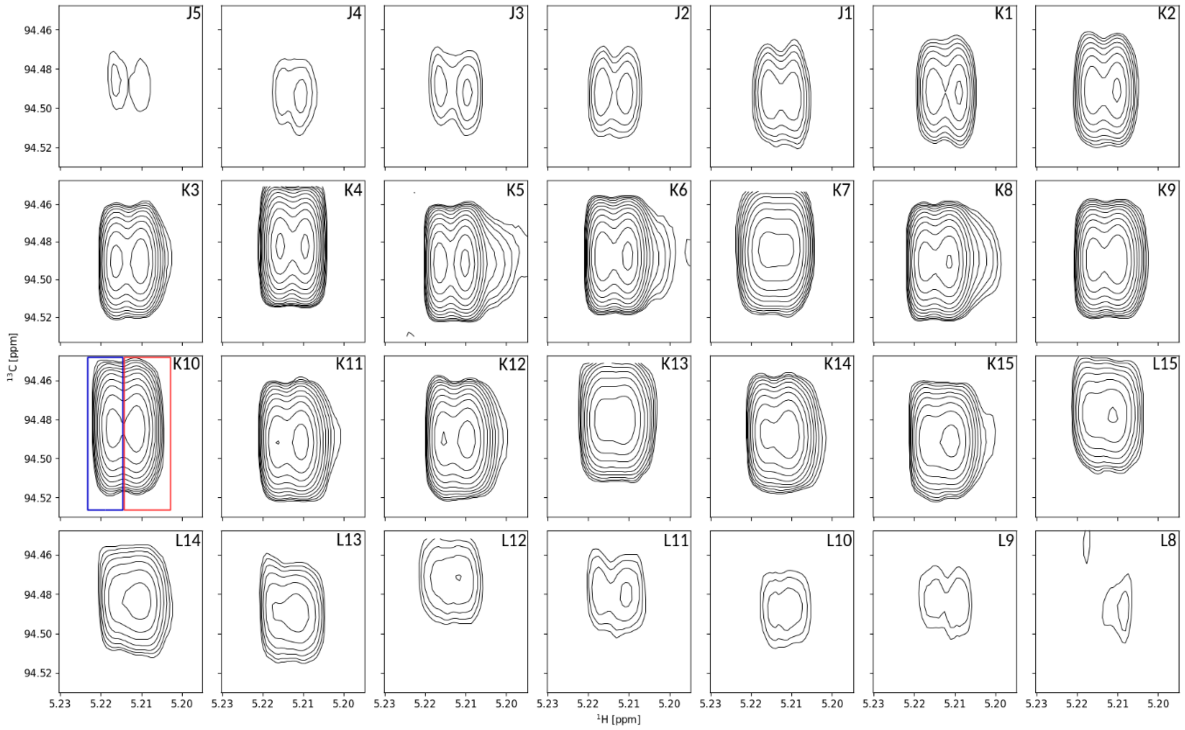
Appendix 2 Figure 4 – Tile plots showing the xyloglucan repeating units' internal side chain further linked xylose residues peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in K10 by the blue box for 2-3Xf-L & 2X(XLG)-L, and red box for 2-3Xf-R & 2X(XLG)-R.



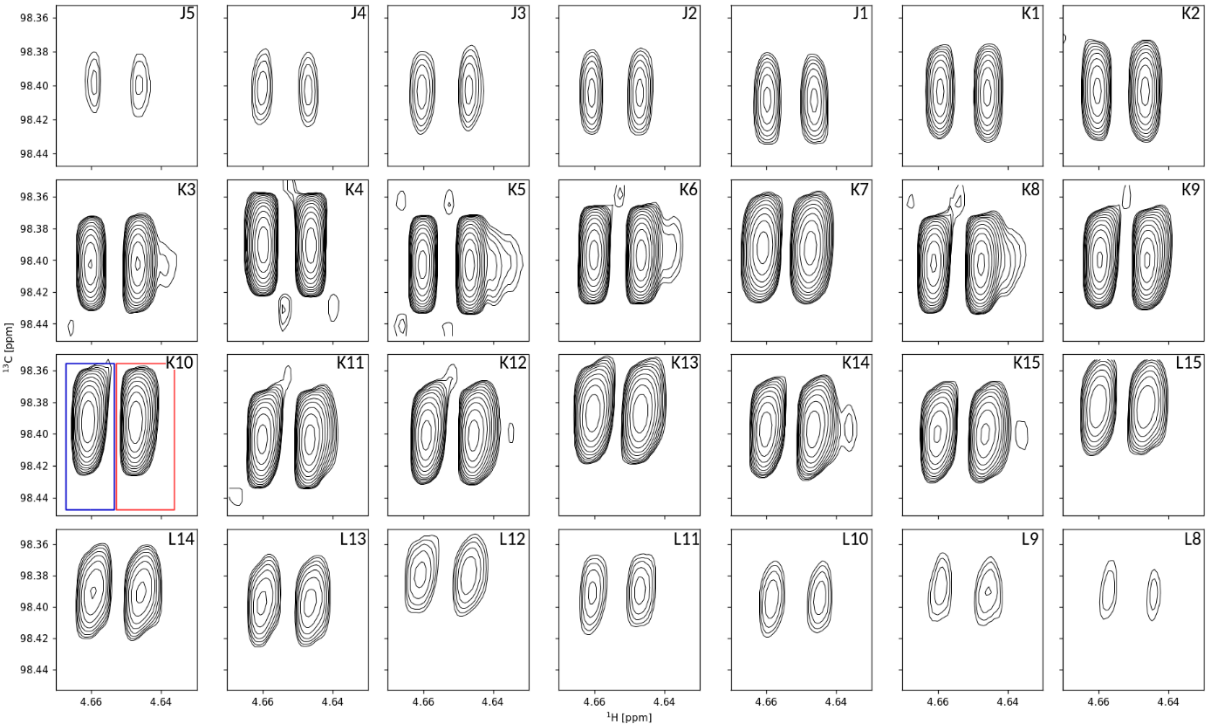
Appendix 2 Figure 5 – Tile plots showing the xyloglucan repeating units' internal side chain terminal xylose residues peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in L14 by the blue box for 2Xt-L, red box for 2Xt-R, purple box for 3Xt-L, and green box for 3Xt-R.



Appendix 2 Figure 6 – Tile plots showing the xyloglucan repeating units' internal side chain terminal galactose residues peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in K10 by the blue box for 3L-L & 2L(XLG)-L, red box for 3L-R & 2L(XLG)-R, purple box for 2L-L and green box for 2L-R.



Appendix 2 Figure 7 – Tile plots showing the xyloglucan repeating units' reducing end backbone glucose residue alpha anomer peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in K10 by the blue box for 4aG-L, and red box for 4aG-R, these also include positions 3bG-L and 3bG-R for the repeating units fragments XLG and XXG.



Appendix 2 Figure 8 – Tile plots showing the xyloglucan repeating units' reducing end backbone glucose residue beta anomer peak development across all sample containing fractions. Integral intensity and maximum amplitude measurement regions are shown in K10 by the blue box for 4bG-L, and red box for 4bG-R, these also include positions 3bG-L and 3bG-R for the repeating units fragments XLG and XXG.

A2.2 – Spectrum Assignments vs. Python Script Detection Areas Tables:

Appendix 2 Table 1 – Xyloglucan repeating unit and fragment anomeric region glucose (G) residue assignments with the regions used to measure integral intensities and maximum amplitudes:

		94.53, 94.45, 5.223, 5.214	94.53, 94.45, 5.214, 5.202	98.45, 98.36, 4.667, 4.653	98.45, 98.36, 4.653, 4.637	105.07, 104.95, 4.572, 4.560	105.07, 104.95, 4.560, 4.548	105.07, 104.95, 4.548, 4.534	105.20, 105.07, 4.575, 4.562	105.20, 105.07, 4.562, 4.549	105.20, 105.07, 4.549, 4.535	105.32, 105.23, 4.559, 4.545	105.32, 105.21, 4.545, 4.528	105.30, 105.21, 4.528, 4.512	105.59, 105.49, 4.552, 4.538	105.59, 105.49, 4.538, 4.521	105.75, 105.64, 4.537, 4.523	105.75, 105.64, 4.523, 4.505
A1G1	L																■	
	R																	■
A2G1	L													■				
	R														■			
A3G1	L					■												
	R						■											
A4αG1	L	■																
	R		■															
A4βG1	L			■														
	R				■													
B1G1	L																■	
	R																	■
B2G1	L												■					
	R													■				
B3G1	L						■											
	R							■										
B4αG1	L	■																
	R		■															
B4βG1	L			■														
	R				■													
C1G1	L														■			
	R															■		
C2G1	L											■						
	R												■					
C3G1	L					■												
	R						■											
C4αG1	L	■																
	R		■															

		94.53, 94.45, 5.223, 5.214	94.53, 94.45, 5.214, 5.202	98.45, 98.36, 4.667, 4.653	98.45, 98.36, 4.653, 4.637	105.07, 104.95, 4.572, 4.560	105.07, 104.95, 4.560, 4.548	105.07, 104.95, 4.548, 4.534	105.20, 105.07, 4.575, 4.562	105.20, 105.07, 4.562, 4.549	105.20, 105.07, 4.549, 4.535	105.32, 105.23, 4.559, 4.545	105.32, 105.21, 4.545, 4.528	105.30, 105.21, 4.528, 4.512	105.59, 105.49, 4.552, 4.538	105.59, 105.49, 4.538, 4.521	105.75, 105.64, 4.537, 4.523	105.75, 105.64, 4.523, 4.505
C4βG1	L			■														
	R				■													
D1G1	L														■			
	R															■		
D2G1	L											■						
	R												■					
D3G1	L						■											
	R							■										
D4αG1	L	■																
	R		■															
D4βG1	L			■														
	R				■													
E1G1	L																■	
	R																	■
E2G1	L								■									
	R									■								
E3αG1	L	■																
	R		■															
E3βG1	L			■														
	R				■													
F1G1	L														■			
	R															■		
F2G1	L									■								
	R										■							
F3αG1	L	■																
	R		■															
F3βG1	L			■														
	R				■													

Appendix 2 Table 2 – Xyloglucan repeating unit and fragment anomeric region xylose (x) residue assignments with the regions used to measure integral intensities and maximum amplitudes:

		101.03, 100.90, 4.940, 4.930	101.03, 100.90, 4.930, 4.920	101.47, 101.36, 5.172, 5.156	101.47, 101.36, 5.156, 5.140	101.56, 101.46, 4.955, 4.946	101.56, 101.46, 4.946, 4.936	101.64, 101.56, 4.956, 4.946	101.64, 101.56, 4.946, 4.937
A1X1	L	■							
	R		■						
A2X1	L			■					
	R				■				
A3X1	L			■					
	R				■				
B1X1	L	■							
	R		■						
B2X1	L			■					
	R				■				
B3X1	L							■	
	R								■
C1X1	L	■							
	R		■						
C2X1	L					■			
	R						■		
C3X1	L			■					
	R				■				
D1X1	L	■							
	R		■						
D2X1	L					■			
	R						■		
D3X1	L							■	
	R								■
E1X1	L	■							
	R		■						
E2X1	L			■					
	R				■				
F1X1	L	■							
	R		■						
F2X1	L					■			
	R						■		

Appendix 2 Table 3 – Xyloglucan repeating unit and fragment anomeric region galactose (L) residue assignments with the regions used to measure integral intensities and maximum amplitudes:

		107.15, 107.07, 4.562, 4.549	107.15, 107.07, 4.549, 4.533	107.23, 107.15, 4.558, 4.544	107.23, 107.15, 4.544, 4.526
A2L1	L			■	
	R				■
A3L1	L	■			
	R		■		
B2L1	L			■	
	R				■
C3L1	L	■			
	R		■		
E2L1	L	■			
	R		■		

A2.3 – Integral Intensities Tabulated Data:

The values contained in the below tables are integral intensities normalised to the greatest value measured i.e. that for 2-3Xf-L & 2X(XLG)-L in sample K7.

Assignment	**J5**	**J4**	**J3**	**J2**	**J1**	**K1**	**K2**
4aG-L	0.0035	0.0047	0.0078	0.0128	0.0174	0.0258	0.0448
4aG-R	0.0051	0.0085	0.0111	0.0121	0.0255	0.0382	0.0593
4bG-L	0.0042	0.0090	0.0153	0.0166	0.0294	0.0433	0.0695
4bG-R	0.0065	0.0085	0.0142	0.0203	0.0291	0.0477	0.0791
3G(L)-L	0.0033	0.0098	0.0163	0.0235	0.0387	0.0565	0.0928
3G(L)-R & 3G(X)-L	0.0059	0.0102	0.0159	0.0224	0.0367	0.0576	0.1004
3G(X)-R	0.0016	0.0011	0.0020	0.0028	0.0064	0.0127	0.0123
2G(XLG)-L	0.0010	0.0015	0.0009	0.0000	0.0001	0.0007	0.0000
2G(XLG)-R & 2G(XXG)-L	0.0012	0.0002	0.0008	0.0011	0.0007	0.0004	0.0006
2G(XXG)-R	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2G(X)-L	0.0019	0.0076	0.0104	0.0080	0.0091	0.0073	0.0050
2G(X)-R & 2G(L)-L	0.0097	0.0174	0.0299	0.0374	0.0489	0.0686	0.1040
2G(L)-R	0.0064	0.0082	0.0185	0.0224	0.0340	0.0616	0.1039
1G(2X)-L	0.0025	0.0044	0.0062	0.0064	0.0079	0.0068	0.0047
1G(2X)-R	0.0026	0.0064	0.0097	0.0092	0.0094	0.0084	0.0086
1G(2L)-L	0.0014	0.0023	0.0065	0.0153	0.0300	0.0522	0.0934
1G(2L)-R	0.0038	0.0062	0.0098	0.0158	0.0322	0.0609	0.1082
2-3Xf-L & 2X(XLG)-L	0.0132	0.0285	0.0471	0.0712	0.1105	0.1536	0.2714
2-3Xf-R & 2X(XLG)-R	0.0277	0.0408	0.0614	0.0707	0.1025	0.1603	0.2235
3Xt-L	0.0064	0.0088	0.0125	0.0180	0.0154	0.0143	0.0142
3Xt-R	0.0057	0.0093	0.0128	0.0123	0.0146	0.0136	0.0100
2Xt-L	0.0000	0.0000	0.0000	0.0000	0.0010	0.0020	0.0023
2Xt-R	0.0018	0.0002	0.0006	0.0000	0.0009	0.0019	0.0038
1X-L	0.0096	0.0214	0.0344	0.0534	0.0753	0.1042	0.1792
1X-R	0.0125	0.0196	0.0294	0.0333	0.0531	0.0873	0.1216
3L-L & 2L(XLG)-L	0.0084	0.0155	0.0255	0.0374	0.0537	0.0788	0.1162
3L-R & 2L(XLG)-R	0.0100	0.0199	0.0320	0.0425	0.0576	0.0907	0.1384
2L-L	0.0083	0.0130	0.0218	0.0312	0.0499	0.0716	0.1134
2L-R	0.0097	0.0171	0.0258	0.0314	0.0453	0.0740	0.1240

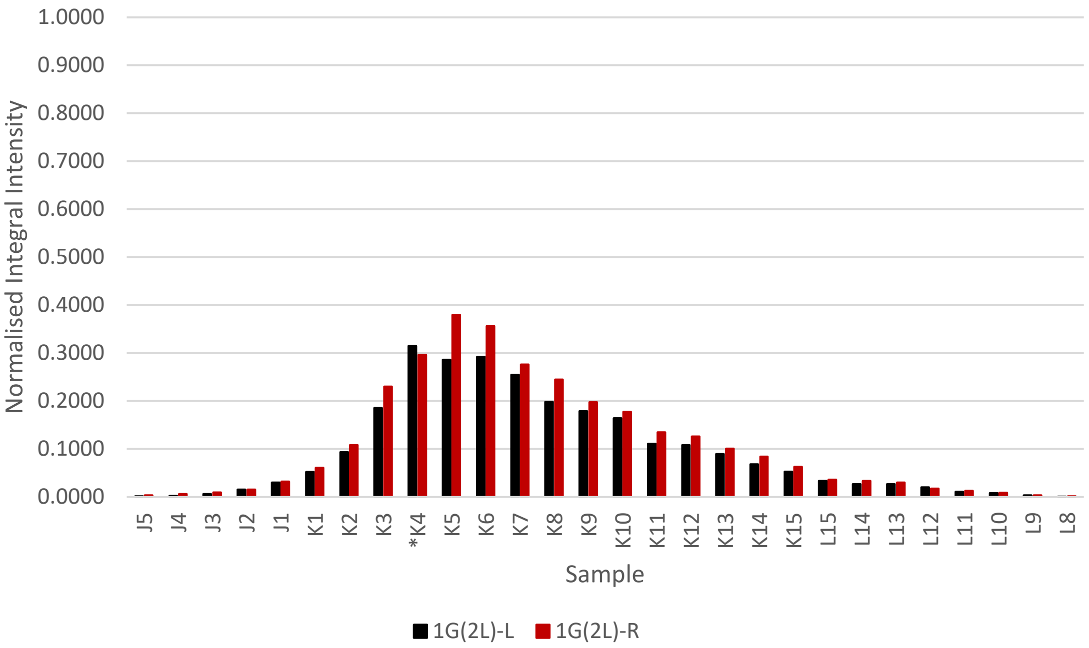
Assignment	**K3**	**K4***	**K5**	**K6**	**K7**	**K8**	**K9**
4aG-L	0.0869	0.1338	0.1331	0.1687	0.1982	0.1820	0.1505
4aG-R	0.1133	0.1611	0.2032	0.2119	0.1622	0.1818	0.1994
4bG-L	0.1241	0.2192	0.2159	0.2533	0.2476	0.2292	0.2416
4bG-R	0.1579	0.2087	0.2743	0.2892	0.2638	0.2804	0.2550
3G(L)-L	0.1772	0.3068	0.2672	0.2704	0.2443	0.1808	0.1547
3G(L)-R & 3G(X)-L	0.2025	0.2880	0.3432	0.3689	0.3573	0.3226	0.3011
3G(X)-R	0.0449	0.0387	0.1005	0.1325	0.1430	0.1685	0.1819
2G(XLG)-L	0.0000	0.0000	0.0000	0.0098	0.0238	0.0300	0.0380
2G(XLG)-R & 2G(XXG)-L	0.0000	0.0000	0.0042	0.0193	0.0473	0.0659	0.0997
2G(XXG)-R	0.0000	0.0000	0.0000	0.0084	0.0096	0.0346	0.0560
2G(X)-L	0.0012	0.0296	0.0132	0.0602	0.1191	0.1175	0.1398
2G(X)-R & 2G(L)-L	0.1987	0.3293	0.3163	0.3645	0.3764	0.3402	0.3145
2G(L)-R	0.2216	0.2791	0.3583	0.3362	0.2408	0.2142	0.1664
1G(2X)-L	0.0088	0.0236	0.0315	0.0731	0.1328	0.1431	0.1750
1G(2X)-R	0.0127	0.0297	0.0454	0.0961	0.1518	0.1849	0.2011
1G(2L)-L	0.1856	0.3151	0.2865	0.2924	0.2551	0.1984	0.1791
1G(2L)-R	0.2304	0.2965	0.3794	0.3564	0.2763	0.2447	0.1973
2-3Xf-L & 2X(XLG)-L	0.4953	0.8586	0.7539	0.8670	1.0000	0.7700	0.6567
2-3Xf-R & 2X(XLG)-R	0.4788	0.5645	0.8108	0.7694	0.4376	0.5184	0.5155
3Xt-L	0.0217	0.0358	0.0550	0.1264	0.2256	0.2410	0.2266
3Xt-R	0.0217	0.0307	0.0659	0.1133	0.1136	0.1696	0.2101
2Xt-L	0.0054	0.0167	0.0128	0.0282	0.0722	0.0764	0.1055
2Xt-R	0.0083	0.0149	0.0190	0.0339	0.0516	0.0722	0.1246
1X-L	0.3345	0.4705	0.5050	0.6149	0.7180	0.6355	0.5367
1X-R	0.2654	0.3940	0.4775	0.4811	0.3394	0.3971	0.4639
3L-L & 2L(XLG)-L	0.2173	0.3816	0.3479	0.3962	0.4086	0.3431	0.3474
3L-R & 2L(XLG)-R	0.2790	0.3553	0.4721	0.5019	0.4460	0.4445	0.3868
2L-L	0.2093	0.3624	0.3087	0.3238	0.2915	0.2184	0.1990
2L-R	0.2555	0.3116	0.4094	0.3805	0.2943	0.2693	0.2079

Assignment	**K10**	**K11**	**K12**	**K13**	**K14**	**K15**	**L15**
4aG-L	0.1794	0.1210	0.1055	0.1285	0.0919	0.0881	0.0532
4aG-R	0.1707	0.1605	0.1800	0.1281	0.1353	0.1047	0.0653
4bG-L	0.2420	0.1847	0.1889	0.1712	0.1379	0.1240	0.0774
4bG-R	0.2553	0.2114	0.2165	0.1814	0.1690	0.1395	0.0858
3G(L)-L	0.1680	0.0925	0.0733	0.0793	0.0470	0.0256	0.0129
3G(L)-R & 3G(X)-L	0.2883	0.2103	0.2115	0.1754	0.1434	0.1136	0.0640
3G(X)-R	0.1732	0.1422	0.1550	0.1213	0.1219	0.1049	0.0650
2G(XLG)-L	0.0691	0.0542	0.0577	0.0708	0.0540	0.0529	0.0416
2G(XLG)-R & 2G(XXG)-L	0.1488	0.1302	0.1496	0.1513	0.1420	0.1405	0.1079
2G(XXG)-R	0.0631	0.0711	0.0919	0.0698	0.0776	0.0803	0.0553
2G(X)-L	0.1532	0.1063	0.1012	0.0961	0.0784	0.0639	0.0316
2G(X)-R & 2G(L)-L	0.3046	0.2239	0.2112	0.1737	0.1380	0.0996	0.0421
2G(L)-R	0.1452	0.1070	0.1018	0.0722	0.0516	0.0268	0.0071
1G(2X)-L	0.2092	0.1569	0.1677	0.1596	0.1368	0.1263	0.0844
1G(2X)-R	0.2343	0.1950	0.2033	0.1852	0.1792	0.1571	0.1033
1G(2L)-L	0.1641	0.1109	0.1082	0.0896	0.0685	0.0531	0.0338
1G(2L)-R	0.1776	0.1347	0.1265	0.1008	0.0842	0.0632	0.0360
2-3Xf-L & 2X(XLG)-L	0.8008	0.4736	0.4189	0.4808	0.3088	0.2513	0.1550
2-3Xf-R & 2X(XLG)-R	0.3085	0.3367	0.3601	0.1540	0.1992	0.1219	0.0275
3Xt-L	0.3081	0.1935	0.1751	0.1730	0.1538	0.1407	0.0682
3Xt-R	0.1301	0.1592	0.1814	0.0797	0.1033	0.0995	0.0200
2Xt-L	0.2163	0.1530	0.1579	0.2716	0.2044	0.1859	0.1953
2Xt-R	0.1500	0.1602	0.2159	0.1576	0.1724	0.1695	0.1163
1X-L	0.6479	0.4373	0.4004	0.4839	0.3420	0.3167	0.2266
1X-R	0.3767	0.3605	0.4235	0.2439	0.2968	0.2389	0.1336
3L-L & 2L(XLG)-L	0.3535	0.2361	0.2355	0.2084	0.1596	0.1298	0.0774
3L-R & 2L(XLG)-R	0.3725	0.2869	0.2770	0.2304	0.2014	0.1553	0.0850
2L-L	0.1889	0.1230	0.1168	0.0860	0.0597	0.0405	0.0143
2L-R	0.1881	0.1403	0.1265	0.0878	0.0710	0.0391	0.0114

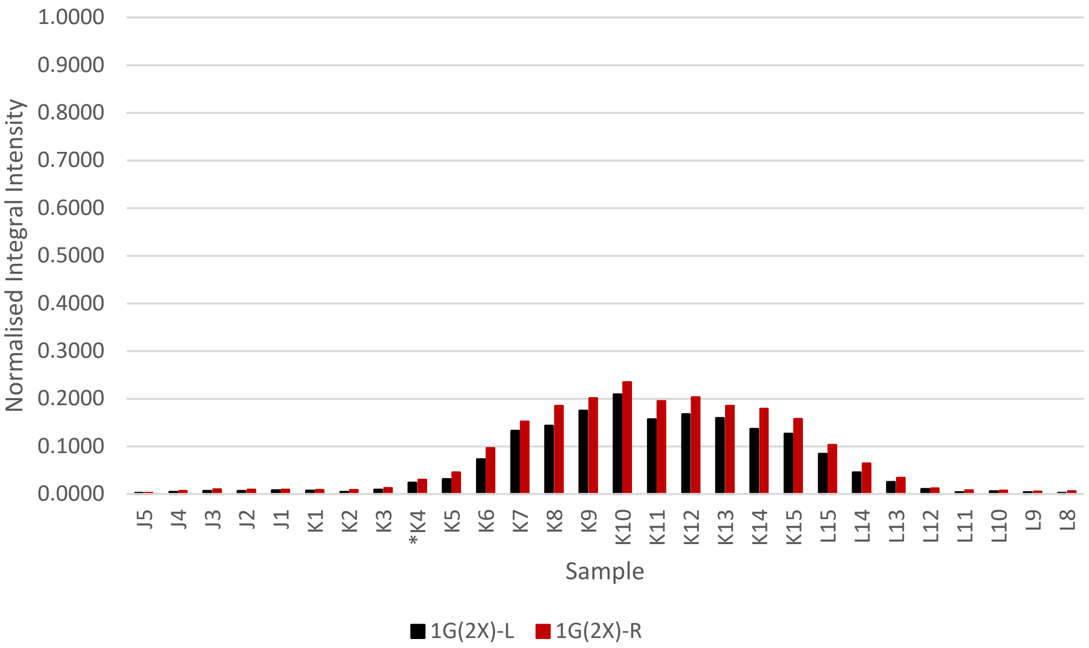
Assignment	**L14**	**L13**	**L12**	**L11**	**L10**	**L9**	**L8**
4aG-L	0.0323	0.0281	0.0147	0.0093	0.0055	0.0056	0.0036
4aG-R	0.0589	0.0305	0.0164	0.0121	0.0106	0.0071	0.0060
4bG-L	0.0551	0.0365	0.0176	0.0144	0.0122	0.0090	0.0045
4bG-R	0.0612	0.0423	0.0219	0.0172	0.0117	0.0104	0.0042
3G(L)-L	0.0050	0.0049	0.0030	0.0023	0.0007	0.0004	0.0002
3G(L)-R & 3G(X)-L	0.0387	0.0252	0.0146	0.0105	0.0078	0.0033	0.0000
3G(X)-R	0.0438	0.0310	0.0178	0.0106	0.0069	0.0022	0.0000
2G(XLG)-L	0.0276	0.0127	0.0045	0.0012	0.0000	0.0000	0.0003
2G(XLG)-R & 2G(XXG)-L	0.0756	0.0440	0.0143	0.0057	0.0040	0.0069	0.0026
2G(XXG)-R	0.0441	0.0253	0.0084	0.0039	0.0030	0.0040	0.0033
2G(X)-L	0.0135	0.0032	0.0009	0.0013	0.0007	0.0006	0.0007
2G(X)-R & 2G(L)-L	0.0203	0.0044	0.0018	0.0010	0.0008	0.0014	0.0014
2G(L)-R	0.0045	0.0010	0.0000	0.0008	0.0005	0.0000	0.0004
1G(2X)-L	0.0454	0.0255	0.0101	0.0041	0.0060	0.0040	0.0026
1G(2X)-R	0.0639	0.0336	0.0119	0.0078	0.0069	0.0053	0.0055
1G(2L)-L	0.0272	0.0268	0.0203	0.0111	0.0085	0.0033	0.0009
1G(2L)-R	0.0333	0.0305	0.0175	0.0132	0.0087	0.0035	0.0014
2-3Xf-L & 2X(XLG)-L	0.0895	0.0692	0.0383	0.0303	0.0189	0.0120	0.0017
2-3Xf-R & 2X(XLG)-R	0.0278	0.0069	0.0010	0.0013	0.0007	0.0007	0.0006
3Xt-L	0.0268	0.0112	0.0000	0.0005	0.0000	0.0000	0.0000
3Xt-R	0.0197	0.0078	0.0000	0.0001	0.0009	0.0004	0.0001
2Xt-L	0.1047	0.0689	0.0407	0.0200	0.0113	0.0125	0.0051
2Xt-R	0.1192	0.0592	0.0165	0.0106	0.0105	0.0080	0.0054
1X-L	0.1292	0.1105	0.0774	0.0476	0.0280	0.0300	0.0170
1X-R	0.1267	0.0724	0.0309	0.0274	0.0305	0.0181	0.0142
3L-L & 2L(XLG)-L	0.0468	0.0316	0.0190	0.0151	0.0103	0.0052	0.0022
3L-R & 2L(XLG)-R	0.0590	0.0368	0.0251	0.0138	0.0100	0.0032	0.0026
2L-L	0.0030	0.0025	0.0000	0.0009	0.0011	0.0000	0.0000
2L-R	0.0021	0.0015	0.0000	0.0000	0.0002	0.0010	0.0001

A2.4 – Integrals Intensities Histograms:

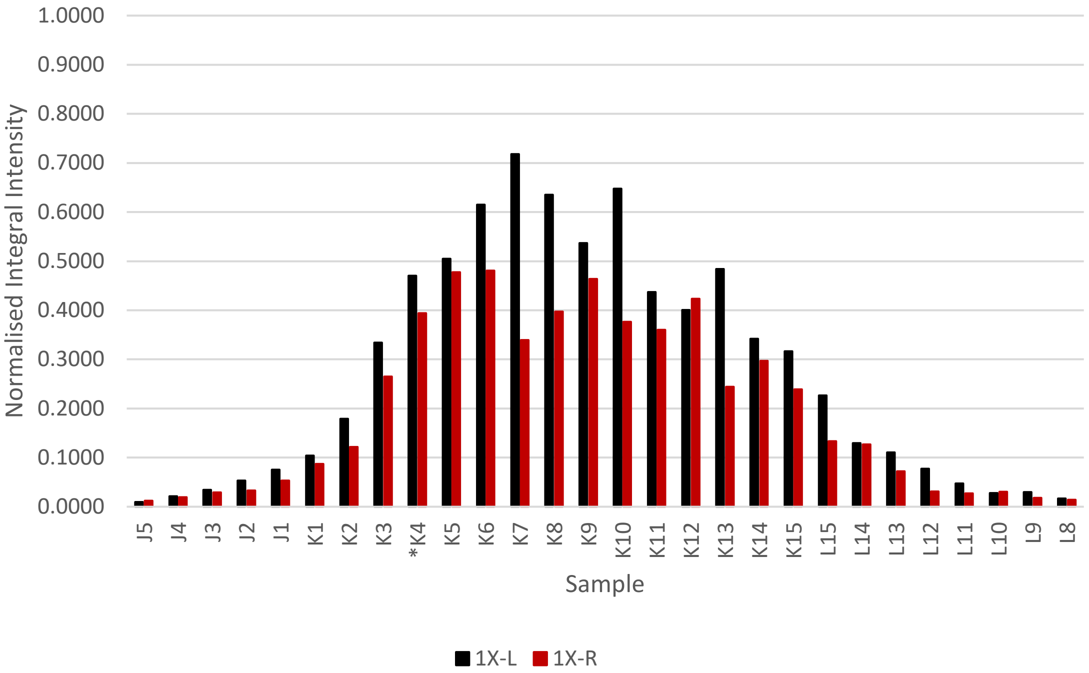
The values presented graphically below are integral intensities normalised to the greatest value measured i.e. that for 2-3Xf-L & 2X(XLG)-L in sample K7.



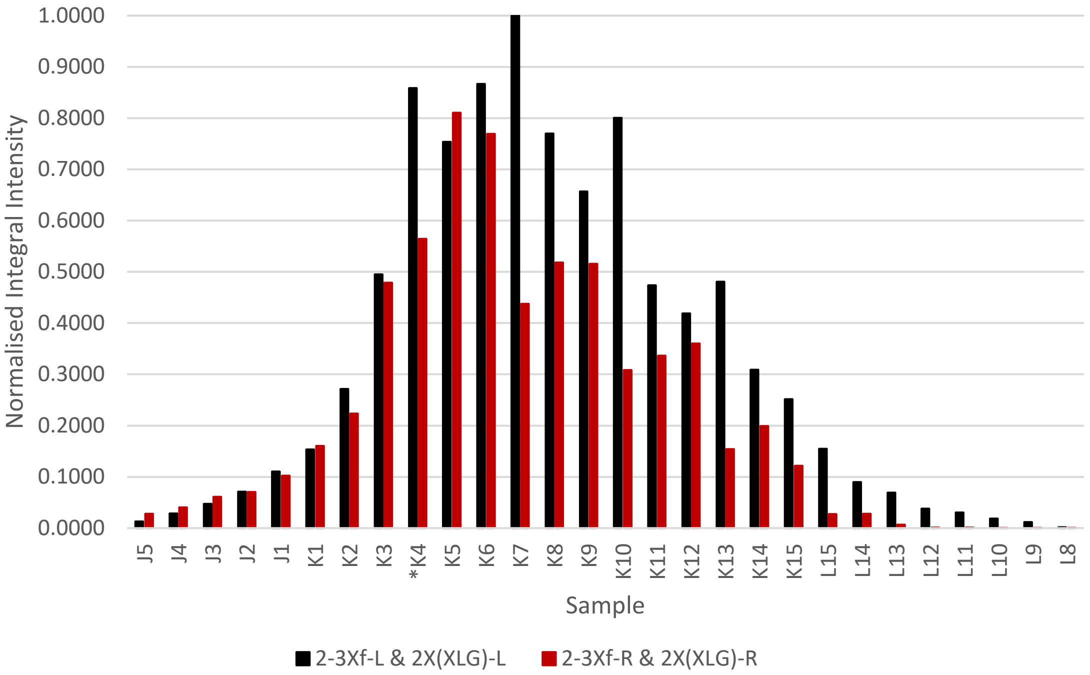
Appendix 2 Figure 9 – Integral intensities for the xyloglucan repeating units' non-reducing end backbone glucose (G) anomeric signal for residues where the repeating units' second side chain contains a terminal galactose (L) residue. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



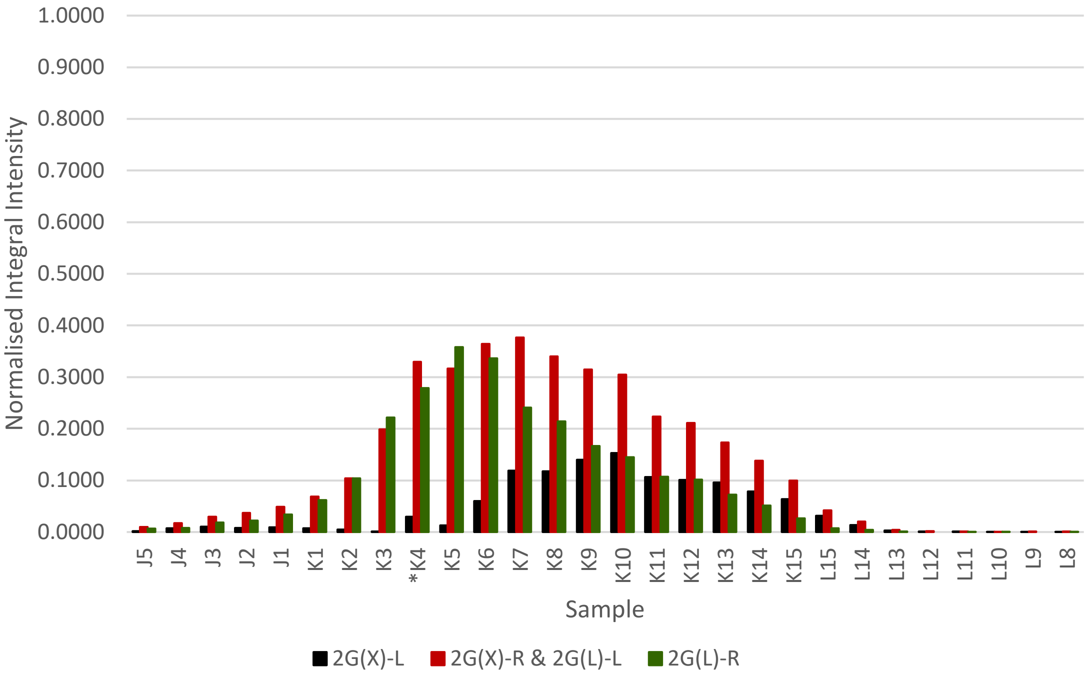
Appendix 2 Figure 10 – Integral intensities for the xyloglucan repeating units' non-reducing end backbone glucose (G) anomeric signal for residues where the repeating units' second side chain contains a terminal xylose (X) residue. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



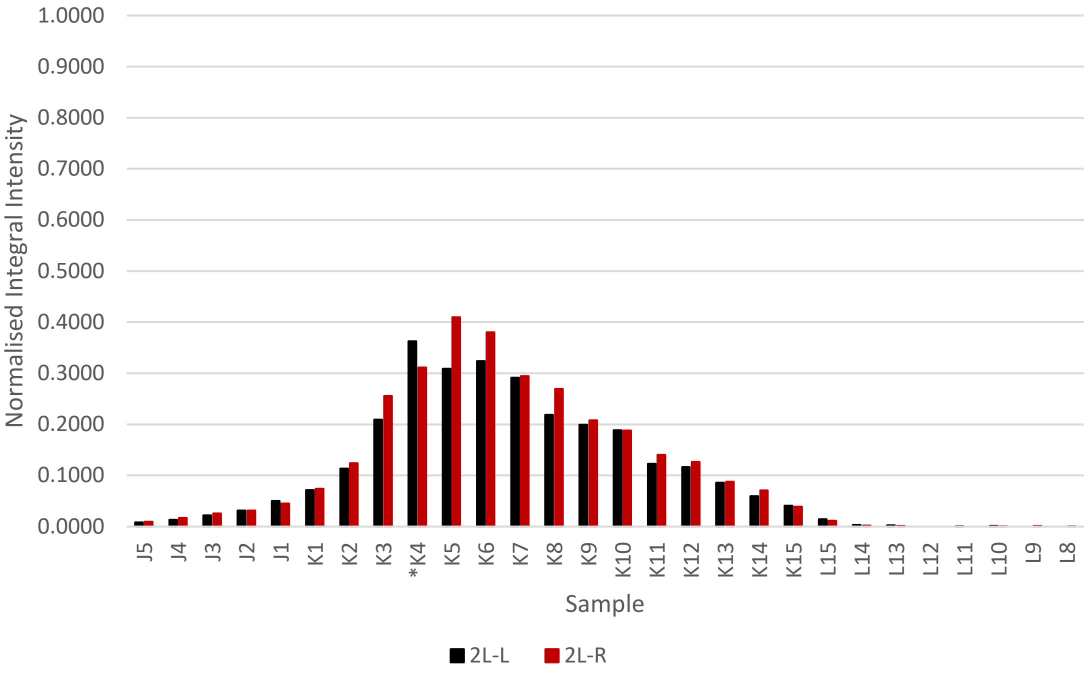
Appendix 2 Figure 11 – Integral intensities for the xyloglucan repeating units' non-reducing end side chain terminal xylose (X) anomeric signal. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



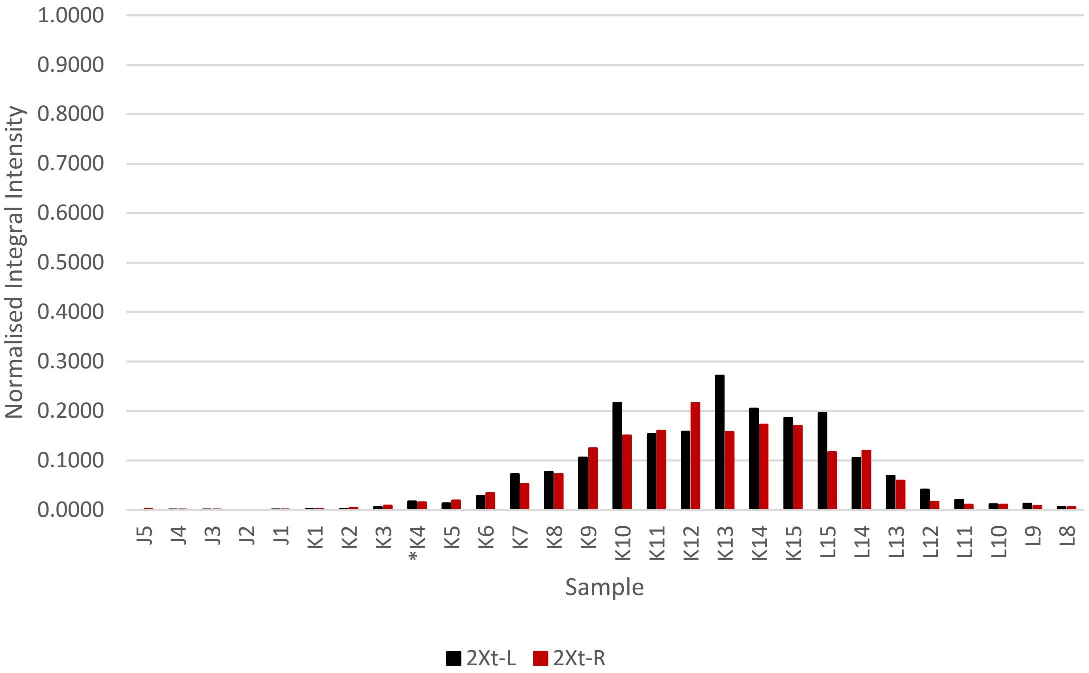
Appendix 2 Figure 12– Integral intensities for the xyloglucan repeating units' internal side chain further linked xylose (X) anomeric signals, alongside the XLG fragment's internal side chain terminal xylose. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



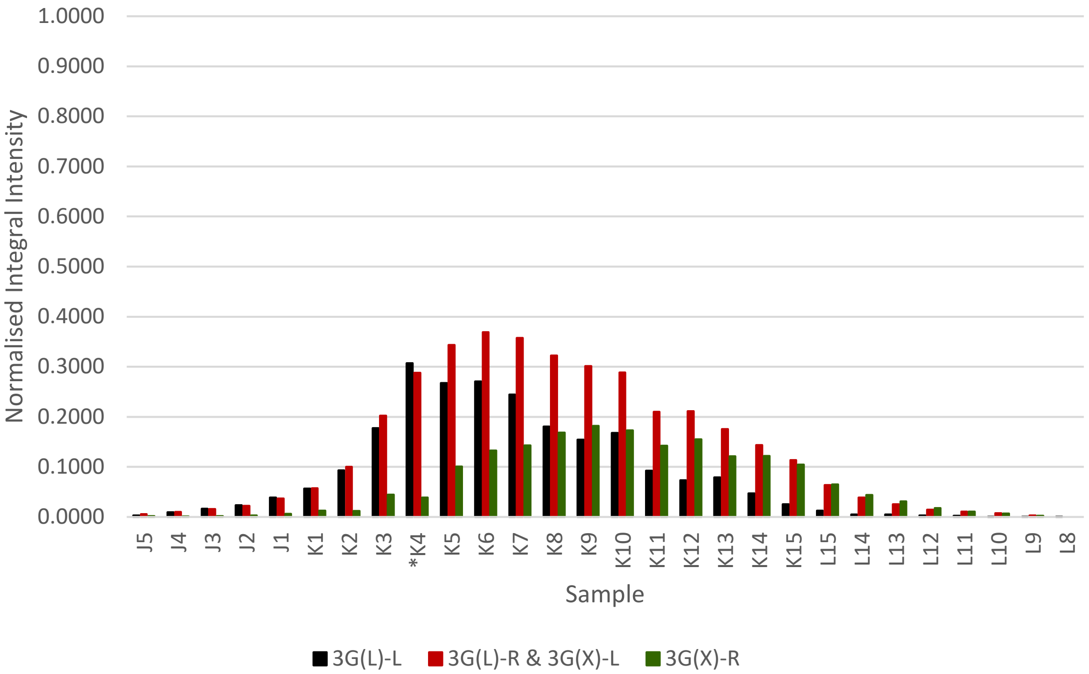
Appendix 2 Figure 13 – Integral intensities for the xyloglucan repeating units' internal (second position) backbone glucose (G) anomeric signal for residues where the linked side chain contains a terminal xylose (X)and those where the linked side chain contains a terminal galactose (L). The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



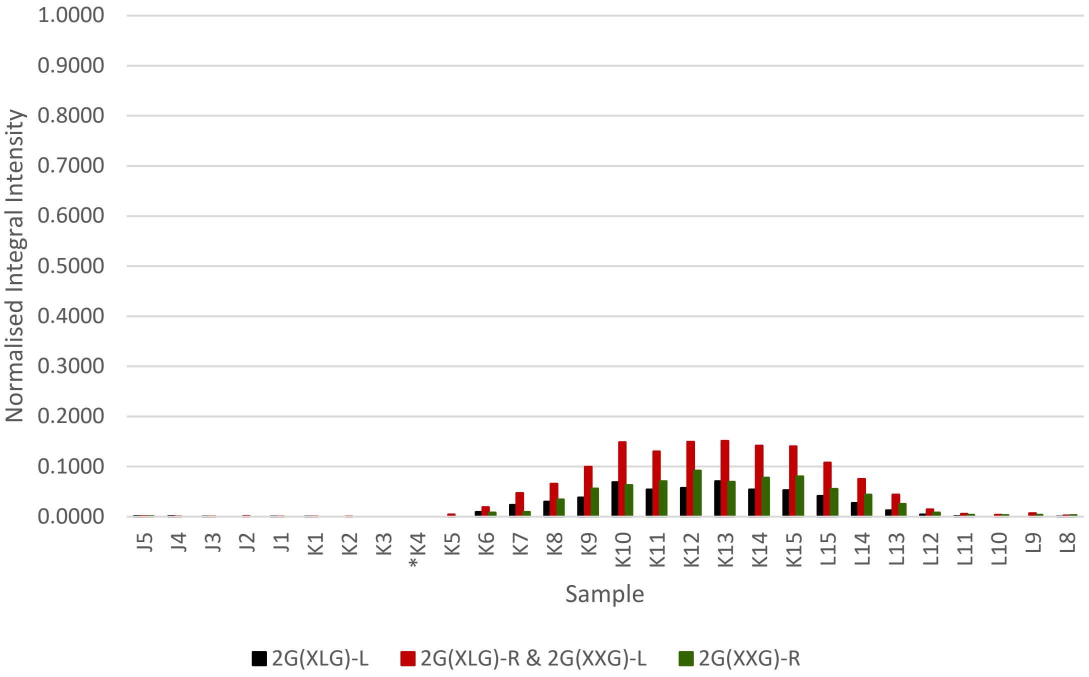
Appendix 2 Figure 14 – Integral intensities for the xyloglucan repeating units' internal (second position) side chain terminal galactose (L) anomeric signal. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



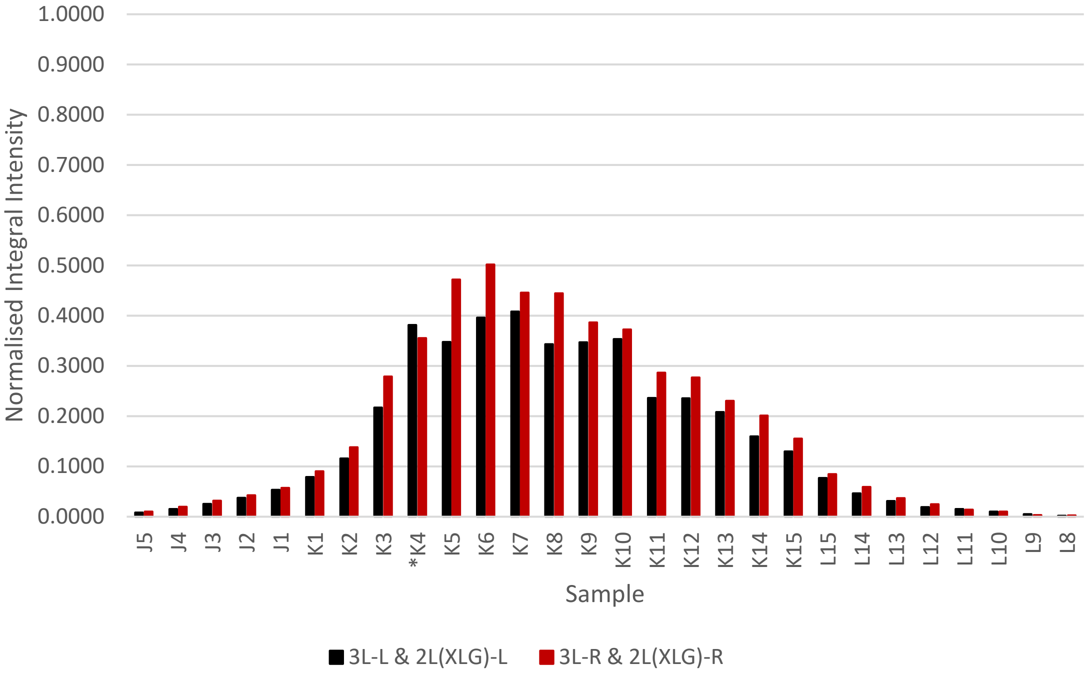
Appendix 2 Figure 15 – Integral intensities for the xyloglucan repeating units' internal (second position) side chain terminal xylose (X) anomeric signal. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



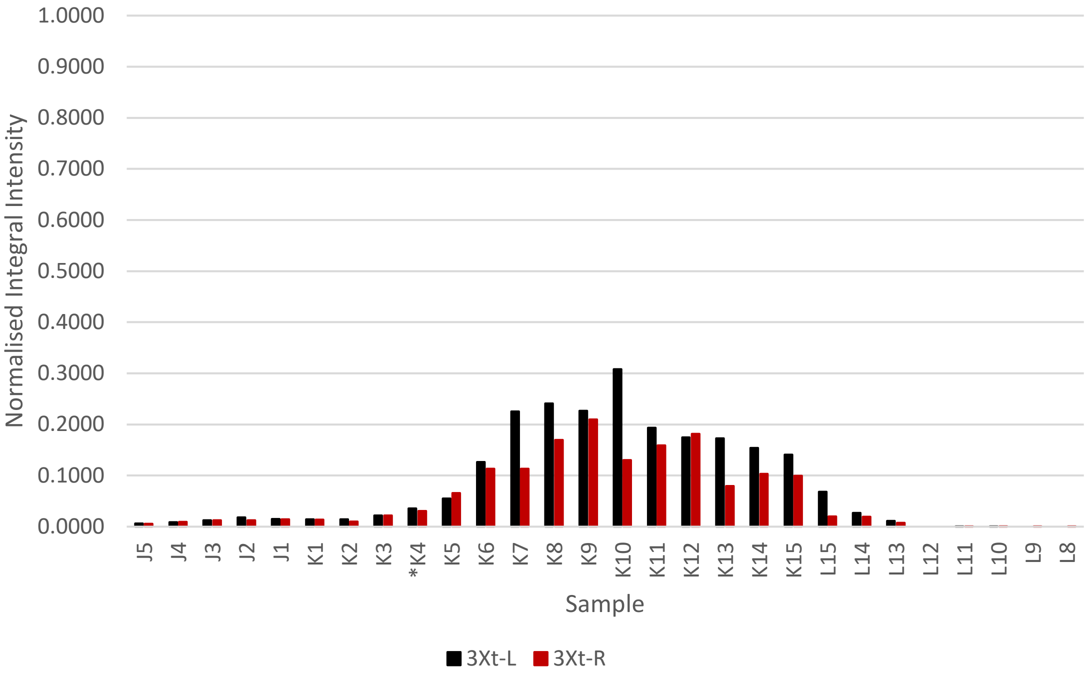
Appendix 2 Figure 16 – Integral intensities for the xyloglucan repeating units' internal (third position) backbone glucose (G) anomeric signal for residues where the linked side chain contains a terminal xylose (X) and those where the linked side chain contains a terminal galactose (L). The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



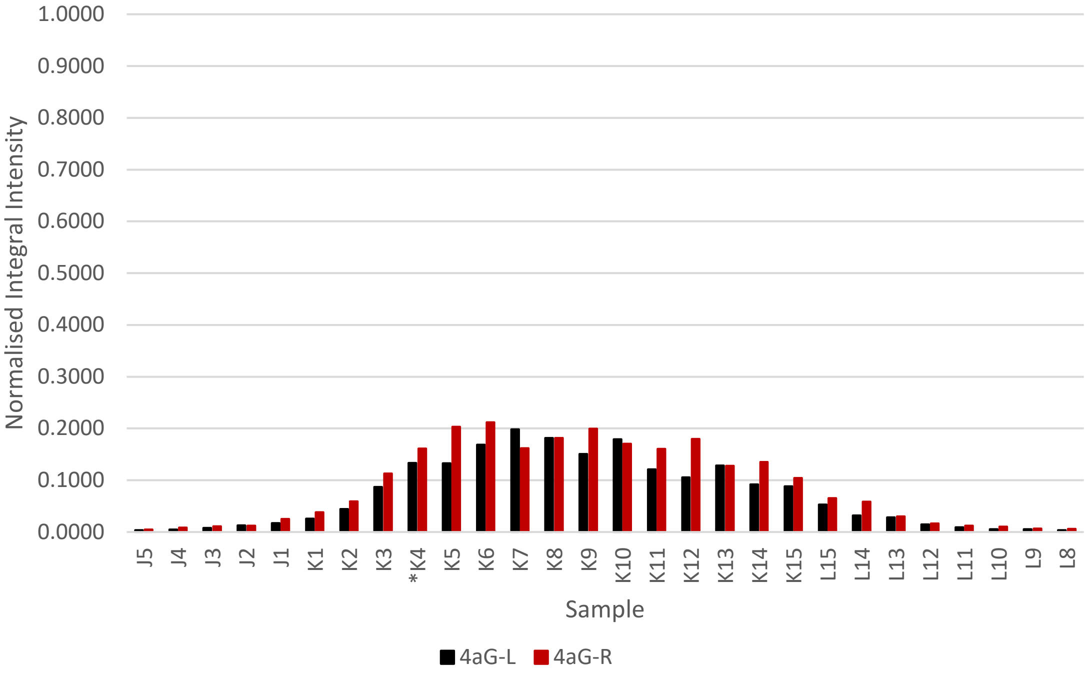
Appendix 2 Figure 17 – Integral intensities for the XLG fragment internal backbone glucose (G) anomeric signal as well as that for the XXG fragment internal backbone glucose (G) residue. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



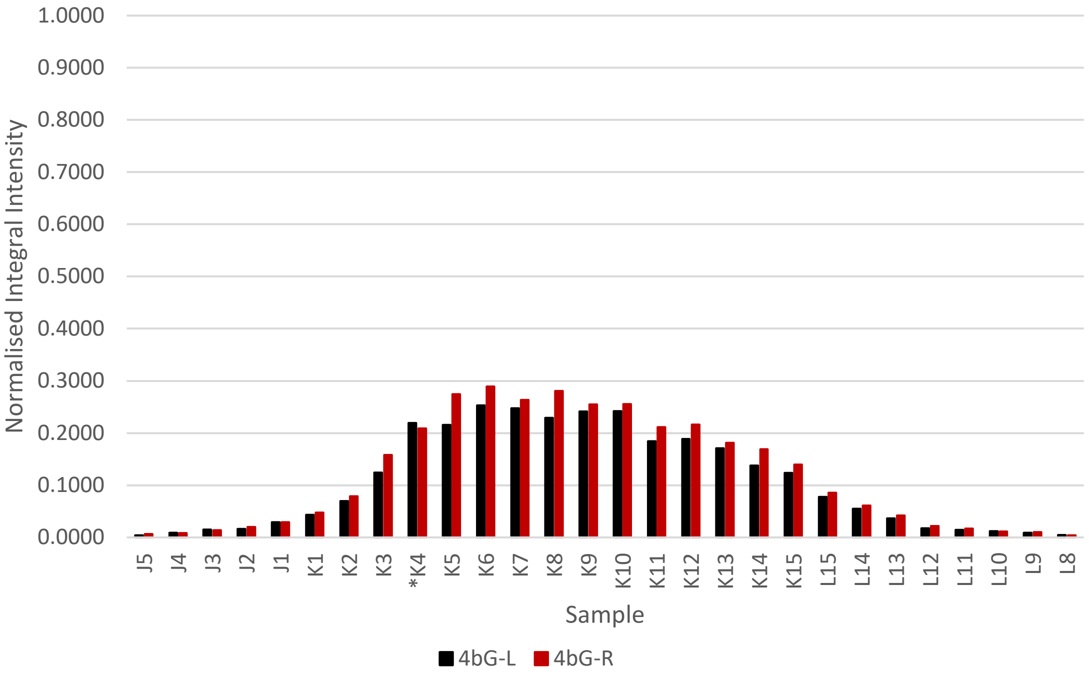
Appendix 2 Figure 18 – Integral intensities for the xyloglucan repeating units' interal (third position) side chain terminal galactose's (L) anomeric signal, as well as that for the XLG fragments internal side chain terminal galactose. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



Appendix 2 Figure 19 – Integral intensities for the xyloglucan repeating units' internal (third position) side chain terminal xylose (X) anomeric signal. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



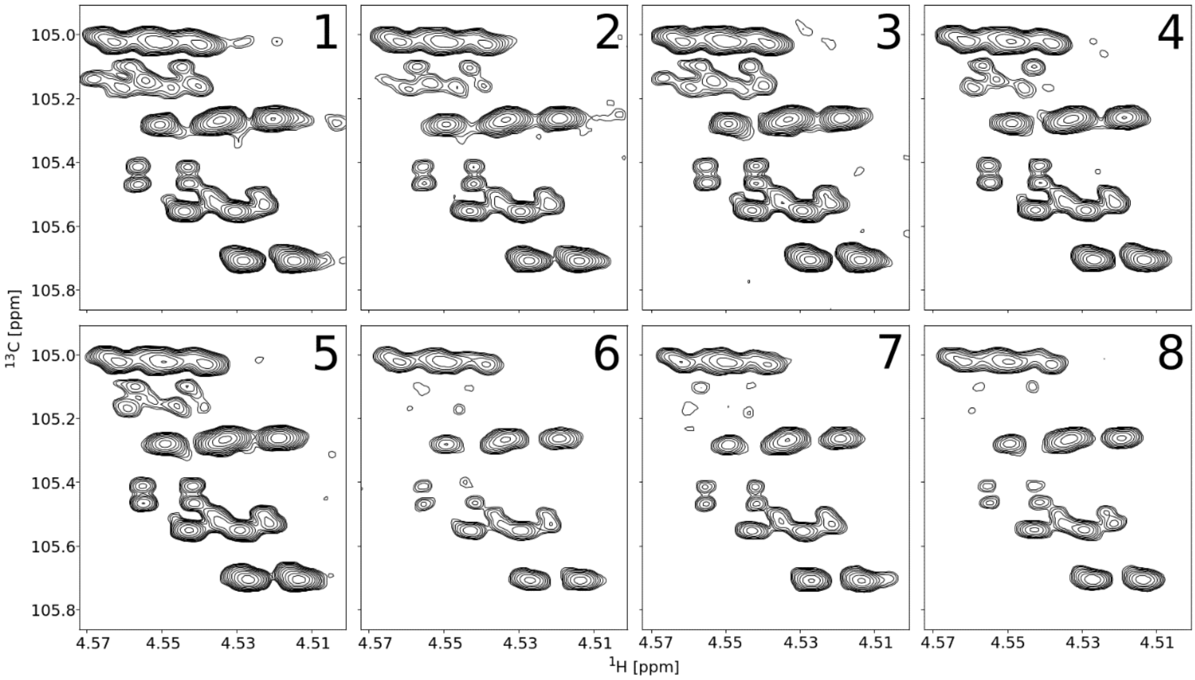
Appendix 2 Figure 20 – Integral intensities for the xyloglucan repeating units' reducing end backbone glucose (G) alpha anomer's anomeric signal as well as that for the XLG and XXG fragments' reducing end alpha anomer. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.



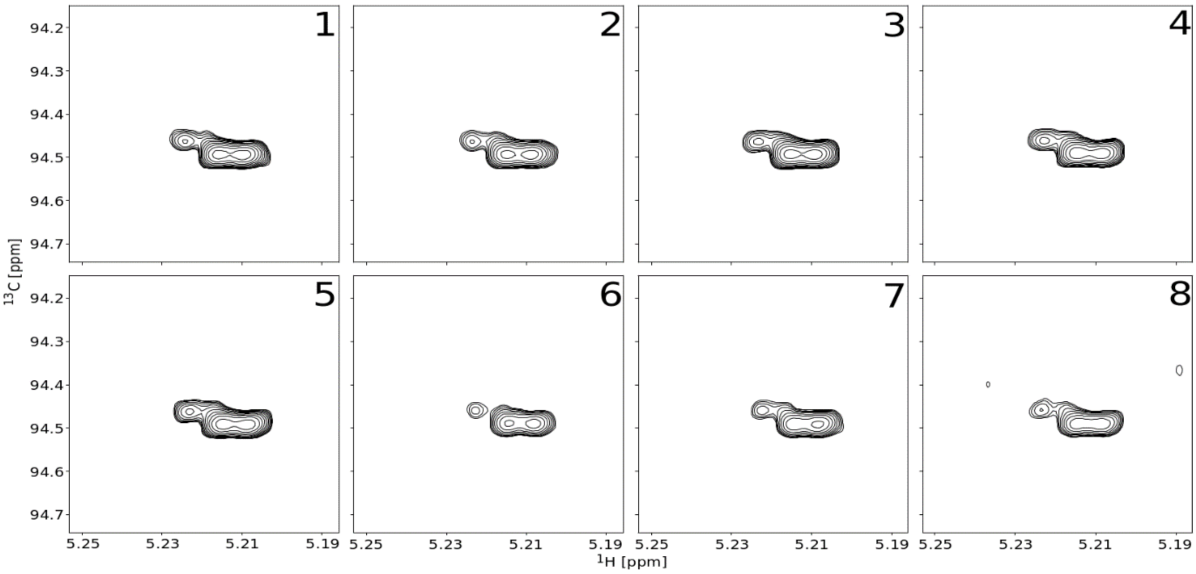
Appendix 2 Figure 21 – Integral intensities for the xyloglucan repeating units' reducing end backbone glucose (G) beta anomer's anomeric signal as well as that for the XLG and XXG fragments' reducing end beta anomer. The fraction K4 intensity has been halved to correct for a scaling error during processing in Bruker TopSpin.

Appendix 3 – Chapter 4 Different Suppliers Samples Spectrum Regions and Tabulated Integral Intensity Data

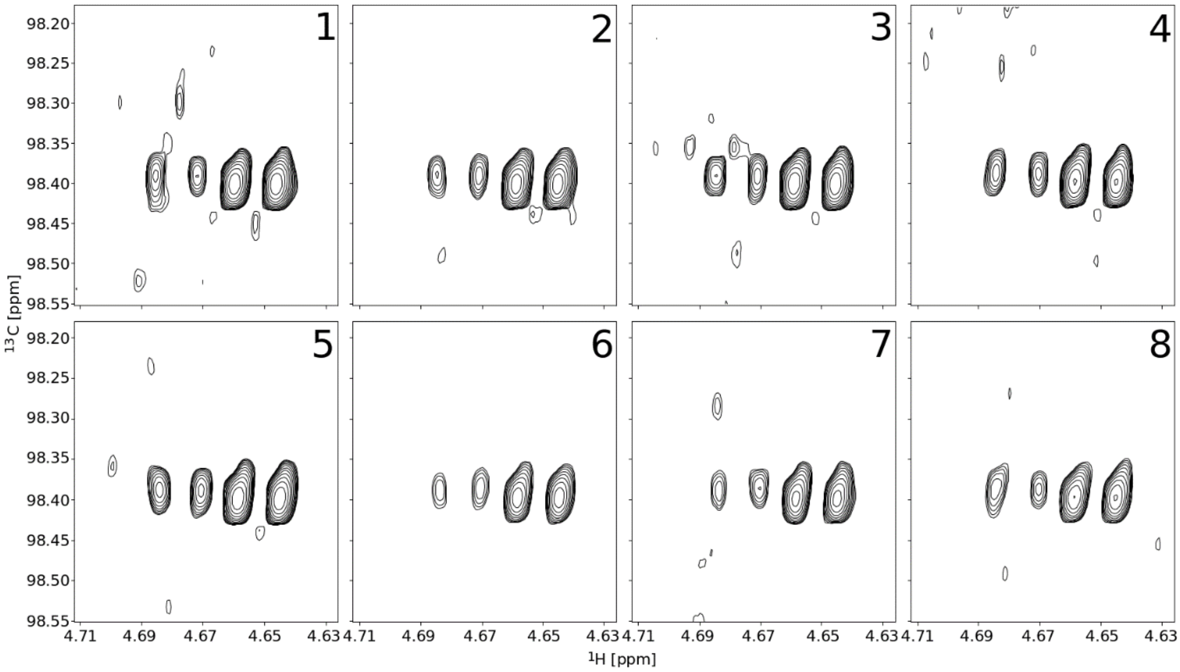
A3.1 – Different Supplier Xyloglucan Major Peak Tile Plots:



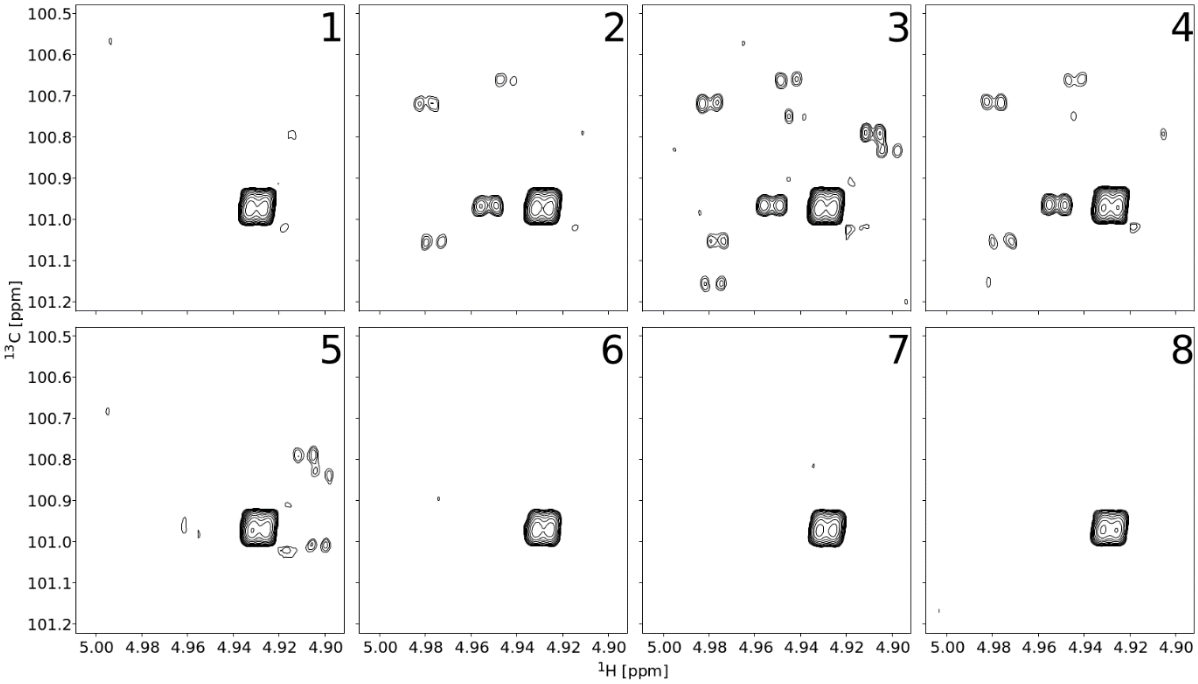
Appendix 3 Figure 1 – Tile plot showing the non-reducing end and internal glucose peaks present in the anomeric region of the eight different suppliers samples.



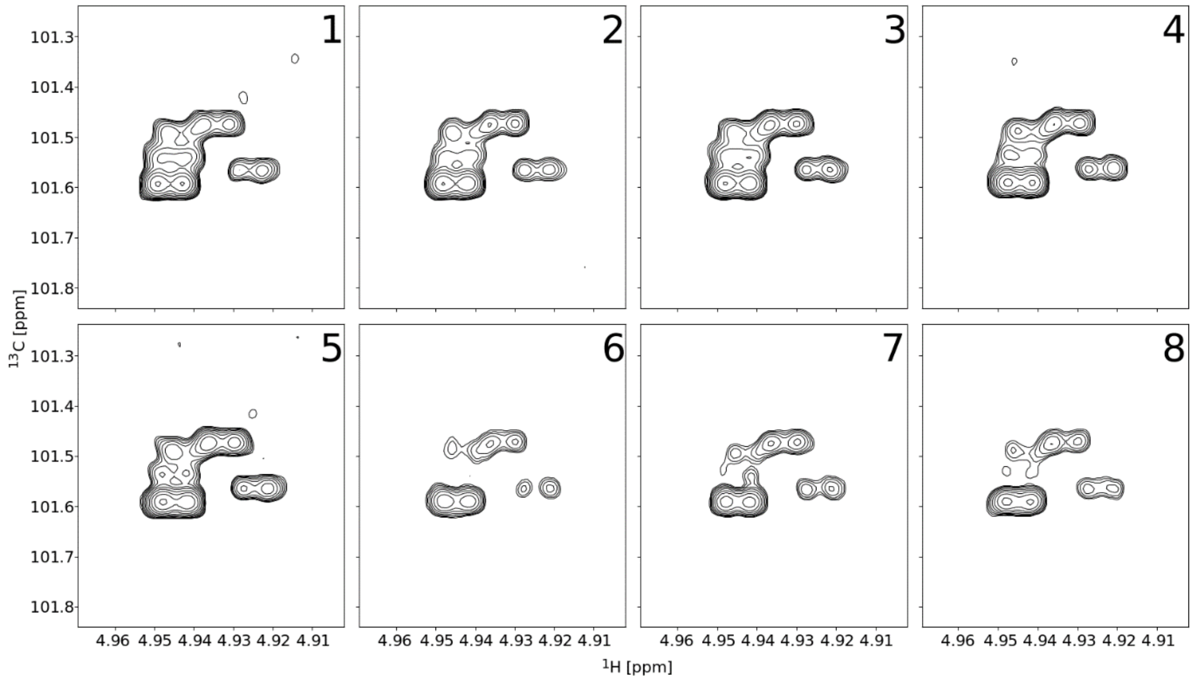
Appendix 3 Figure 2 – Tile plot showing the reducing end α-glucose peak present in the anomeric region of the eight different suppliers samples.



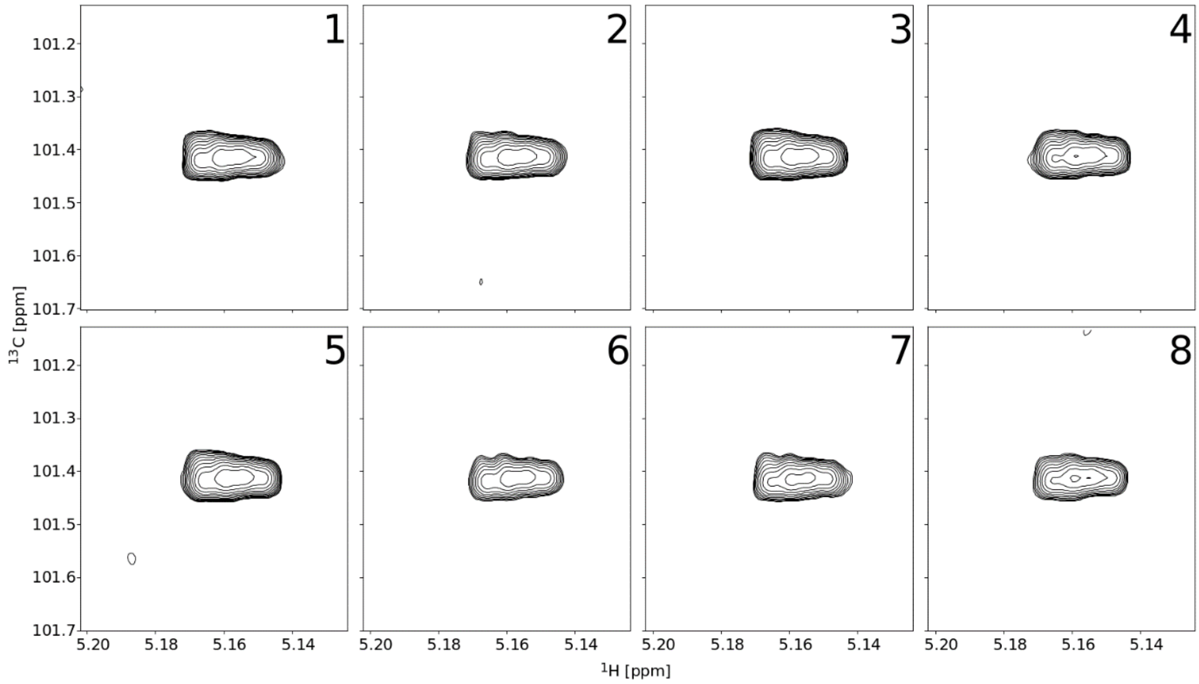
Appendix 3 Figure 3 – Tile plot showing the reducing end β-glucose peak present in the anomeric region of the eight different suppliers samples.



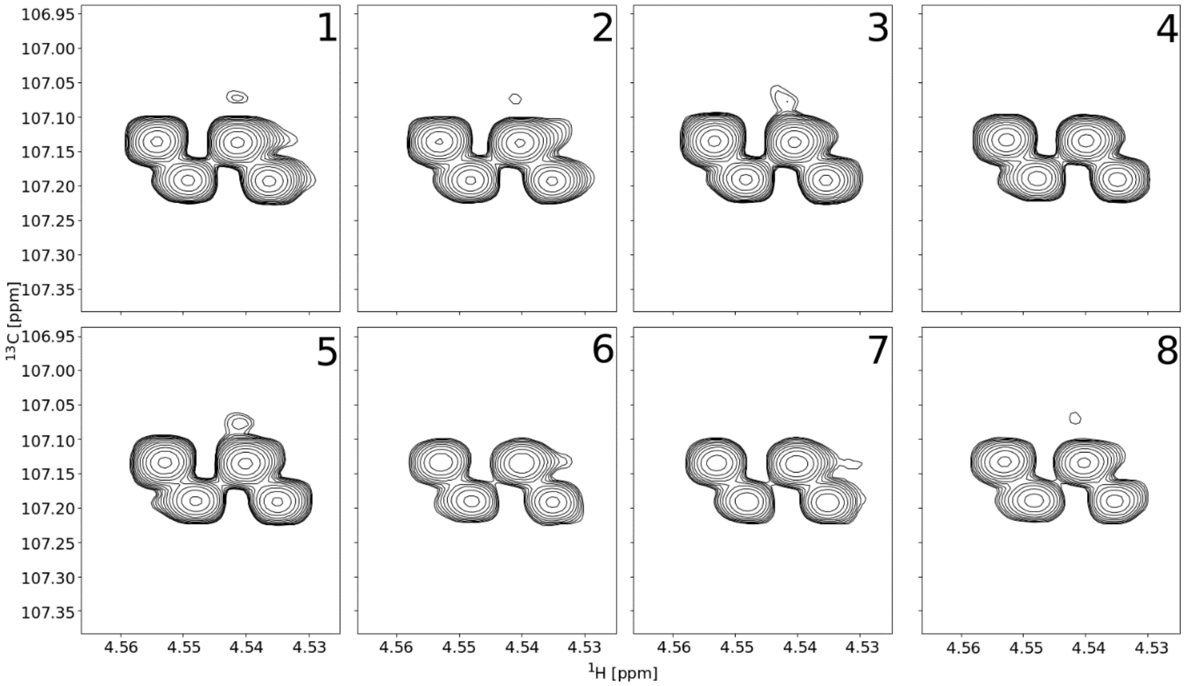
Appendix 3 Figure 4 – Tile plot showing the non-reducing end side chain terminal xylose peaks present in the anomeric region of the eight different suppliers samples.



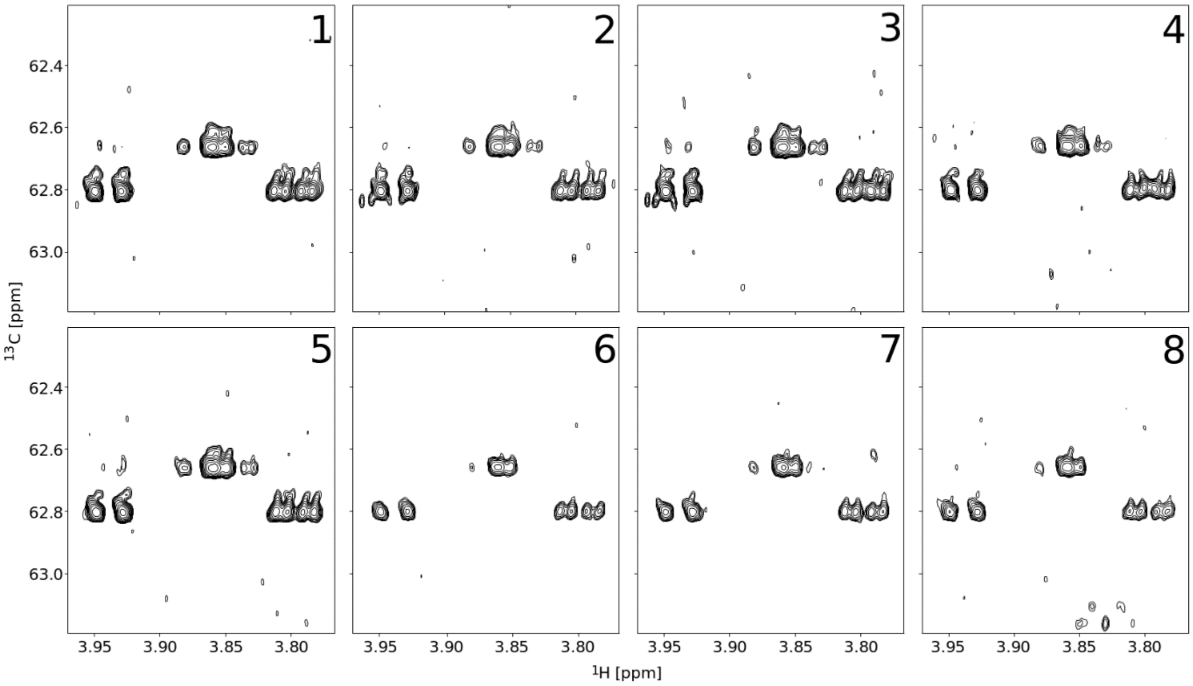
Appendix 3 Figure 5 – Tile plot showing the internal side chain terminal xylose peaks present in the anomeric region of the eight different suppliers samples.



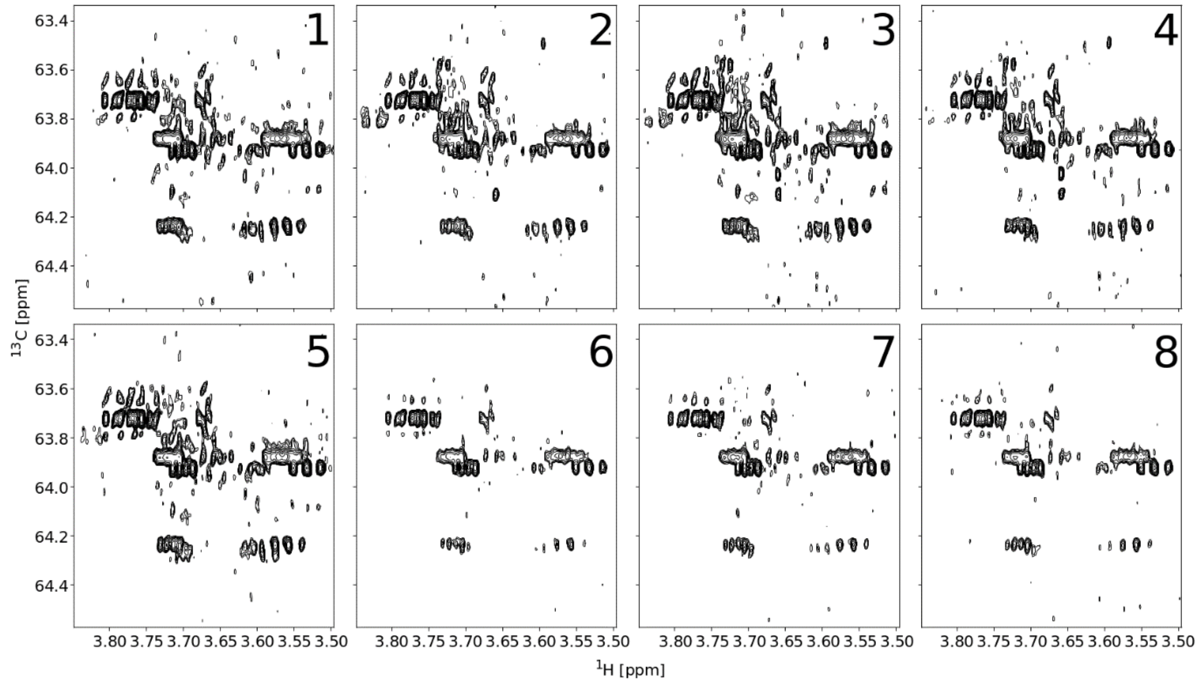
Appendix 3 Figure 6 – Tile plot showing the internal side chain further linked xylose peaks present in the anomeric region of the eight different suppliers samples.



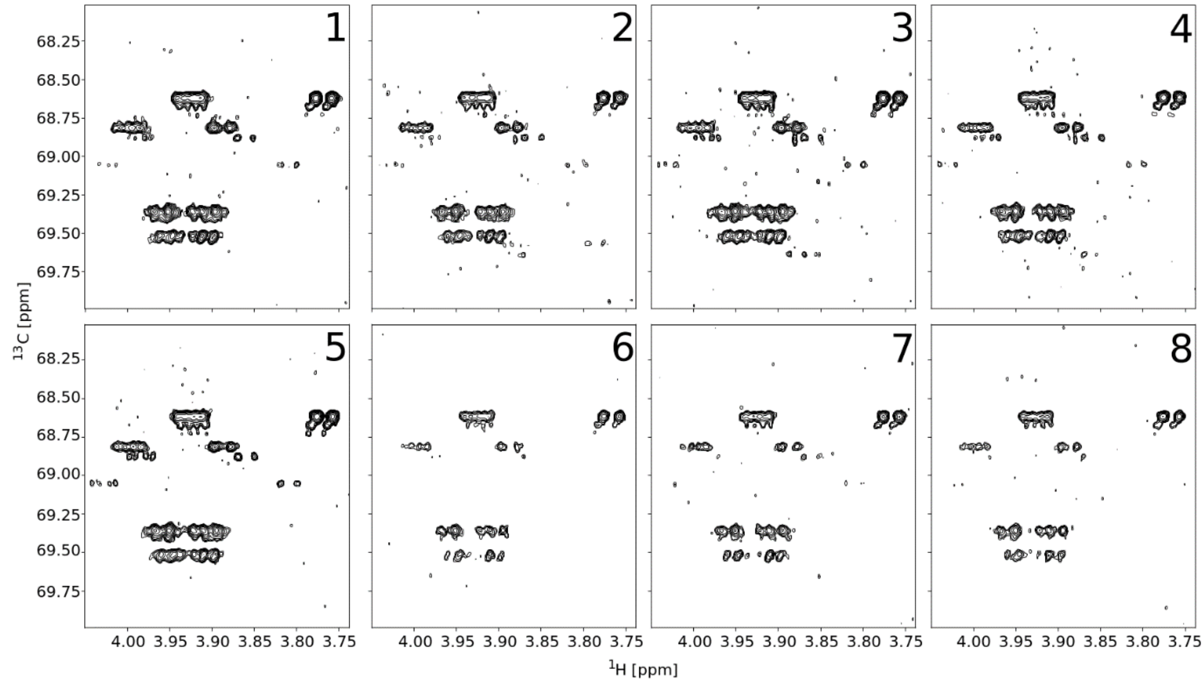
Appendix 3 Figure 7 – Tile plot showing the internal side chain terminal galactose peaks present in the anomeric region of the eight different suppliers samples.



Appendix 3 Figure 8 – Tile plot showing the unbound glucose peaks present in the CH_2 region of the eight different suppliers samples.

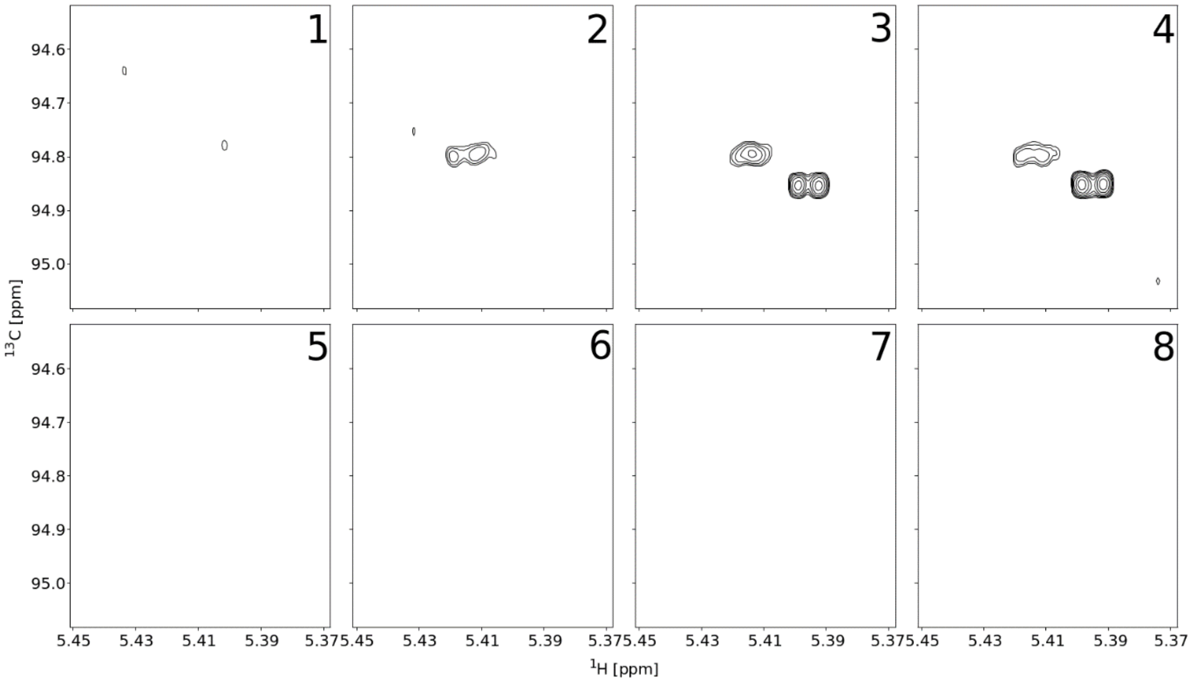


Appendix 3 Figure 9 – Tile plot showing the unbound xylose and galactose peaks present in the CH$_2$ region of the eight different suppliers samples.

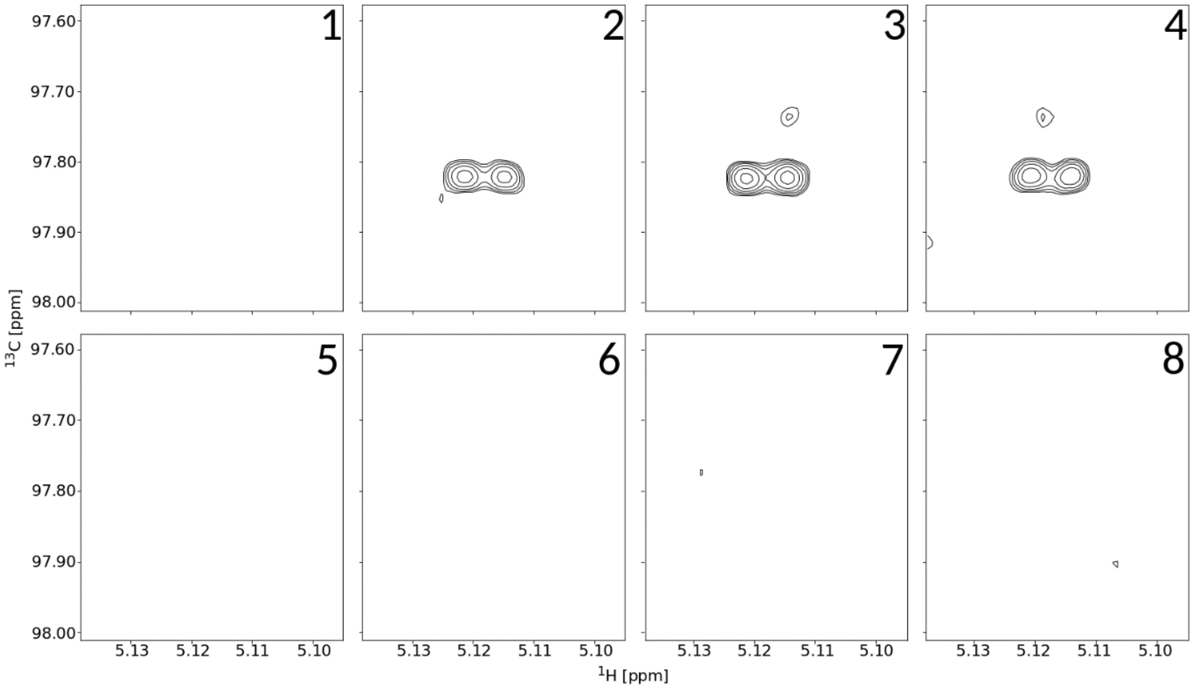


Appendix 3 Figure 10 – Tile plot showing the bound glucose peaks present in the CH$_2$ region of the eight different suppliers samples.

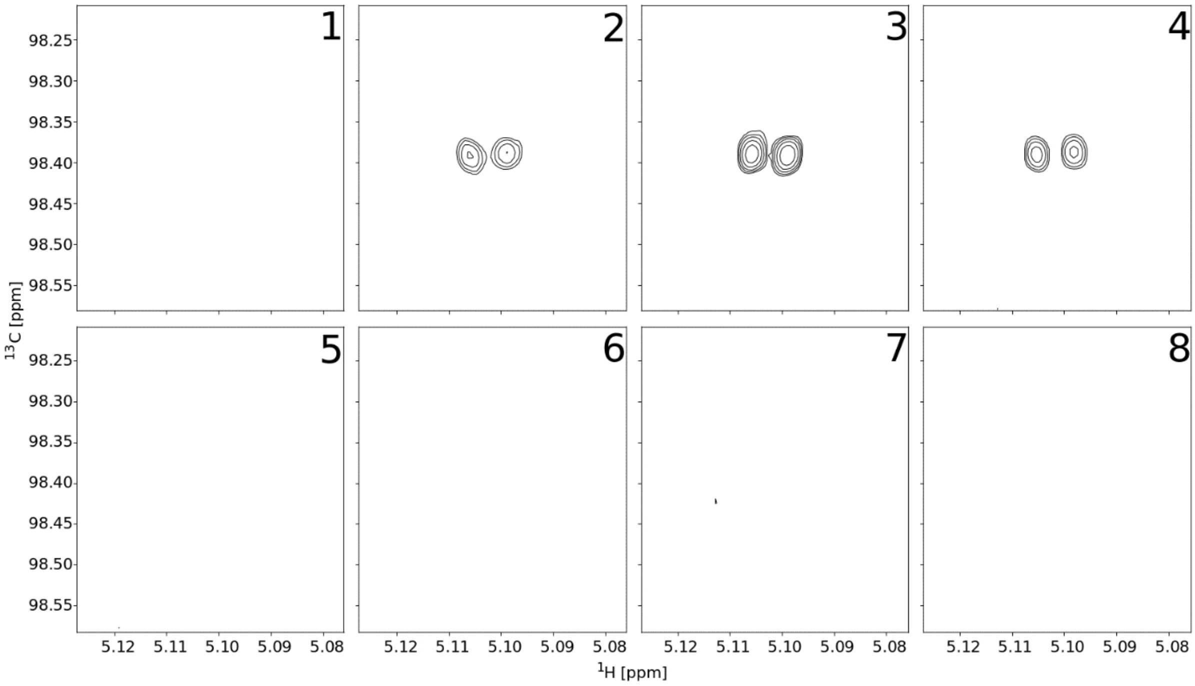
A3.2 – Different Supplier Xyloglucan Additional Peak Tile Plots:



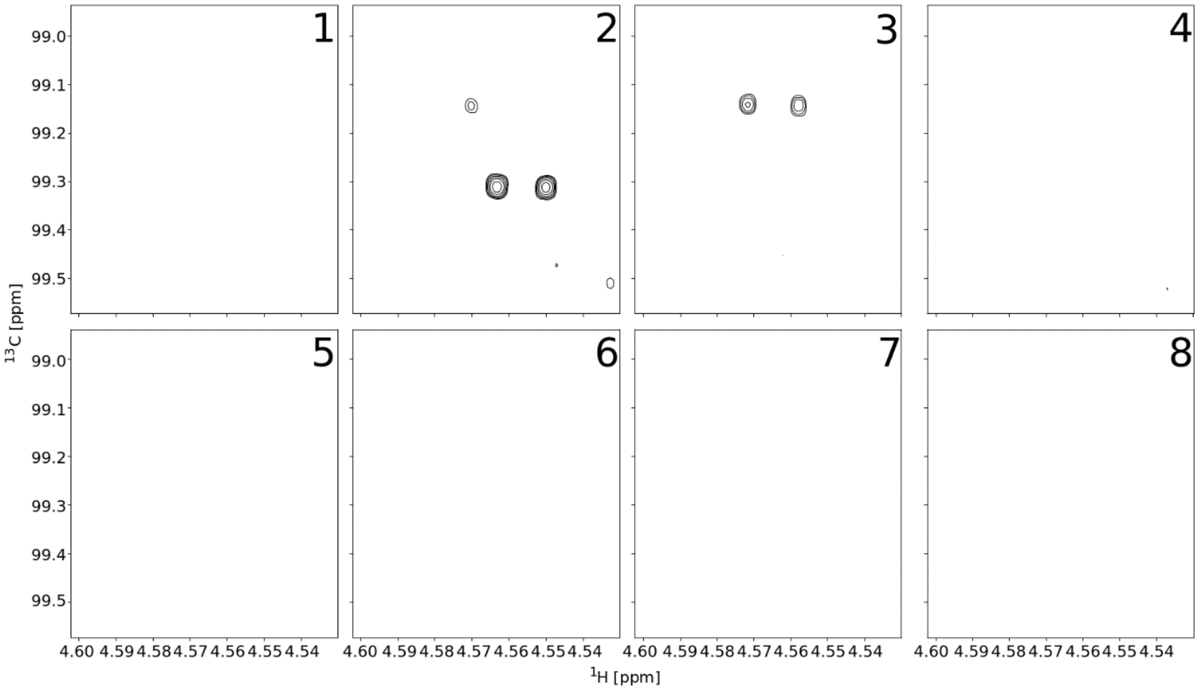
Appendix 3 Figure 11 – Tile plot showing the additional peak present at 5.42 – 5.39ppm and 98.8 – 98.9ppm in the ^{1}H and ^{13}C dimensions respectively of the anomeric region within the eight different suppliers samples.



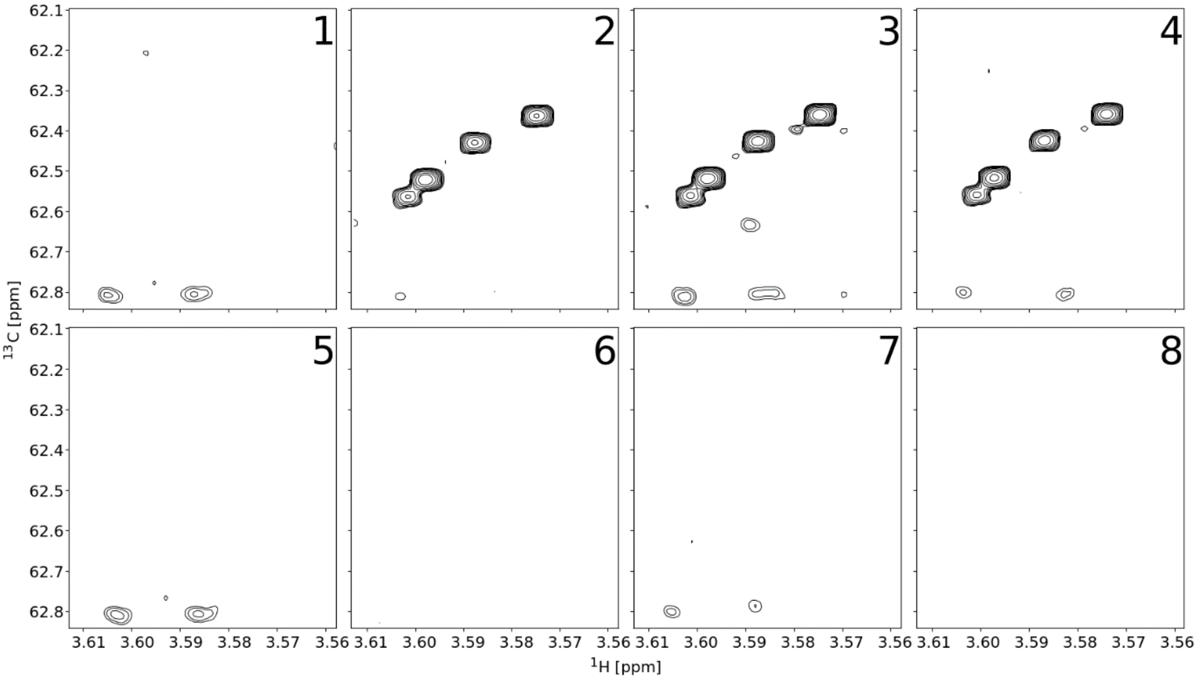
Appendix 3 Figure 12 – Tile plot showing the additional peak present at 5.12ppm and 97.80ppm in the ^{1}H and ^{13}C dimensions respectively of the anomeric region within the eight different suppliers samples.



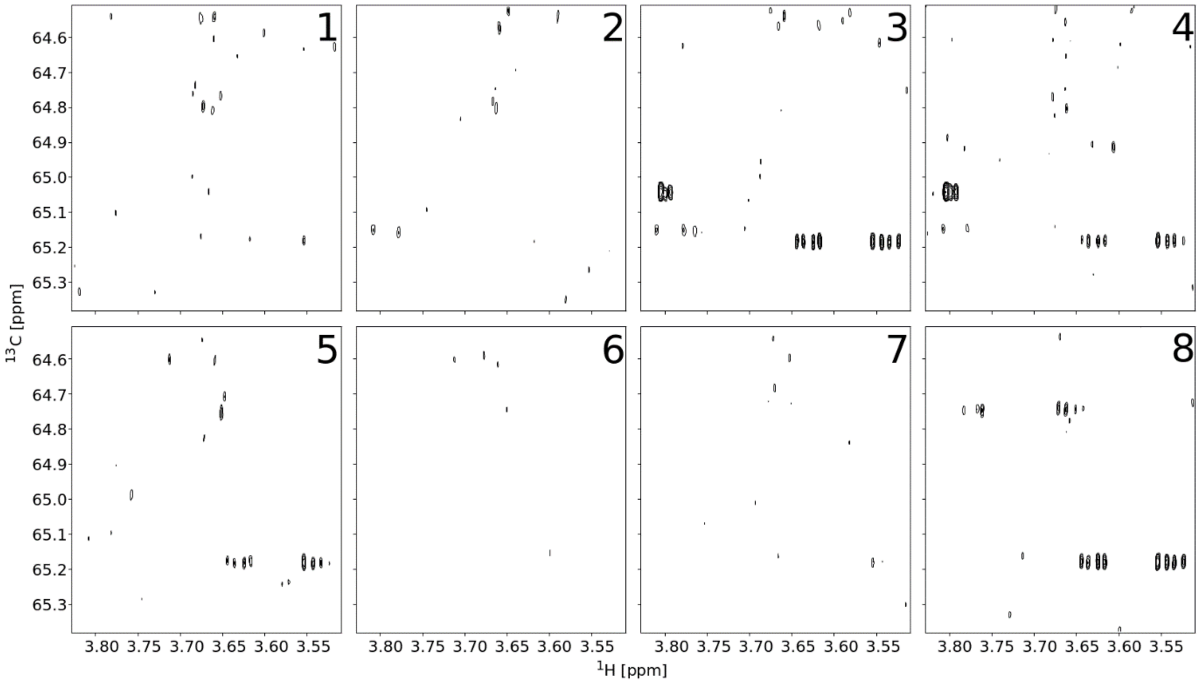
Appendix 3 Figure 13 – Tile plot showing the additional peak present at 5.10ppm and 98.40ppm in the ^{1}H and ^{13}C dimensions respectively of the anomeric region within the eight different suppliers samples.



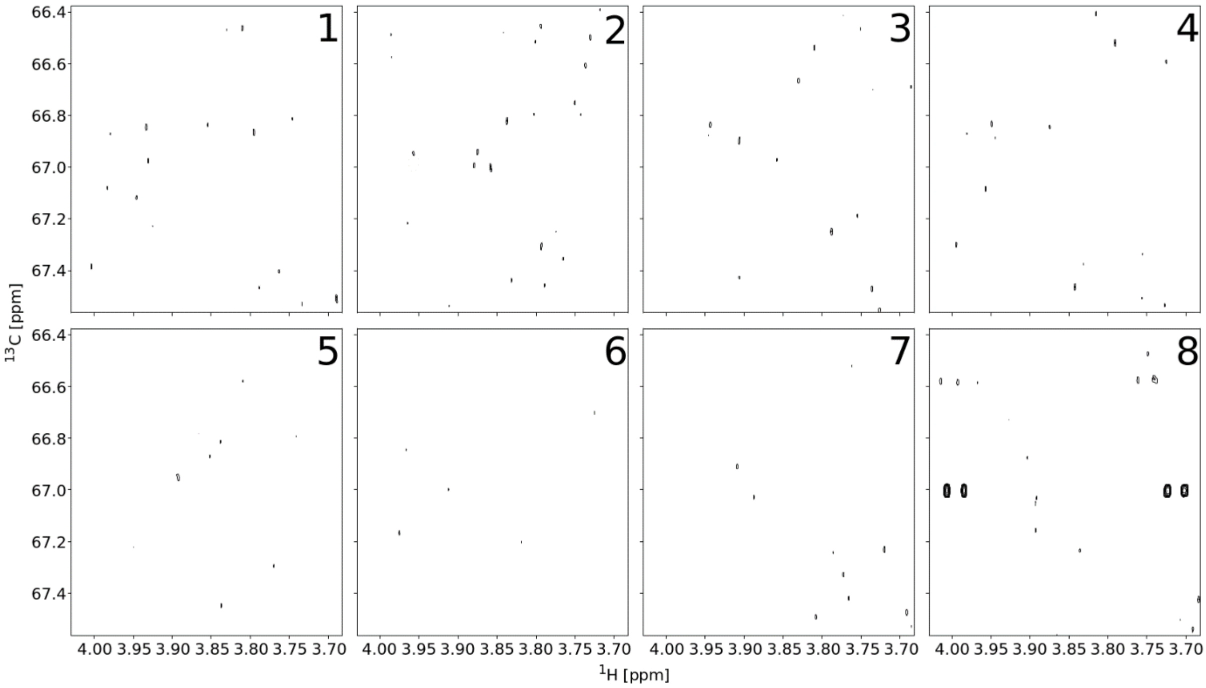
Appendix 3 Figure 14 – Tile plot showing the additional peak present at 4.57 – 4.55ppm and 99.1 – 99.4ppm in the ^{1}H and ^{13}C dimensions respectively of the anomeric region within the eight different suppliers samples.



Appendix 3 Figure 15 – Tile plot showing the additional peak present at 3.61 – 3.57ppm and 62.3 – 62.8ppm in the ^{1}H and ^{13}C dimensions respectively of the CH_2 region within the eight different suppliers samples.



Appendix 3 Figure 16 – Tile plot showing the additional peak present at 3.80 – 3.55ppm and 65.0 – 65.2ppm in the ^{1}H and ^{13}C dimensions respectively of the CH_2 region within the eight different suppliers samples.



Appendix 3 Figure 17 – Tile plot showing the additional peak present at 4.00 – 3.70ppm and 67.0ppm in the ^{1}H and ^{13}C dimensions respectively of the CH_2 region within the eight different suppliers samples.

A3.3 – Integral Intensities Tabulated Data:

Appendix 3 Table 1 – Different supplier xyloglucan samples anomeric region’s residue proportions table:

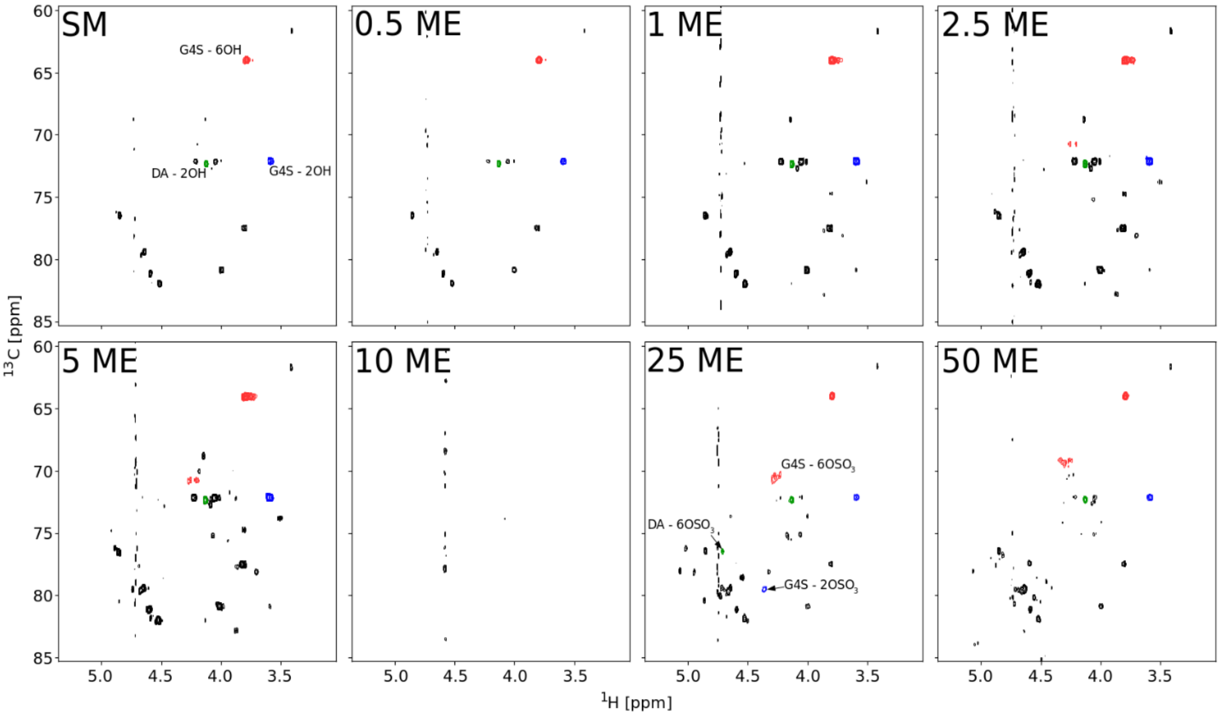
Assignment	Supplier 1	Supplier 2	Supplier 3	Supplier 4	Supplier 5	Supplier 6	Supplier 7	Supplier 8
α Glucose 4 P1	5.12	5.42	5.32	5.31	5.50	5.54	5.67	5.42
Unknown P1 (94.7 - 94.9ppm)	0.00	0.24	0.00	0.61	1.10	0.00	0.00	0.24
Unknown P1 (97.8ppm)	0.00	0.49	0.00	0.49	0.52	0.00	0.00	0.49
Unknown P1 (98.4ppm)	0.00	0.26	0.00	0.26	0.30	0.00	0.00	0.26
β Glucose 4 P1	8.41	8.07	8.24	8.18	8.55	8.59	8.48	8.07
Unknown P1 (99.1 - 99.4ppm)	0.00	0.85	0.00	0.36	0.00	0.00	0.00	0.85
Xylose (Further Linked 2 and 3) P1	17.22	16.19	16.41	15.85	15.98	16.15	16.64	16.19
Xylose (Terminal 1) P1	18.06	16.41	15.52	16.37	16.83	16.25	17.85	16.41
Xylose (Terminal 2 and 3) P1	5.62	6.23	6.96	6.47	5.98	6.32	5.71	6.23
Glucose 1, 2, and 3 P1	28.84	29.84	31.36	30.04	29.34	30.76	29.60	29.84
Galactose (Terminal 2 and 3) P1	16.73	15.99	16.19	16.07	15.90	16.39	16.05	15.99

Appendix 3 Table 2 – Different supplier xyloglucan samples CH_2 region’s residue proportions table:

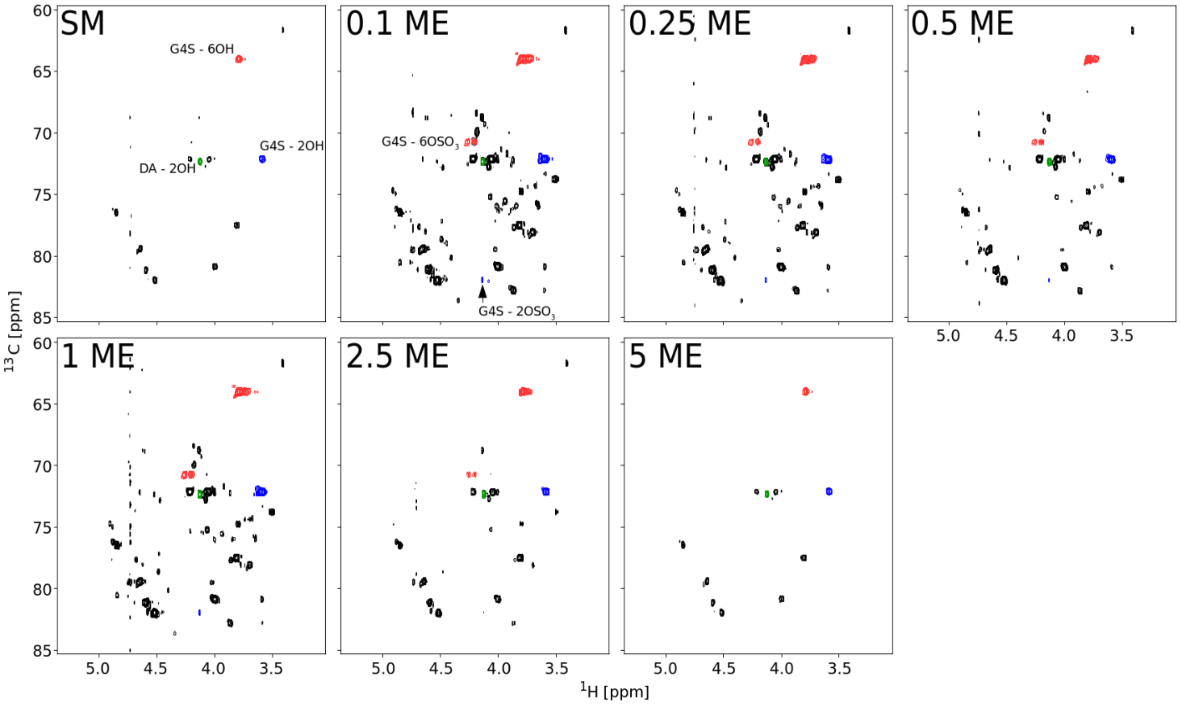
Assignment	Supplier 1	Supplier 2	Supplier 3	Supplier 4	Supplier 5	Supplier 6	Supplier 7	Supplier 8
Unbound Glucose P6	12.26	12.09	12.72	12.16	12.56	12.77	13.15	12.26
Unknown P6 (62.3 - 62.6ppm)	0.00	1.80	0.00	2.13	1.95	0.00	0.00	0.00
Xylose and Galactose P6	54.97	54.69	55.28	54.06	53.38	55.09	55.15	54.97
Unknown P6 (64.2ppm)	0.00	0.00	0.00	0.80	1.24	0.00	0.00	0.00
Unknown P6 (67.0ppm)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bound Glucose P6	32.77	31.42	32.01	30.85	30.86	32.14	31.70	32.77

Appendix 4 – Chapter 5 Sulfation Spectra

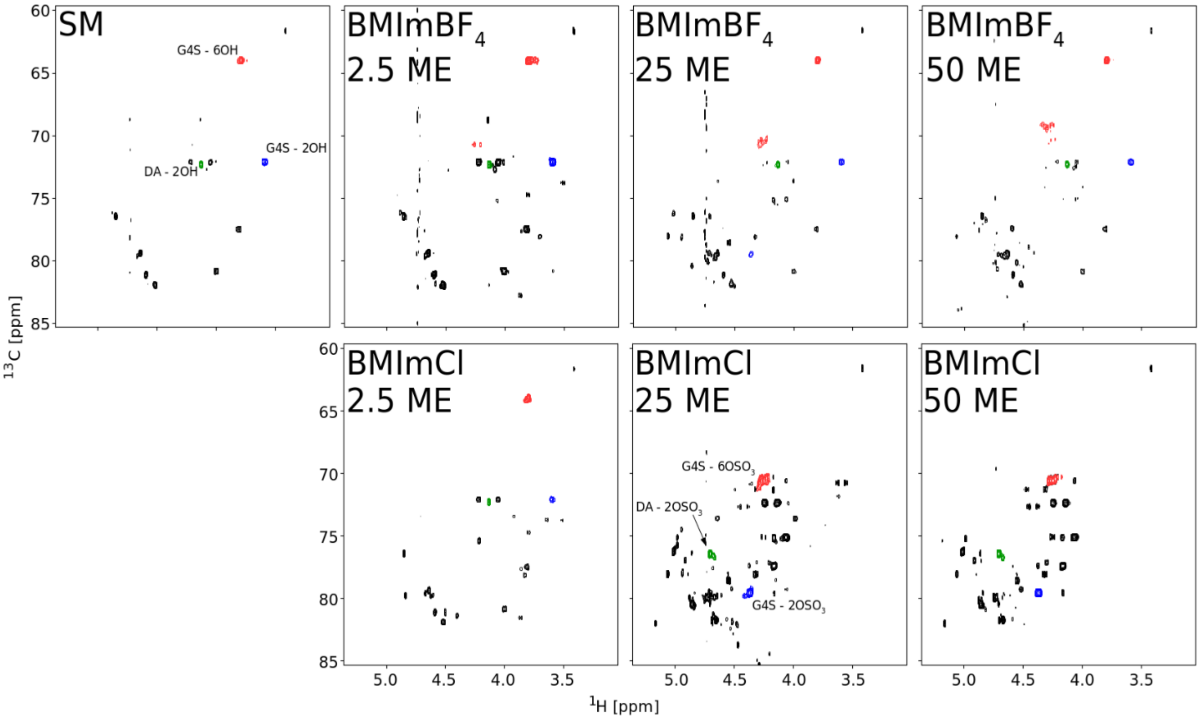
A4.1 – Sulfation in BMImCl Reaction Optimisation using κ-Carrageenan:



Appendix 4 Figure 1: κ-Carrageenan in BMImBF$_4$ reaction PyrSO$_3$ optimisation spectra for the starting material and Pyr·SO$_3$ concentrations between 0.5 and 50 molar equivalents with the galactose-4-sulfate (G4S) positions six, red, and two, blue, and 3,6-anhdrogalactose (DA) position two, green, hydroxyl and sulfated hydroxyl group peaks indicated.

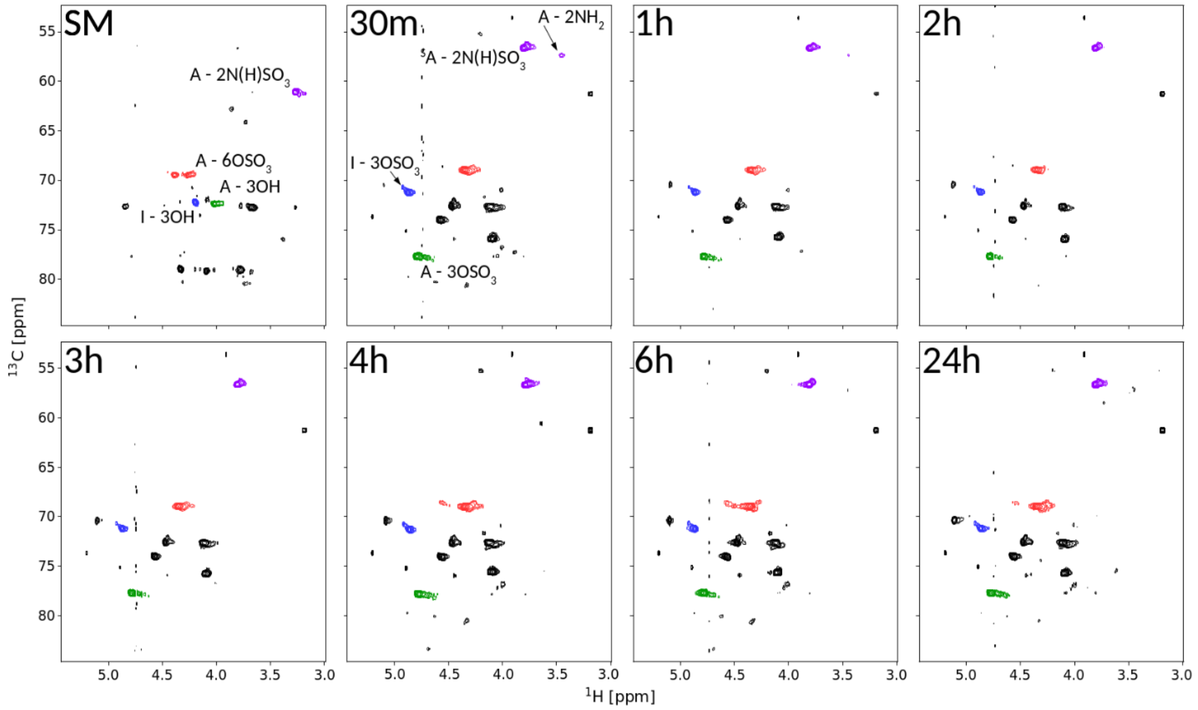


Appendix 4 Figure 2: κ-Carrageenan in BMImBF$_4$ reaction DMAP optimisation spectra for the starting material and DMAP concentrations between 0.1 and 5 molar equivalents with the galactose-4-sulfate (G4S) positions six, red, and two, blue, and 3,6-anhydrogalactose (DA) position two, green, hydroxyl and sulfated hydroxyl group peaks indicated.

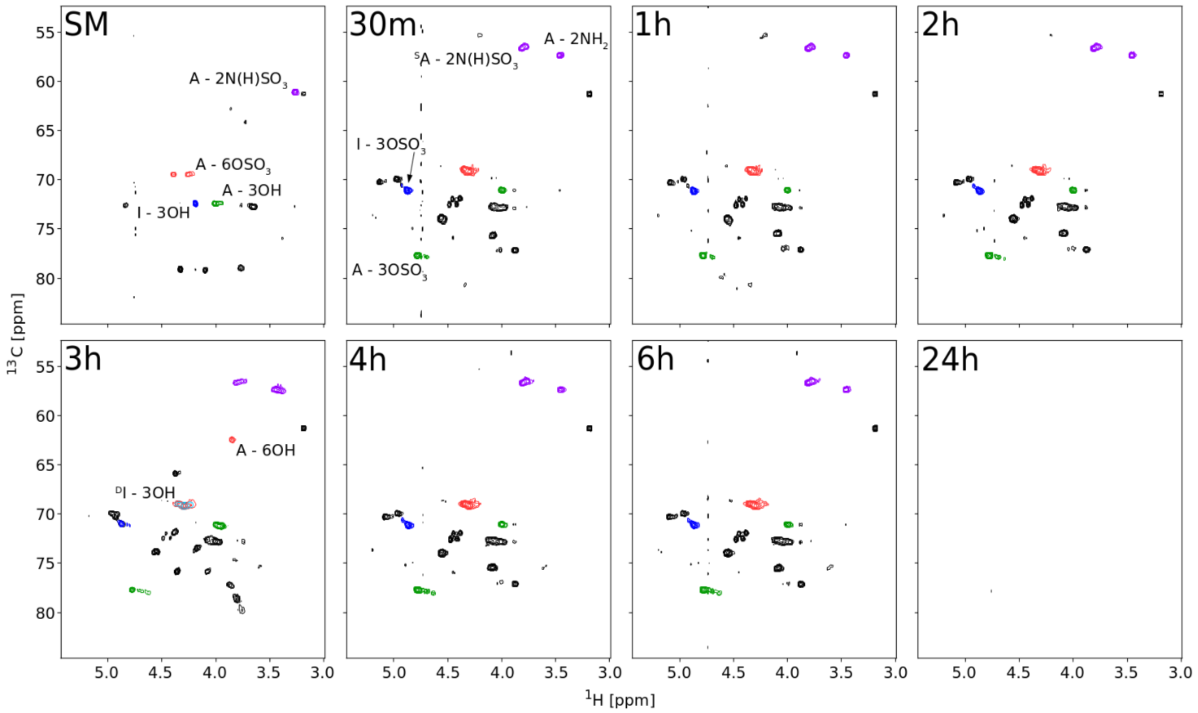


Appendix 4 Figure 3: κ-Carrageenan solvent, BMImBF$_4$ or BMImCl, optimisation spectra for the starting material and Pyr·SO$_3$ concentrations of 2.5, 25, and 50 molar equivalents in both solvents with the galactose-4-sulfate (G4S) positions six, red, and two, blue, and 3,6-anhydrogalactose (DA) position two, green, hydroxyl and sulfate group peaks indicated.

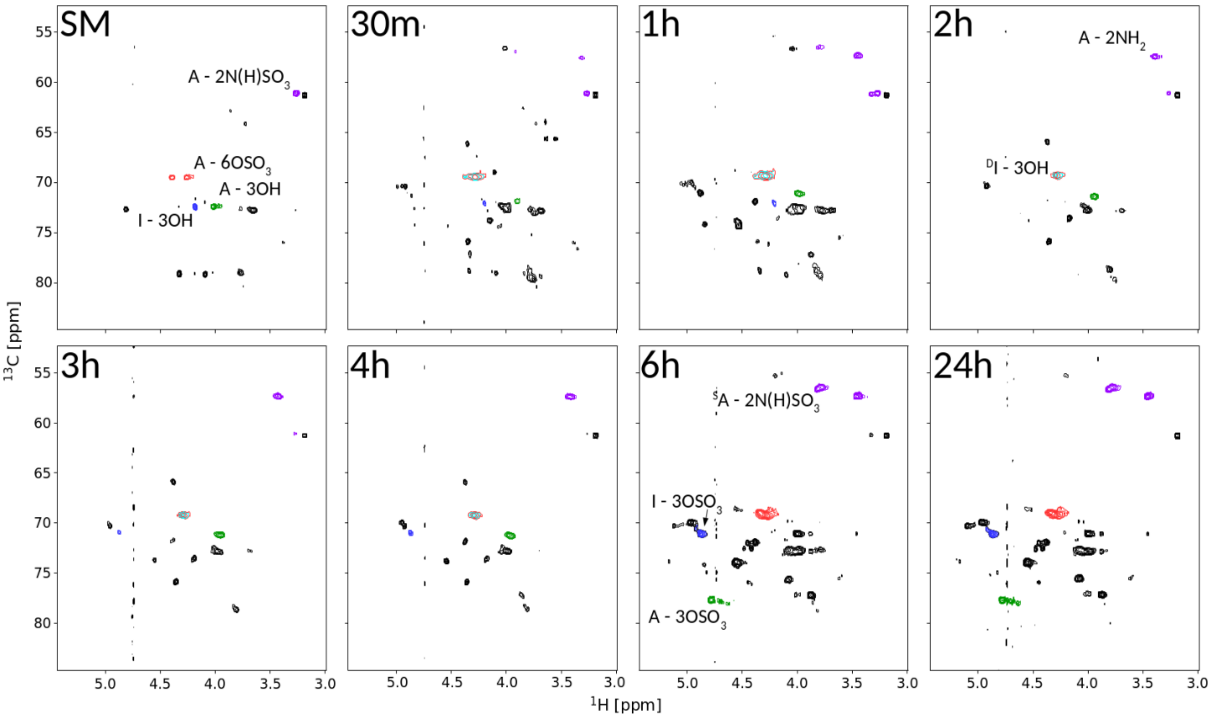
A4.2 – Heparin Pyr·SO$_3$ Sulfation in BMImCl and DMF Reaction Spectra:



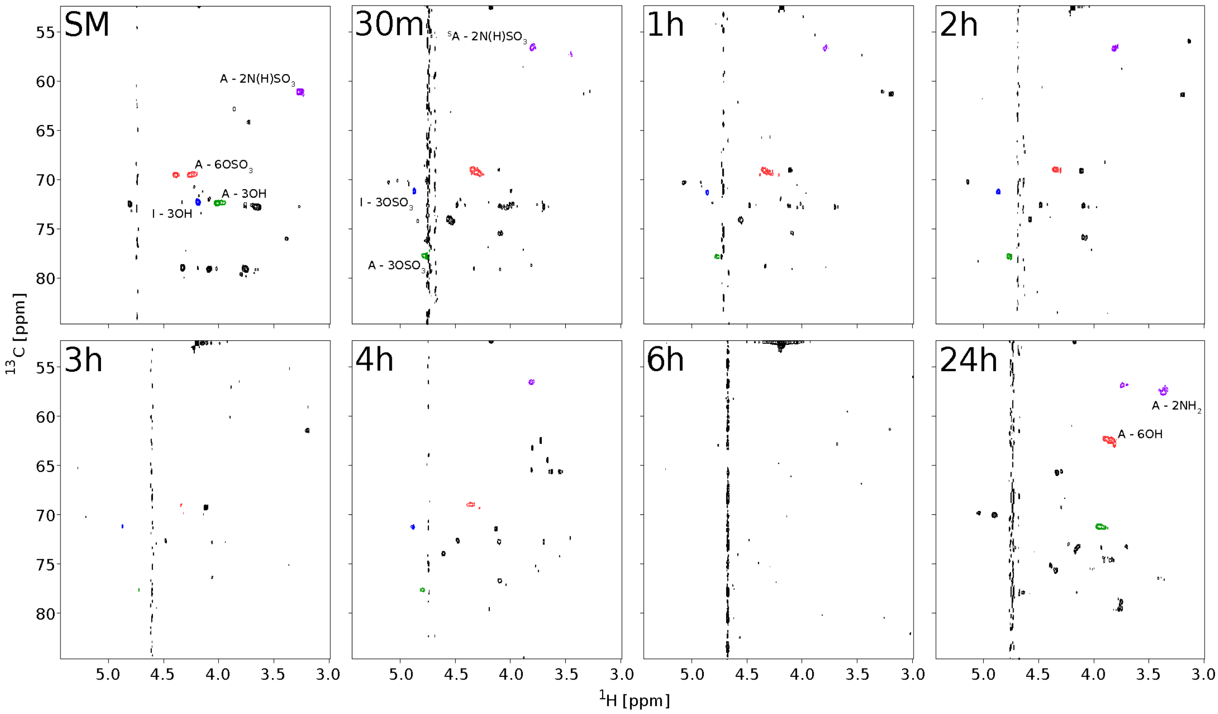
Appendix 4 Figure 4: TBA$^+$ heparin 85ºC sulfation time course, using DMF solvation, reaction spectra for the starting material and thirty minutes as well as one, two, three, four, six, and twenty-four hour timepoints with the glucosamine (A) positions six, red, and three, green, and purple, green, as well as the iduronic acid (I) position two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.



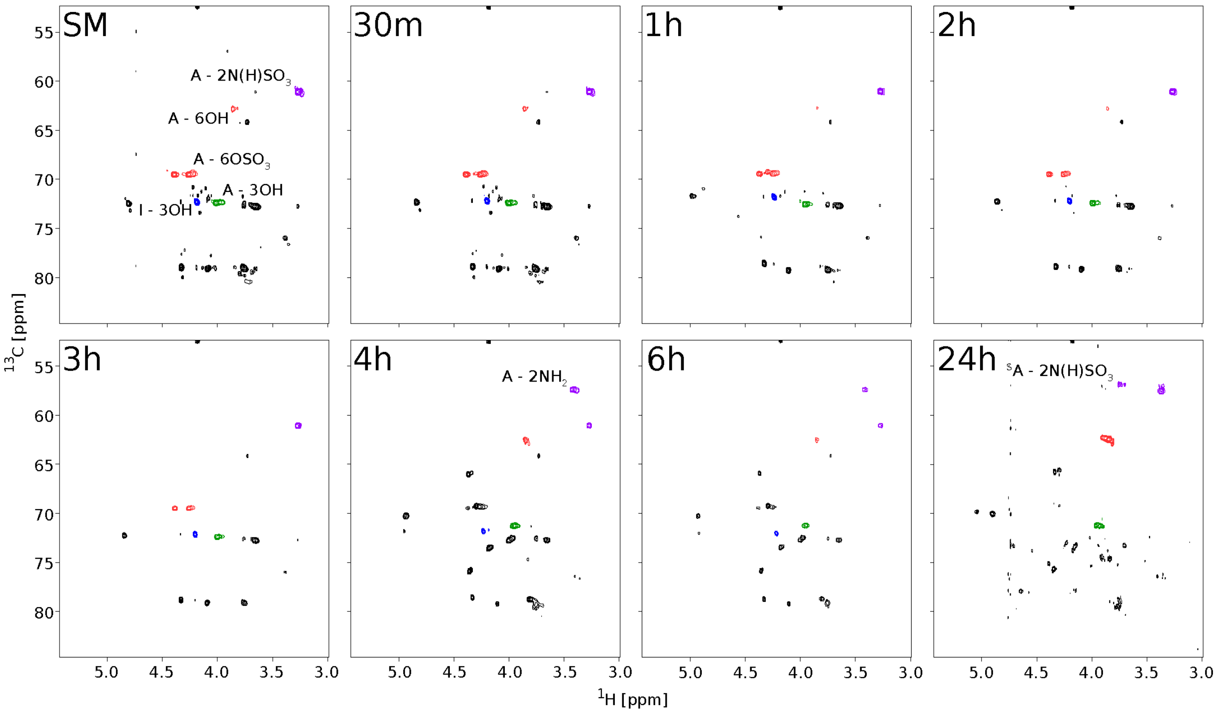
Appendix 4 Figure 5: TBA$^+$ heparin 55ºC sulfation time course, using DMF solvation, reaction spectra for the starting material and thirty minutes as well as one, two, three, four, six, and twenty-four hour timepoints with the glucosamine (A) positions six, red, and three, green, and purple, green, as well as the iduronic acid (I) position two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.



Appendix 4 Figure 6: TBA⁺ heparin 35ºC sulfation time course, using DMF solvation, reaction spectra for the starting material and thirty minutes as well as one, two, three, four, six, and twenty-four hour timepoints with the glucosamine (A) positions six, red, and three, green, and purple, green, as well as the iduronic acid (I) position two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.



Appendix 4 Figure 7: TBA⁺ heparin 85ºC sulfation time course, using the optimised conditions in BMImCl, reaction spectra for the starting material and thirty minutes as well as one, two, three, four, six, and twenty-four hour timepoints with the glucosamine (A) positions six, red, and three, green, and purple, green, as well as the iduronic acid (I) position two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.



Appendix 4 Figure 8: Na$^+$ heparin 85°C sulfation time course, using the optimised conditions in BMImCl, reaction spectra for the starting material and thirty minutes as well as one, two, three, four, six, and twenty-four hour timepoints with the glucosamine (A) positions six, red, and three, green, and purple, green, as well as the iduronic acid (I) position two, blue, hydroxyl and sulfated hydroxyl group peaks indicated.

A4.3 – Pyr·SO_3 Sulfation in BMImCl and DMF Integral Intensity and Residue Proportion Values:

The integral intensity values recorded were generated courtesy of scripts provided by Dr. Rudi Grosman, these were then used to calculate the proportion of sulfated and unsulfated residues at each atom position.

Optimisation:

Appendix 4 Table 1 – Integral intensity values recorded for κ-Carrageenan Pyr·SO_3 Optimisation spectra from the scripts provided:

Assignment	Starting Material	0.5 Mol. Eq.	1 Mol. Eq.	2.5 Mol. Eq.	5 Mol. Eq.	10 Mol. Eq.	25 Mol. Eq.	50 Mol. Eq.
G4S - 6OH	18075500.2	18075500.2	43097845.3	54582251.7	64093439.2		13412804.7	15399689.9
G4S - 6OSO3	0.0	0.0	1623361.7	5442917.3	10718625.5		13512936.2	17864616.3
G4S - 2OH	7144696.7	7144696.7	21439070.8	28620810.1	35878379.7		4918884.8	6951494.8
G4S - 2OSO3	194438.5	194438.5	851405.2	983560.2	2093534.5		7198357.1	5153352.4
DA - 2OH	5700138.7	5700138.7	16357920.7	21717481.2	28655507.4		6137444.0	6111053.4
DA - 2OSO3	0.0	0.0	199859.0	0.0	256468.6		1303484.0	0

Appendix 4 Table 2 – Sulfated and Unsulfated residue proportions for κ-Carrageenan Pyr·SO_3 Optimisation spectra, calculated from the integrals in Appendix 4 Table 1:

Assignment	Starting Material	0.5 Mol. Eq.	1 Mol. Eq.	2.5 Mol. Eq.	5 Mol. Eq.	10 Mol. Eq.	25 Mol. Eq.	50 Mol. Eq.
DA - 2OH	100.0	100.0	98.8	100.0	99.1		82.5	100
DA - 2OSO3	0.0	0.0	1.2	0.0	0.9		17.5	0
G4S - 2OH	97.4	97.4	96.2	96.7	94.5		40.6	57.4
G4S - 2OSO3	2.6	2.6	3.8	3.3	5.5		59.4	42.6
G4S - 6OH	100.0	100.0	96.4	90.9	85.7		49.8	46.3
G4S - 6OSO3	0.0	0.0	3.6	9.1	14.3		50.2	53.7

Appendix 4 Table 3 – Integral intensity values recorded for κ-Carrageenan DMAP Optimisation spectra from the scripts provided:

Assignment	Starting Material	0.1 Mol. Eq.	0.25 Mol. Eq.	0.5 Mol. Eq.	1 Mol. Eq.	2.5 Mol. Eq.	5 Mol. Eq.
G4S - 6OH	18075500.2	203837904.6	153379184.2	91843675.7	176467620.7	49693179.8	19706495.0
G4S - 6OSO3	0.0	45753065.3	29011551.8	14469604.5	30443503.7	8241860.5	1641964.7
G4S - 2OH	7144696.7	115540051.0	80967485.9	48605578.6	98313560.4	27785135.2	9587585.3
G4S - 2OSO3	194438.5	7588942.6	4324219.7	1589164.4	3503494.2	970647.0	760978.9
DA - 2OH	5700138.7	65851945.8	50689900.1	34737375.6	64083847.3	19947591.6	7235727.5
DA - 2OSO3	0.0	366851.0	0.0	0.0	0.0	21404.0	140829.4

Appendix 4 Table 4 – Sulfated and Unsulfated residue proportions for κ-Carrageenan DMAP Optimisation spectra, calculated from the integrals in Appendix 4 Table 3:

Assignment	Starting Material	0.1 Mol. Eq.	0.25 Mol. Eq.	0.5 Mol. Eq.	1 Mol. Eq.	2.5 Mol. Eq.	5 Mol. Eq.
DA - 2OH	100.0	99.4	100.0	100.0	100.0	99.9	98.1
DA - 2OSO3	0.0	0.6	0.0	0.0	0.0	0.1	1.9
G4S - 2OH	97.4	93.8	94.9	96.8	96.6	96.6	92.6
G4S - 2OSO3	2.6	6.2	5.1	3.2	3.4	3.4	7.4
G4S - 6OH	100.0	81.7	84.1	86.4	85.3	85.8	92.3
G4S - 6OSO3	0.0	18.3	15.9	13.6	14.7	14.2	7.7

Appendix 4 Table 5 – Integral intensity values recorded for κ-Carrageenan Solvent Optimisation spectra from the scripts provided:

Assignment	Starting Material	BMImBF$_4$			BMImCl		
		2.5 Mol. Eq.	25 Mol. Eq.	50 Mol. Eq.	2.5 Mol. Eq.	25 Mol. Eq.	50 Mol. Eq.
G4S - 6OH	18075500.2	54582251.7	13412804.7	15399689.9	98772941.6	1952841.6	792912.7
G4S - 6OSO3	0.0	5442917.3	13512936.2	17864616.3	0.0	413177928.6	437035038.8
G4S - 2OH	7144696.7	28620810.1	4918884.8	6951494.8	44785367.1	28194448.2	6786003.1
G4S - 2OSO3	194438.5	983560.2	7198357.1	5153352.4	3399529.9	195532032.0	209256183.0
DA - 2OH	5700138.7	21717481.2	6137444.0	6111053.4	29242372.9	115193510.7	97194116.6
DA - 2OSO3	0.0	0.0	1303484.0	0.0	0.0	129198770.8	168922458.8

Appendix 4 Table 6 – Sulfated and Unsulfated residue proportions for κ-Carrageenan Solvent Optimisation spectra, calculated from the integrals in Appendix 4 Table 5:

Assignment	Starting Material	BMImBF$_4$			BMImCl		
		0.1 Mol. Eq.	0.25 Mol. Eq.	0.5 Mol. Eq.	1 Mol. Eq.	5 Mol. Eq.	5 Mol. Eq.
DA - 2OH	100.0	90.9	49.8	46.3	100.0	0.5	0.2
DA - 2OSO3	0.0	9.1	50.2	53.7	0.0	99.5	99.8
G4S - 2OH	97.4	96.7	40.6	57.4	92.9	12.6	3.1
G4S - 2OSO3	2.6	3.3	59.4	42.6	7.1	87.4	96.9
G4S - 6OH	100.0	100.0	82.5	100.0	100.0	47.1	36.5
G4S - 6OSO3	0.0	0.0	17.5	0.0	0.0	52.9	63.5

Plant-based Carbohydrate Time Courses:

Appendix 4 Table 7 – Integral intensity values recorded for κ-Carrageenan time course spectra from the scripts provided:

Assignment	Starting Mat.	30m	1h	2h	3h	4h	6h	24h
G4S - 6OH	18075500.2	46091463.0		14656499.7	29537874.9	1742519.3	1397931.0	1630658.7
G4S - 6OSO3	0.0	4586028.1		5439823.9	7030134.5	153689442.2	169161564.7	333463116.9
G4S - 2OH	7144696.7	8618571.4		6506189.1	6744505.0	3228471.8	3538533.9	7113870.9
G4S - 2OSO3	194438.5	931604.3		532566.9	2331858.8	68974833.6	81895712.4	169073249.2
DA - 2OH	5700138.7	6861634.9		4855424.6	6049289.3	25337528.3	39826813.9	94712755.1
DA - 2OSO3	0.0	399400.9		0.0	427149.1	49843027.2	61830217.0	131115719.2

Appendix 4 Table 8 – Sulfated and Unsulfated residue proportions for κ-Carrageenan time course spectra, calculated from the integrals in Appendix 4 Table 7:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
DA - 2OH	100.0	91.0		72.9	80.8	1.1	0.8	0.5
DA - 2OSO3	0.0	9.0		27.1	19.2	98.9	99.2	99.5
G4S - 2OH	97.4	90.2		92.4	74.3	4.5	4.1	4.0
G4S - 2OSO3	2.6	9.8		7.6	25.7	95.5	95.9	96.0
G4S - 6OH	100.0	94.5		100.0	93.4	33.7	39.2	41.9
G4S - 6OSO3	0.0	5.5		0.0	6.6	66.3	60.8	58.1

Appendix 4 Table 9 – Integral intensity values recorded for ι-Carrageenan time course spectra from the scripts provided:

Assignment	Starting Mat.	30m	1h	2h	3h	4h	6h	24h
G4S - 6OH	180300304.2	5597098.0	5787233.4	7195473.1	4095975.5	1476842.4	2965581.8	8175112.7
G4S - 6OSO3	5506716.5	5430744.8	9946221.6	6459041.4	9154279.3	4548199.3	8214518.8	18900113.9
G4S - 2OH	54309715.1	253240.5	1486925.8	1108353.3	840440.0	1005026.7	2043951.5	28039.6
G4S - 2OSO3	4186925.5	1639416.8	4063257.1	2272674.5	3227043.9	1557830.8	3319631.6	8484997.4

Appendix 4 Table 10 – Sulfated and Unsulfated residue proportions for ι-Carrageenan time course spectra, calculated from the integrals in Appendix 4 Table 9:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
DA - 2OH	92.8	13.4	26.8	32.8	20.7	39.2	38.1	0.3
DA - 2OSO3	7.2	86.6	73.2	67.2	79.3	60.8	61.9	99.7
G4S - 2OH	97.0	50.8	36.8	52.7	30.9	24.5	26.5	30.2
G4S - 2OSO3	3.0	49.2	63.2	47.3	69.1	75.5	73.5	69.8

Appendix 4 Table 11 – Integral intensity values recorded for Cellulose time course spectra from the scripts provided:

Assignment	Starting Mat.	30m	1h	2h	3h	4h	6h	24h
2 - OH	5901488433.3	133451541.2	11421445.7	3878062.5	1724355.4	962766.7	4761824.0	4680736.8
2 - OSO3	48251455.0	1527750.9	59376825.4	85377220.0	77592242.8	89735512.0	66901338.2	91978917.0
3 - OH	4364971143.1	132360441.8	7412165.2	1413714.8	1360754.7	120877.6	3174430.1	1942801.5
3 - OSO3	269110034.1	2390416.5	44775951.4	64023270.0	60976063.3	71195755.4	53406016.2	68871064.4
6 - OH	10783541819.0	58560562.9	92345.2	0.0	0.0	0.0	0.0	0.0
6 - OSO3	354982285.8	110147937.7	96400012.9	105482870.0	100889064.8	131978914.9	110439220.1	147078433.0

Appendix 4 Table 12 – Sulfated and Unsulfated residue proportions for Cellulose time course spectra, calculated from the integrals in Appendix 4 Table 11:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
2 - OH	99.2	98.9	16.1	4.3	2.2	1.1	6.6	4.8
2 - OSO3	0.8	1.1	83.9	95.7	97.8	98.9	93.4	95.2
3 - OH	94.2	98.2	14.2	2.2	2.2	0.2	5.6	2.7
3 - OSO3	5.8	1.8	85.8	97.8	97.8	99.8	94.4	97.3
6 - OH	96.8	34.7	0.1	0.0	0.0	0.0	0.0	0.0
6 - OSO3	3.2	65.3	99.9	100.0	100.0	100.0	100.0	100.0

Appendix 4 Table 13 – Integral intensity values recorded for Xyloglucan time course spectra from the scripts provided:

Peak	Starting Material	30m	1h	2h	3h	4h	6h	24h
1	0	152521	269513.3	61897.57	242008.8	188879.8	293647.4	5786235
2	165354.8314	80916.66	0	0	289806	158000.6	9355400	6780422
3	0	258594.2	737306.9	1644496	2294766	1978421	16857086	9354902
4	268716.0065	0	0	141137.5	1824996	767319.5	33283538	26137091
5	0	111313.8	379065.6	933338.5	1824043	1117220	19392404	18895031
6	189921.7445	0	0	0	2607268	653846.8	24081462	11226774
7	158308.086	118413.8	310665.3	0	0	0	6554855	22795943
8	78274.35244	84209.64	0	320369.2	3956169	1587098	32953155	15917056
9	232221.7988	203576.7	423187.6	1914884	10934060	7363849	80959194	43708070
10	0	113706.9	114461	0	660460.3	259370.9	6990560	26745769
11	0	68435.36	466538	369675.1	2060462	1183121	27674111	44420001
12	84840.65796	0	56325.12	236565.9	1106771	427721.5	13702417	4022462

Heparin Time Courses:

Appendix 4 Table 14 – Integral intensity values recorded for Heparin 85°C DMF solvated time course spectra from the scripts provided:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	369594.6	8393844.0	5272035.0	3038003.5	3633849.5	6849531.0	1131986.1	9470101.0
A - NSO3 SM	15832840.0	258714.3	0.0	224477.6	0.0	0.0	723233.9	0.0
A - NSO3 Slp	3308210.0	81802770.0	51353080.0	46423024.0	59415450.0	90159384.0	79150790.0	93290370.0
A - 6OH	8881180.0	1471699.6	1611607.8	1721427.2	1083949.9	2088502.1	2357823.0	2089379.8
A - 6OSO3	25616476.0	94948620.0	64393380.0	56799070.0	74236140.0	123764410.0	109545670.0	127522760.0
A - 3OH	24958360.0	3472486.8	2063615.6	1556003.1	2040381.6	4078950.2	379163.1	4336798.0
A - 3OSO3	32019.5	75044170.0	46523100.0	44531584.0	54700344.0	90297020.0	95233300.0	89770056.0
I - 3OH	16030477.0	47539304.0	30846364.0	28943404.0	37232860.0	65663570.0	71993250.0	54980640.0
I - 3OSO3	615563.4	51328900.0	31343028.0	29283376.0	36156560.0	57798760.0	53102504.0	56502052.0

Appendix 4 Table 15 – Sulfated and Unsulfated residue proportions for Heparin 85°C DMF solvated time course spectra, calculated from the integrals in Appendix 4 Table 14:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	1.9	9.3	9.3	6.1	5.8	7.1	1.4	9.2
A - NSO3 SM	81.1	0.3	0.0	0.5	0.0	0.0	0.9	0.0
A - NSO3 Slp	17.0	90.4	90.7	93.4	94.2	92.9	97.7	90.8
A - 6OH	25.7	1.5	2.4	2.9	1.4	1.7	2.1	1.6
A - 6OSO3	74.3	98.5	97.6	97.1	98.6	98.3	97.9	98.4
A - 3OH	99.9	4.4	4.2	3.4	3.6	4.3	0.4	4.6
A - 3OSO3	0.1	95.6	95.8	96.6	96.4	95.7	99.6	95.4
I - 3OH	96.3	48.1	49.6	49.7	50.7	53.2	57.6	49.3
I - 3OSO3	3.7	51.9	50.4	50.3	49.3	46.8	42.4	50.7

Appendix 4 Table 16 – Integral intensity values recorded for Heparin 55°C DMF solvated time course spectra from the scripts provided:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	369510.7	27386343.6	18794868.1	26097470.1	57791769.0	26702103.7	29689673.1	
A - NSO3 SM	15832894.2	0.0	0.0	0.0	0.0	0.0	0.0	
A - NSO3 Slp	3307875.7	41293294.2	46592101.0	50284554.8	34148624.9	64098312.1	61041296.8	
A - 6OH	8880859.0	2453908.7	1457679.7	2119479.9	19546428.7	2274505.4	2282236.1	
A - 6OSO3	25617190.8	105366767.3	98059248.1	106003638.2	104702082.3	116417570.2	121590818.9	
A - 3OH	24958595.9	8163406.2	5231999.2	4301279.9	13038580.4	5736705.6	7235764.1	
A - 3OSO3	32030.6	34391875.6	42730319.3	46663983.2	16075949.8	53075064.1	50469308.7	
I - 3OH	16030764.8	14314285.3	26133046.1	25151897.9	30525292.2	31783608.9	28002522.1	
I - 3OSO3	615584.5	42851741.3	41614308.2	46466169.4	25546722.2	49791650.4	50120287.5	

Appendix 4 Table 17 – Sulfated and Unsulfated residue proportions for Heparin 55°C DMF solvated time course spectra, calculated from the integrals in Appendix 4 Table 16:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	1.9	39.9	28.7	34.2	62.9	29.4	32.7	
A - NSO3 SM	81.2	0.0	0.0	0.0	0.0	0.0	0.0	
A - NSO3 Slp	17.0	60.1	71.3	65.8	37.1	70.6	67.3	
A - 6OH	25.7	2.3	1.5	2.0	15.7	1.9	1.8	
A - 6OSO3	74.3	97.7	98.5	98.0	84.3	98.1	98.2	
A - 3OH	99.9	19.2	10.9	8.4	44.8	9.8	12.5	
A - 3OSO3	0.1	80.8	89.1	91.6	55.2	90.2	87.5	
I - 3OH	96.3	25.0	38.6	35.1	54.4	39.0	35.8	
I - 3OSO3	3.7	75.0	61.4	64.9	45.6	61.0	64.2	

Appendix 4 Table 18 – Integral intensity values recorded for Heparin 35°C DMF solvated time course spectra from the scripts provided:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	369510.7	6609286.7	19023784.3	16875854.0	20273373.9	20697743.4	44191586.1	42624522.3
A - NSO3 SM	15832894.2	10161399.7	17175602.5	3372147.9	3191584.6	1353390.8	3353840.3	-2293660.0
A - NSO3 Slp	3307875.7	6416722.3	13178451.2	4448939.2	6666327.3	4698456.4	38881082.5	49866737.6
A - 6OH	8880859.0	19456561.3	3879431.3	5450483.9	2992510.9	1630917.5	3401208.3	2825257.3
A - 6OSO3	25617190.8	63281410.3	92897678.6	42145502.7	39518911.5	45057653.7	132371032.7	132160288.5
A - 3OH	24958595.9	31409503.6	33587125.2	11606644.4	7642154.0	6138870.6	13099234.6	7642189.5
A - 3OSO3	32030.6	205210.1	2577958.1	0.0	1033792.3	117347.5	25736697.3	37886302.4
I - 3OH	16030764.8	19651540.5	17480870.6	17228711.9	14243234.5	11825737.9	13391018.0	19600638.5
I - 3OSO3	615584.5	4125552.7	16283244.0	1862509.8	4631742.5	8683347.6	47106498.8	54456251.7

Appendix 4 Table 19 – Sulfated and Unsulfated residue proportions for Heparin 35°C DMF solvated time course spectra, calculated from the integrals in Appendix 4 Table 18:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	1.9	28.5	38.5	68.3	67.3	77.4	51.1	46.1
A - NSO3 SM	81.2	43.8	34.8	13.7	10.6	5.1	3.9	0.0
A - NSO3 Slp	17.0	27.7	26.7	18.0	22.1	17.6	45.0	53.9
A - 6OH	25.7	23.5	4.0	11.5	7.0	3.5	2.5	2.1
A - 6OSO3	74.3	76.5	96.0	88.5	93.0	96.5	97.5	97.9
A - 3OH	99.9	99.4	92.9	100.0	88.1	98.1	33.7	16.8
A - 3OSO3	0.1	0.6	7.1	0.0	11.9	1.9	66.3	83.2
I - 3OH	96.3	82.6	51.8	90.2	75.5	57.7	22.1	26.5
I - 3OSO3	3.7	17.4	48.2	9.8	24.5	42.3	77.9	73.5

Appendix 4 Table 20 – Integral intensity values recorded for TBA$^+$ Heparin 85^{o}C BMImCl solvated time course spectra from the scripts provided:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	369510.7	1796387.3	1524427.8	524336.0	149699.6	1701.3		7651436.8
A - NSO3 SM	15832894.2	1976438.2	1090115.5	17169.6	928037.7	307238.9		0.0
A - NSO3 Slp	3307875.7	5875765.8	3573410.9	4670185.5	998568.0	3741964.7		4732141.6
A - 6OH	8880859.0	1427224.0	466689.8	698158.7	191649.1	10161597.5		17882944.2
A - 6OSO3	25617190.8	15381877.9	11930109.9	9207255.5	6298608.9	5713637.6		3996587.8
A - 3OH	24958595.9	11632917.2	3169070.3	841421.6	453801.1	1283728.8		2545721.0
A - 3OSO3	32030.6	4415759.0	3137580.0	5465090.0	405725.4	2699061.0		0.0
I - 3OH	16030764.8	2132477.7	2883904.5	1037080.3	436579.0	3597189.7		4964633.9
I - 3OSO3	615584.5	3679055.6	2522504.1	3280571.9	914246.7	2349727.2		0.0

Appendix 4 Table 21 – Sulfated and Unsulfated residue proportions for TBA$^+$ Heparin 85^{o}C BMImCl solvated time course spectra, calculated from the integrals in Appendix 4 Table 20:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	1.9	18.6	24.6	10.1	7.2	0.0		61.8
A - NSO3 SM	81.2	20.5	17.6	0.3	44.7	7.6		0.0
A - NSO3 Slp	17.0	60.9	57.7	89.6	48.1	92.4		38.2
A - 6OH	25.7	8.5	3.8	7.0	3.0	64.0		81.7
A - 6OSO3	74.3	91.5	96.2	93.0	97.0	36.0		18.3
A - 3OH	99.9	72.5	50.2	13.3	52.8	32.2		100.0
A - 3OSO3	0.1	27.5	49.8	86.7	47.2	67.8		0.0
I - 3OH	96.3	36.7	53.3	24.0	32.3	60.5		100.0
I - 3OSO3	3.7	63.3	46.7	76.0	67.7	39.5		0.0

Appendix 4 Table 22 – Integral intensity values recorded for Na^+ Heparin 85°C BMImCl solvated time course spectra from the scripts provided:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	369510.7	4791224.9	13262002.2	3255818.1	1208041.9	60902482.8	28340462.3	7347375.8
A - NSO3 SM	15832894.2	97861649.9	60292723.9	61647946.2	44742140.4	26700889.9	18723548.2	0.0
A - NSO3 Slp	3307875.7	12712582.5	19057685.0	13401971.1	8438005.8	12281176.5	9994864.8	5302815.5
A - 6OH	8880859.0	50067827.9	31319261.8	31162028.5	22161414.6	49778787.5	31978618.1	19979765.0
A - 6OSO3	25617190.8	146359133.6	124121414.6	98774846.2	71525651.1	108202328.5	63007855.2	2871892.8
A - 3OH	24958595.9	162092389.4	92009953.0	90238924.8	64091962.3	58109332.1	33345578.6	3205981.5
A - 3OSO3	32030.6	515739.2	7331662.6	5184304.1	2206424.3	257883.3	770142.1	0.0
I - 3OH	16030764.8	80867618.7	56020086.0	49849864.5	37752785.6	65638295.2	36707873.8	5824652.4
I - 3OSO3	615584.5	3213613.2	11044031.1	6886862.7	4359985.1	363076.9	1771418.5	0.0

Appendix 4 Table 23 – Sulfated and Unsulfated residue proportions for Na^+ Heparin 85°C BMImCl solvated time course spectra, calculated from the integrals in Appendix 4 Table 22:

Assignment	Starting Material	30m	1h	2h	3h	4h	6h	24h
A - NH2	1.9	4.2	14.3	4.2	2.2	61.0	49.7	58.1
A - NSO3 SM	81.2	84.8	65.1	78.7	82.3	26.7	32.8	0.0
A - NSO3 Slp	17.0	11.0	20.6	17.1	15.5	12.3	17.5	41.9
A - 6OH	25.7	25.5	20.1	24.0	23.7	31.5	33.7	87.4
A - 6OSO3	74.3	74.5	79.9	76.0	76.3	68.5	66.3	12.6
A - 3OH	99.9	99.7	92.6	94.6	96.7	99.6	97.7	100.0
A - 3OSO3	0.1	0.3	7.4	5.4	3.3	0.4	2.3	0.0
I - 3OH	96.3	96.2	83.5	87.9	89.6	99.4	95.4	100.0
I - 3OSO3	3.7	3.8	16.5	12.1	10.4	0.6	4.6	0.0

Appendix 5 – Chapter 6 Differential Scanning Fluorimetry Melting Temperature Values

A5.1 – Differential Scanning Fluorimetry Protein-Carbohydrate Interaction Screening:

Appendix 5 Table 1: The melting temperatures and standard deviations of the FGFs and Cyclophilins Alone.

	FGF-1	FGF-2	CypB	CypD
Melting Temperature	48.4	54.6	55.1	50.3
Standard Deviation	0.05	0.07	0.03	0.05

A5.2 – DSF Analysis of the FGFs and Cyclophilins Alone:

Appendix 5 Table 2: The melting temperatures and standard deviations of the FGFs and Cyclophilins tested when combined with unmodified heparin, ι-carrageenan, κ-carrageenan, and xyloglucan polysaccharides.

Carbohydrate		FGF-1	FGF-2	CypB	CypD
Heparin	Melting Temperature	68.8	83.0	52.4	45.9
	Standard Deviation	0.04	0.10	0.02	0.12
ι-carrageenan	Melting Temperature	61.6	65.9	51.7	48.2
	Standard Deviation	0.28	0.33	0.03	0.11
κ-carrageenan	Melting Temperature	53.0	52.7	50.8	49.6
	Standard Deviation	0.03	0.12	0.02	0.04
Xyloglucan	Melting Temperature	48.9	50.2	54.5	50.0
	Standard Deviation	0.07	0.02	0.01	0.09

A5.3 – Effect of Carbohydrate Sulfation on FGF and Cyclophilin Stability:

Appendix 5 Table 3: The melting temperatures and standard deviations of the FGFs and Cyclophilins tested when combined with unmodified heparin and its modified sulfation patterns one, two, three, and four.

Carbohydrate		FGF-1	FGF-2	CypB	CypD
Unmodified Heparin	Melting Temperature	68.8	83.0	52.4	45.6
	Standard Deviation	0.04	0.10	0.02	0.10
Modified Sulfation Pattern One	Melting Temperature	64.9	83.6	51.7	45.9
	Standard Deviation	0.11	0.06	0.01	0.05
Modified Sulfation Pattern Two	Melting Temperature	65.0	82.5	52.4	45.8
	Standard Deviation	0.05	0.03	0.01	0.03
Modified Sulfation Pattern Three	Melting Temperature	66.4	82.3	51.5	48.8
	Standard Deviation	0.09	0.05	0.02	0.03
Modified Sulfation Pattern Four	Melting Temperature	66.2	78.4	51.5	48.8
	Standard Deviation	0.09	0.28	0.01	0.02

Appendix 5 Table 4: The melting temperatures and standard deviations of the FGFs and Cyclophilins tested when combined with unmodified κ-carrageenan and its modified sulfation patterns one, two, and three.

Carbohydrate		FGF-1	FGF-2	CypB	CypD
Unmodified κ-Carrageenan	Melting Temperature	53.0	52.7	50.8	49.7
	Standard Deviation	0.03	0.12	0.02	0.04
Modified Sulfation Pattern One	Melting Temperature	60.3	82.7	47.8	46.3
	Standard Deviation	0.08	2.65	0.01	0.05
Modified Sulfation Pattern Two	Melting Temperature	61.2	70.7	48.8	46.3
	Standard Deviation	0.13	0.37	0.02	0.08
Modified Sulfation Pattern Three	Melting Temperature	60.9	83.3	48.4	43.6
	Standard Deviation	0.10	2.82	0.03	0.06

Appendix 5 Table 5: The melting temperatures and standard deviations of the FGFs and Cyclophilins tested when combined with three modified cellulose polysaccharides.

Carbohydrate		FGF-1	FGF-2	CypB	CypD
Modified Sulfation Pattern One	Melting Temperature	52.6	67.1	52.6	50.2
	Standard Deviation	0.05	0.17	0.01	0.06
Modified Sulfation Pattern Two	Melting Temperature	50.4	58.4	54.2	50.2
	Standard Deviation	0.04	0.07	0.01	0.06
Modified Sulfation Pattern Three	Melting Temperature	53.4	62.9	53.8	50.4
	Standard Deviation	0.03	0.06	0.01	0.07